

*Genome Editing using Custom Endonucleases
and Dystrophin cDNA Insertions Targeting
Intron 1 of the Duchenne Muscular Dystrophy
Gene*

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE UNIVERSITY OF LONDON



December 2017

By

Marc Moore

Statement

Unless otherwise stated in text all the work contained in this thesis was conducted by the author, working in the School of Biological Sciences Royal Holloway University of London, between September 2013 and December 2017. All the work is original unless referenced by text. This work has not been submitted for any other degree at this or any other University.

Signed:

Date:

Abstract:

Duchenne muscular dystrophy (DMD) is a hereditary X-linked neuromuscular disease resulting from mutations across the *DMD* gene. The subsequent absence of the encoded dystrophin protein prevents the correct formation of the dystrophin-associated protein complex (DAPC), a structural link between the sarcolemmal actin cytoskeleton and the myofibre extracellular matrix. Dystrophin deficiency compromises myofibre stability giving rise to progressive muscle wasting, the prominent pathological feature of this disease. Over time, muscle deterioration results in loss of ambulation and upper body movement, decline in respiratory and cardiac function and ultimately premature death in the third decade of life. Current treatment is limited to corticosteroid use, which delays clinical manifestations but does not address the underlying aetiology of the disease. In addition, many therapies currently under development address specific mutation types or are reliant upon repeated administration; this serves to reduce their applicability to patients.

Genome editing refers to modification of genetic material at a precise genomic locus using special enzymes called customisable endonucleases. Recently, interest in genome editing approaches has increased with the advent of the clustered regularly interspaced palindromic repeats - Cas9 system (CRISPR/Cas9), in which an RNA guide can be used to direct the Cas9 nuclease to cleave the genome at a specific sequence.

In the present study a genome editing strategy was designed and evaluated to restore functional dystrophin expression from the endogenous *DMD* gene. The approach involved modifying intron 1 of the locus to insert dystrophin cDNAs downstream of the endogenous full-length dystrophin promoter and exon 1 elements, and was developed and verified in HEK293T cell and human DMD patient myoblast cultures. The approach was developed to permit full-length, mini-dystrophin and microdystrophin cDNA insertions to be examined. Specific CRISPR/Cas9 and guide RNAs (gRNAs) were designed and assessed for their ability to introduce double strand breaks (DSBs) at the *DMD* intron 1 locus. An exogenous DNA repair template construct, targeted to the DSB region in intron 1 was designed and further modified to contain a microdystrophin coding cDNA from *DMD* exon 2 onwards

Co-delivery of CRISPR/Cas9, gRNA and exogenous repair template constituents in HEK293T cells, resulted in specific integration of the microdystrophin cDNA at the *DMD* intron 1 site. This was confirmed at genomic, transcriptomic and protein levels in polyclonal and zeocin-selected clonal cell populations. Application of the genome editing system in human DMD myoblasts also gave rise to evidence of specific microdystrophin cDNA integration, although further refinement of the

methodology is required to ensure myoblast survival following positive selection with the zeocin antibiotic.

The research reported demonstrates the potential of genome editing the 5' end of the DMD gene and offers a strategy towards permanent corrective gene therapy applicable to almost all DMD patient genotypes.

Acknowledgements

What a rollercoaster undertaking this PhD has been, with its own inherent highs and lows. Completion of this thesis, feels much like disembarking a ride I am sporting additional frayed nerves, extra grey hairs and few years less life expectancy, but for the most part remain happy. First and foremost, I would like to thank my supervisory team and work colleagues from the Dickson/Popplewell lab groups. Notably, George and Linda for critical appraisal of the manuscript, for their patience, continued friendship and advice throughout the course of the PhD, your understanding of the DMD field and creative research ideas continue to both inspire and drive me to be a better scientist. Linda, I will particularly miss grabbing a cuppa, setting the science world to rights and discussing the next series of cool research proposals.

In addition the post-doctoral staff and PhD students past and present of the Dickson/Popplewell lab have been fantastic to work with. Three notable post-docs to mention are Hanna, Susie and Anita; thank you for not 'killing me in the face with a spoon' for taking over the cold-room and/or tissue culture. I have been fortunate to have your continued support, guidance and expertise throughout the PhD. My many peers have also had a significant role in keeping me almost sane; among the top culprits are Lukasz, Golnoush, James, Jade, Neda, Anila and Ilda. They are among many others but to include you all would be to write another thesis!

Further to my esteemed colleagues, I would also like to say a massive thanks to my family, who have been instrumental in me getting this far. First my parents, who not only fought tirelessly throughout my youth to ensure I had access to education, but have dedicated themselves to encouraging and supporting my continued academic pursuits as an "adult". Whether it be preventing me from starving with the delivery of food "I'll pay you next week" or listening to me resolving this week's insurmountable research problem out-loud and blindly agreeing; you truly have been a powerhouse of support for me. Although dad, contrary to your belief gained through flicking through my undergraduate practical books – after 7 years of work in the Biological Sciences, I can conclude that Eppendorf is in fact not Dumbledore's cousin and indeed of no relation to Dumbledore.

Next up – My sister who in addition to being a fantastic mum herself, continues to find time to be my best friend. You have never failed to make me laugh, even on the most trying days of optimising protocols in the lab and continue to remind me the importance of perspective and not taking yourself too seriously. Your kindness and compassion definitely make you the superior sibling – although I will deny this on all future occasions. This brings me to mention my two nephews Tyler and Charlie and my Niece Darcey; who during the course of my PhD and thesis have tried to keep me up to date with PlayStation games, introduced me to jump scares with 5 nights of Freddy's (run bruv,

run!), made creative additions to my unsupervised notes and collectively have provided me with comic genius and happiness – good job guys.

In fear of this sounding like an Oscar speech, I best conclude with my friends. Firstly, Ellen whom whilst on the other side of the world – has always remained only a skype or phone call away. I have lost count of how many times you have kept me sane or enthused with the prospect of travel! Secondly, Katie, Andrew, Liam, Danielle and Roberto (my HCS crew) – you have ensured the PhD experience had the right infusion of good food, alcohol and bad humour. In spite of the continued mocking of the way I speak, you remind me of my roots and keep me true to myself. (Marc out *mic drops*)

Notable Achievements

Publications:

Benedetti, S. Hoshiya, H. Uno, N. Ragazzi, M. Ferrari, G. Kazuki, Y. Moyle, L. Tonlorenzi, R. Lombardo, A. Chaouch, S. Mouly, C. **Moore, M.** Popplewell, L. Kazuki, K. Kato, M. Naldini, L. Dickson, G. Messina, G. Oshimura, M. Cossu, G. Tedesco, F. Reversible Immortalisation Enables Genetic Correction and Engineering of Next-Generation Human Artificial Chromosomes for Duchenne Muscular Dystrophy (Pre-publication EMBO 2017).

Prakash, V, **Moore, M.** Yanez-Munoz, J. R. (2016) Current Progress in Therapeutic Gene Editing for Monogenic Diseases. **Molecular Therapy**. Pages 465-474, 24 (3)

Moore, M. Vallese, D. Dickson, G. Popplewell, L. (2015) Exciting Developments in CRISPR/Cas9-mediated Genome Engineering Approaches for Duchenne Muscular Dystrophy. **Gene and Cell Therapy Insights**. Pages 215-230, 1 (2)

Loperfido, M. Jarmin, S. Dastidar, S. Perinis, I. Matteo, D. M. Perini, I. **Moore, M.** Nair, N. Samarakuko, E. Athanasopoulos, T. Tedesco, S. F. Dickson, G. Sampaolesi, M. Driesshe, V. T. Chuach, K, M. (2015) piggyBac transposons expressing full-length human dystrophin enable genetic correction of dystrophic mesoangioblasts. **Nucleic Acids Research**

Grants and Awards:

Santander Travel Grant (2016) Awarded £1720 was provided to facilitate the production of CRISPR Cas9 Protein and the development of a range of Cas9 protein techniques with Jean Paul Concordet at the Muséum National d'Histoire Naturelle (Paris).

University of London Post Graduate Research Fund (2016): Awarded £600 was provided to enable attendance of the FASEB Science Research Conference and facilitate the dissemination of my research.

Patents

Sequence optimised full length dystrophin patent

Abbreviations

6MWT	6 Minute walk test
AAV	Adeno-associated virus
AdV	Adenoviral
ActRIIB	Activin receptor IIB
ALK5	Activin linked kinase 5
AON	Antisense oligonucleotide
ATM	Ataxia-telangiectasia mutated
BGH	Bovine growth hormone
BLAST	Basic local alignment search tool
BMD	Becker muscular dystrophy
Bp	Base pairs
BRCA1	Breast and ovarian cancer type 1 susceptibility protein
CBh	Chicken beta actin hybrid promoter
CFTR	Cystic fibrosis transmembrane conductance regulator
CK	Creatine Kinase
CMV	Cytomegalovirus
CRE	Recombinase enzyme (causes recombination)
CRISPR/Cas9	Clustered regularly interspaced palindromic repeat/Cas9
crRNA	CRISPR RNA
CTGF	Connective tissue growth factor
CtIP	CtBP-interacting protein
DAPC	Dystrophin-associated protein complex
DNA	Deoxyribonucleic acid
DDR	DNA damage response
dHJ	Double holliday junction
DMD	Duchenne muscular dystrophy
DMEM	Dulbecco's modified Eagle's medium
DNA-PKcs	DNA dependent protein kinases catalytic subunit
DSB	Double Stranded Break
ECM	Extracellular matrix
EMBL/EBI	European bioinformatics institute
FACs	Fluorescence activated cell sorting
FCS	Foetal calf serum

GASP-1	Growth and differentiation factor associated serum protein
GD8	Growth and differentiation factor 8
GFP	Green Fluorescent Protein
HDR	Homology directed repair
HEK293T	Human Embryonic Kidney Cell Line containing Sv40 Large T antigen
HEK293T.Cas9	HEK293T containing Cas9 integrated at the AAVSI locus
HuDys-SO	Sequence optimised dystrophin cDNA
hDys	Native dystrophin cDNA
IL	Interleukin
InDels	Micro insertions or deletions
LIG4	Ligase IV
LV	Lentiviral
MD1	Microdystrophin 1
MD1 RT	Microdystrophin 1 Repair template
MD1ΔEx1	Microdystrophin 1 transgene with exon 1 removed
MD	Muscular dystrophies
MGN	Meganucleases
MMEJ	Microhomology-mediated end joining
Nbs1	Nibrin
NF-κβ	Nuclear factor – κβ
NHEJ	Non-homologous end joining
NMJ	Neuro-muscular junction
nNOS	Neuronal nitric oxide synthase
NOX2	NADH oxidase
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PolyA	Polyadenylation
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPA	Replication protein A
RTEL1	Regulation telomere elongation helicase 1
RYR1	Ryanodine receptor 1
SCID	Severe combined immunodeficiency

SMC	Structural maintenance of chromosomes
SSA	Single strand annealing
SSB	Single strand break
ssODN	Single stranded oligonucleotide
SV40	Simian virus 40
T7EI	T7 endonuclease I
TALEs	Transcription activator-like effectors
TALENs	Transcriptional activator like effector nucleases
TGFβ1	Transforming growth factor β1
TIMP	Tissue inhibitors or metalloproteases
tracrRNA	Trans-activating CRISPR RNA
mWRPE	Mutated woodchuck hepatitis virus post transcriptional element
XRCC4	X-ray cross-complementing protein 4
Zeo	Zeocin resistance cassette
ZFNs	Zinc Finger Nucleases

Table of Contents

Statement	2
Abstract:	3
Acknowledgements	5
Notable Achievements.....	7
Abbreviations.....	8
Chapter 1 Introduction	18
1.1 Muscular Dystrophies	18
1.2 Duchenne Muscular Dystrophy (DMD):	18
1.3 Molecular Pathogenesis of DMD.....	19
1.3.1 Calcium Dysregulation and Protease Activation:.....	19
1.3.2 Reactive Oxygen Species	20
1.3.3 Fibrosis and Inflammation.....	21
1.3.4 Multiple Combined Pathogenic Mechanisms:.....	23
1.4 DMD Diagnosis, Clinical Signifiers and Disease Progression and Treatment:	25
1.5 Gene Augmentation and Gene-Editing	26
1.6 Experimental Therapeutic Strategies for the Treatment of DMD:	28
1.6.1 Gene Augmentation Approaches:	28
1.6.2 Utrophin Modulation Strategies.....	29
1.6.3 Nonsense Mutation Read-Through Therapies:.....	30
1.6.4 Exon Skipping by Antisense Oligonucleotides	32
1.6.5 Myostatin Inhibition Studies:	33
1.7 Gene Editing Background and the DNA Damage Response	35
1.7.1 The Canonical Non-Homologous End Joining Pathway (NHEJ).....	36
1.7.2 The Homology Directed Repair Pathway:.....	37
1.8 Gene Editing Background: Customisable Nucleases and the Induction of a DSB:	41
1.8.1 Meganucleases.....	41
1.8.2 Zinc Finger Nucleases	42
1.8.3 Transcriptional Affector Like Effector Nucleases (TALENs).....	42
1.8.4 CRISPR/Cas9	43
1.9 Current Gene-Editing Attempts in the Context of DMD:.....	47
1.10 Project Aims:	52
Chapter 2 Materials and Methods	55
2.1 General Laboratory Reagents.....	55

2.2	Bioinformatics Assessments of the 5' End of the <i>DMD</i> gene	56
2.2.1	Common Intron 1 Sequence Identification	56
2.2.2	Rigorous Local Sequence Alignment:	56
2.2.3	Endonuclease Design.....	56
2.3	Bacterial Cultures	56
2.3.1	Materials	56
2.3.2	Heat Shock Transformation of Bacteria:.....	57
2.3.3	Production of Starter Cultures and Glycerol Stocks	57
2.3.4	Mini, Maxi and Mega Preparations:	57
2.3.5	Picking Colonies for Colony PCR:	58
2.4	Plasmids	58
2.4.1	Gene- Editing Plasmids.....	58
2.4.2	Dystrophin Plasmids.....	60
2.4.3	Repair Template Plasmids	61
2.5	Computer Software used throughout Thesis:	61
2.6	Descriptive Statistical analysis:.....	61
2.6.1	Inferential Data Analysis.....	62
2.7	Molecular Sub-Cloning Methods.....	62
2.8	Guide Sub-Cloning Methods	64
2.8.1	Background	64
2.8.2	Human Primers	64
2.8.3	Mouse Primers	64
2.8.4	Preparation of the U6 Guide Plasmid	64
2.8.5	Self- Annealing Oligonucleotides.....	64
2.8.6	Ligation.....	65
2.8.7	U6 Guide Colony PCR	65
2.9	Sub-cloning Microdystrophin into the pUC57-hINT1-RT.....	65
2.9.1	Preparation of plasmids	65
2.9.2	Microdystrophin Colony PCR.....	66
2.10	Tissue Culture.....	66
2.10.1	General Tissue Culture Reagents:.....	66
2.11	Transient Transfection Protocols	69
2.11.1	Cell Densities for 6 well Plate	69
2.11.2	Transfection Reagents and Ratios:	69

2.11.3	Viafect Transient Transfection	69
2.11.4	Lipofectamine Based Transient Transfection:	70
2.12	Flow Cytometric Analysis	71
2.12.1	Materials	71
2.12.2	Analysis of GFP Expression in Cell Culture:.....	71
2.12.3	Analysis of Flow Cytometric Data	72
2.13	DNA/RNA Extraction and Quantification from Cell Culture:.....	72
2.13.1	Materials:	72
2.13.2	Cell Pelleting for Extraction	72
2.13.3	DNA Extraction	73
2.13.4	RNA Extraction:	73
2.13.5	Nucleic Acid Quantification	73
2.14	Protein Extraction and Quantification:.....	74
2.14.1	Materials	74
2.14.2	Protein Extraction:	74
2.14.3	Protein Quantification:.....	74
2.15	Western Blotting	75
2.15.1	Materials	75
2.15.2	Antibodies	75
2.15.3	Sample Preparation.....	76
2.15.4	Gel Preparation and Electrophoresis.....	76
2.15.5	Electro-transfer to Nitrocellulose Membrane.	76
2.15.6	Post-Transfer Nitrocellulose Membrane Checks and Blocking	77
2.15.7	Visualisation of the Nitrocellulose Membrane: The ECL Method	77
2.15.8	Visualisation of the Nitrocellulose Membrane: The Odyssey Method	78
2.16	Quantification of Dystrophin.....	78
2.16.1	Software.....	78
2.16.2	Methods.....	78
2.17	Polymerase Chain Reactions (PCR).....	79
2.17.1	Materials	79
2.18	Mismatch Detection Assays:	80
2.18.1	Materials	80
2.18.2	Human Intron 1 High Fidelity PCR:	80
2.18.3	Mouse Intron 1 High Fidelity PCR:	80

2.18.4	Surveyor Assay	81
2.18.5	The T7 Endonuclease I.....	81
2.19	Indel Quantification Approaches.....	82
2.19.1	Gel Based Quantification Methods:	82
2.19.2	Agilent Bioanalyser Analysis	83
2.19.3	Tracking of Indels by DEcomposition (TIDE)	84
2.20	In Vitro Transcription of mRNA guides	84
2.20.1	PCR Amplification and Purification of Extended Guide Oligonucleotides:.....	85
2.21	In Vitro RNA Transcription of Guides and Clean-up:	85
2.22	In Vitro Transcription of Cas9	86
2.22.1	Materials	86
2.23	Cas9 Protein Production.....	88
2.23.1	Materials (Work Undertaken in Paris):.....	88
2.23.2	In Vitro Cleavage Assay using Cas9 protein and In Vitro Transcribed RNA:	90
2.24	Zeocin Selection Methods	90
2.24.1	Materials	90
2.24.2	Generation of a Zeocin Kill-Curve:.....	90
2.25	Production of Polyclonal MD1 Cell lines.....	90
2.26	Generation of HEK293T-MD1 Monoclonal Cultures.....	91
2.26.1	Materials	91
2.26.2	Culturing Cells at Low Density to Attain Single Colonies:	92
2.26.3	Identification and Picking of Individual Colonies	92
2.27	Molecular Characterisation of Colonies	93
2.27.1	Integration PCRs	93
2.27.2	Reverse Transcriptase PCR	93
2.27.3	MD1 Western Blotting	94
Chapter 3 In Silico Analysis of the 5'-End of the <i>DMD</i> Gene and Design of Customisable Endonucleases:		95
3.1	Introduction	95
3.1.1	Human <i>DMD</i> Gene	95
3.1.2	Murine <i>Dmd</i> Gene:.....	97
3.2	Chapter Aims:.....	99
3.3	Results:.....	100

3.3.1	Mapping the 5' End of the <i>DMD</i> Gene to Locate Full Length Dystrophin Promoters and Exon 1 Elements Relative to the Common Exon 2 Sequence:	100
3.3.2	Rigorous Local Pairwise Sequence Alignment of Human and Mouse Intron 1 Sequences: the Identification of a Region of Homology within <i>DMD</i> Intron 1.....	103
3.3.3	Identification of TALEN Target Sites and CRISPR Guide Designs to the 5' End of the <i>DMD</i> Gene:	104
3.3.4	Sub-Cloning of Guide Sequences.....	118
3.3.5	Retrospective Bioinformatics Analysis of the Specificity of Guide RNA Sequences Directed to the Homologous Region in <i>DMD</i> Intron 1:	123
3.4	Discussion	128
3.5	Conclusion.....	131
Chapter 4 Establishing a Delivery System for the Cas9 Nuclease and Validation of Guide RNA Efficacy at the Human <i>DMD</i> Intron 1 Site		
		133
4.1	Introduction	133
4.2	Chapter Aims:.....	134
4.3	Results:.....	135
4.3.1	An Examination of Cas9 Nuclease Delivery and Expression using HEK293T Cells as a Model System:	135
4.3.2	Establishing a Western Blotting Protocol for the Detection of Cas9 Protein.....	135
4.3.3	Examination of Cas9 Protein Expression Following Transient Transfection in HEK293T Cells with Dose Escalation.....	136
4.4	Establishing the Surveyor and T7 Endonuclease (T7E1) Cleavage Assays for the Detection of Targeted Genome Modification.....	141
4.4.1	Efficient PCR Amplification of the Human <i>DMD</i> and Mouse <i>Dmd</i> Intron 1 Locus	141
4.4.2	Efficacy of Human Guides and Optimisation of the Surveyor and T7E1 Enzyme Mismatch Detection Assays.....	141
4.5	Assessing Non-Viral Delivery Strategies for Cas9 Nuclease into Immortalised Δ 45-52 Patient Myoblasts.....	153
4.5.1	Transient Transfection of Px458 to Immortalised Patient Myoblasts Harbours Deletions of Exon 45 -52	153
4.6	Establishing In-Vitro Transcription of guide RNAs and Delivery Approaches in the Context of Gene-Editing.....	169
4.6.2	Production and Genome Editing Efficiency of In-Vitro Transcribed Guide RNAs:.....	169
4.6.3	Production of In-Vitro Transcribed Cas9 mRNA and Genome Editing Efficacy in Immortalised Δ 45-52 Patient Myoblasts:.....	170
4.6.4	Production and Assessment of Cas9 Protein	178
4.7	Discussion	188

4.7.1	Establishing Mismatch Detection Assays.....	188
4.7.2	Assessing the Efficacy of In Silico Guide Designs	189
4.7.3	Px458-Guide 1 Efficacy	189
4.7.4	Alternative Cas9 Delivery Approaches	191
4.7.5	Strategies to Enhance Gene-modification:.....	193
4.8	Conclusion:.....	195
Chapter 5 Dystrophin Transgene Sequence Optimisation and Putative Design of cDNA-based Repair Templates: 196		
5.1	Introduction:	196
5.1.1	Minidystrophin and Microdystrophin Transgenes	196
5.1.2	Sequence Optimisation of Dystrophin Transgenes.....	200
5.1.3	Chapter Aims:.....	202
5.2	Results:.....	203
5.2.1	Assessment of Dystrophin Expression from Native and Optimised Fused Dystrophin-eGFP Expression Constructs.....	203
5.2.2	Quantification of the Difference in Expression of Native and Optimised Dystrophin Constructs Driven by a CMV Promoter:	206
5.2.3	Quantifying the Difference in Expression of Native and Optimised Dystrophin Constructs Driven by a Spc512 Promoter:	212
5.2.4	An Examination of Native and Optimised Microdystrophin Constructs Driven by the Spc512 Promoter:	212
5.2.5	Design of an Exogenous Repair Template Plasmid Backbone Compatible with multiple dystrophin cDNA Variants:.....	217
5.2.6	Sub-cloning MD1 into the pUC57 Intron 1 Exogenous Repair Template:	222
5.3	Discussion	228
5.3.1	Sequence Optimised Dystrophin cDNAs:.....	228
5.3.2	Design and Construction of Exogenous Repair Template Directed to DMD Intron 1: 231	
5.4	Conclusions	233
Chapter 6 An Examination of CRISPR-mediated Microdystrophin Knock-in Downstream of Full-length Endogenous DMD Promoters.		
6.1	Introduction:	234
6.2	Chapter Aims:.....	235
6.3	Results.....	235
6.3.1	Determination of Zeocin Sensitivity of HEK293T and Immortalised DMD Patient Myoblasts.....	235

6.3.2	Transient Transfection of HEK293T Cells with Cas9, Guide RNA and Repair Template Plasmids and Polyclonal Enrichment using Zeocin Selection:	240
6.3.3	Dual Transfection Experimental Design	240
6.3.4	Positive Selection: The Enrichment of Stably Transfected HEK293T-MD1 Cells	246
6.3.5	Molecular Characterisation of Zeocin-Enriched Polyclonal Cultures and Establishing Single Colonies Allowing for Clonal Expansion:	253
6.3.6	Establishing Single Colonies: Further Enrichment and Molecular Characterisation: .	260
6.3.7	Patient Myoblasts: A Parallel CRISPR Cas9- HDR Mediated Knock-in of the MD1 Transgene:	272
6.3.8	Assessment of Transient Transfection Efficiency:	272
6.4	Discussion	278
6.4.1	HEK293T, HDR-mediated MD1 Transgene Integration:.....	278
6.4.2	Stability of HDR-mediated MD1 Transgene Integration in Δ 45-52 Patient Myoblasts, 282	
6.4.3	Transfection Efficiency of Immortalised Δ 45-55 Patient Myoblasts:.....	282
6.5	Conclusion.....	287
Chapter 7	General Discussion	289
7.1	Current Research Considerations:.....	289
7.1.1	Identify and Resolve the Source of the MD1 Truncation.....	289
7.1.2	Assess the Region of Homology Identified between Human and Mouse Intron 1 for Potential Functionality:	290
7.2	Enhancement of Transgene Integration Approaches	296
7.2.1	Advances in CRISPR Guide Design: A Prospect for Improved CRISPR guides	296
7.2.2	Improved Homology Directed Repair and the Prospect for Non-Homologous End Joining-mediated Integration Approaches	300
7.3	In Vivo Translation:	304
7.3.1	Cell Engraftment Approaches in Pre-clinical Animal Models and DMD Patients.....	304
7.3.2	Viral Delivery Strategies	308
7.3.3	Non-viral Delivery Approaches:.....	312
7.4	Concluding Remarks.....	313
Bibliography	314

Chapter 1 Introduction

1.1 Muscular Dystrophies

Muscular Dystrophies (MDs) is an encompassing term for more than 30 genetically distinct conditions that affect muscle-associated proteins, culminating in alterations in muscle morphology, muscle wasting and paralysis (Emery, 2002; Mercuri & Muntoni 2013). Classified on the basis of the clinical manifestation, muscle wasting, these dystrophies have heterogeneous clinical onset, progression and varying degrees of cardiac involvement (Emery 2002; Mercuri & Muntoni 2013). Currently, there is no cure for muscular dystrophies, only palliative care such as corticoid steroids to treat the secondary manifestations; leaving an unmet medical need to treat the underlying genetic aetiology of the disease (Bushby et al. 2010a; Bushby et al, 2010b).

1.2 Duchenne Muscular Dystrophy (DMD):

Duchenne Muscular Dystrophy (DMD) is an X-linked inherited condition affecting 1 in 5000 live male births (Guiraud et al, 2015; Bladen et al. 2015). The condition arises from mutations across the *DMD* gene that results in the destruction or premature termination of the *DMD* open reading frame, as a consequence, dystrophin protein is not synthesised (Monaco et al. 1988). The complex expression profile of the *DMD* gene, being driven by multiple promoters and producing multiple isoforms, accounts for the progressive muscle wasting and non-progressive neurological manifestations of the disease observed in patients (Nudel, 2005). Whilst the full gene was cloned three decades ago, and identified as a candidate for gene therapy, the genetic aetiology of the disease has proved challenging to address. This is due to the large size of the gene, with 79 exons spanning 2.4Mbs of the genome and encoding an 11.2kb mRNA (Koenig et al. 1987).

Another challenge to the effective treatment of this disease is the diverse mutation profile exhibited by DMD patients, with mutation subsets and their occurrence being: large deletions 68%, duplications 12% and small mutations 20% (Bladen et al. 2015). These mutations are observed to show clustering between exons 2-20 and 45-55, but can occur throughout the full length of the *DMD* gene, being either inherited or arising spontaneously (Bladen et al. 2015). In contrast, Becker Muscular Dystrophy (BMD), a clinically milder form of muscular dystrophy, occurs in instances where the mutation is present without the disruption of the reading frame (Monaco et al. 1988; England et al. 1990). Subsequently, an internally truncated form of dystrophin is synthesised with partial function; the discovery of this allelic form, prompted interest towards mutation-specific therapeutic approaches that would restore the *DMD*

reading frame. This would enable the conversion from a Duchenne to a Becker Muscular phenotype, with the outlook of stabilising or slowing disease progression.

1.3 Molecular Pathogenesis of DMD

In muscle, dystrophin protein forms a structural link between F-actin and the sarcolemmal dystrophin-associated protein complex (DAPC) and thus an indirect connection to laminin, a component of extracellular matrix (Hoffman et al. 1987; Koenig et al. 1988; Ervasti et al. 1990). It speculated that dystrophin, allows for lateral dispersion of mechanical stress throughout the muscle, that occurs during exercise; thereby preventing contractile-induced injury (Lovering & De Deyne 2004; Bushby et al. 2010a; Ervasti et al. 1990). However, growing research has implicated dystrophin in signalling roles, in addition to the maintenance of muscle integrity as reviewed by (Allen et al. 2016).

1.3.1 Calcium Dysregulation and Protease Activation:

In muscle contraction, stimulation at the neuromuscular junction (NMJ), causes electrical excitation of the voltage sensing dihydropyridine receptor (DHPR) a modified L-type calcium channel, in the T-tubules (Allen et al. 2016; Anderson & Meissner 1995). These dihydropyridine receptors interact with the ryanodine receptor (RYR1), promoting calcium release from the sarcoplasmic reticulum (Blat & Blat 2015). Ultimately, within a sarcomere this culminates in calcium binding to troponin, the displacement of tropomyosin and myosin-actin mediated muscle contraction. As a consequence of this contraction mechanism, cytosolic concentrations of calcium are tightly regulated in muscle.

Interestingly, the cytosolic concentration of calcium ions appears to be elevated in dystrophin-deficient muscle (Turner et al. 1991; Turner et al. 1993; Hopf et al. 1996); this is speculated to be due to microlesions in the sarcolemma (Clarke et al. 1993; Blat & Blat 2015) or the localised effects of mechanosensitive channels (Whitehead et al. 2006; Iwata et al. 2009), which serve to increase sarcolemma permeability. The prolonged increase in intracellular calcium concentration, in turn, is implicated in the interference of normal excitation-contraction coupling and is thought to render muscle more susceptible to necrosis (Allen et al. 2016). In combination, this would give rise to muscle weakness, the most prominent clinical manifestation of DMD (Bushby et al. 2010a).

Evidence in support of high calcium levels interfering with excitation-contraction coupling was afforded by observations in the *mdx* mouse muscle where calcium ion release was reduced following stimulation at the NMJ by both single and successive action potentials (Hollingworth et al. 2008). Further studies indicated that the reduced release of calcium ions could be accounted for by less frequent RYR1 opening or a depleted calcium store as a consequence of

impaired calsequestrin buffering (Doran et al. 2004). Interestingly, this trend is speculated to be retained in human muscle, where following strenuous exercise the increased cytosolic concentration of calcium, is associated with a reduction in muscle force following low frequency stimulation at the NMJ (Edwards et al. 1977). Thus in dystrophic muscle, where the sarcolemma is speculated to be more permeable, this could provide one explanation of muscle weakness.

An alternative theory is that calcium cytotoxicity, renders muscle more susceptible to necrosis (Tidball & Spencer 2000). Whilst evidence suggests calpains play a role in muscular dystrophies, the exact mechanism of action remains to be elucidated. It is theorised these proteases could directly target and degrade muscle proteins or target transcriptional factors resulting in dysregulated muscle protein expression patterns (Allen et al. 2016). One interesting study implicated calpain in the direct destruction of junctophilin 1 and 2 proteins, a structural protein that maintains correct separation between the sarcoplasmic reticulum and the T-Tubules (Murphy et al. 2013). If the structural integrity of this protein were compromised it could certainly account for depleted sarcoplasmic and higher cytosolic calcium levels. While a plethora of theories exist as to molecular mechanisms, the research undertaken certainly suggests the perturbation of calcium homeostasis has a role to play in muscular dystrophy pathogenesis.

1.3.2 Reactive Oxygen Species

Oxidative stress largely refers to increased cytosolic concentrations of reactive oxygen species (ROS), a reduction in ROS scavengers and ROS-mediated dysregulation of signalling pathways (Allen et al. 2016; Jones 2006). Interestingly, dystrophic muscles are theorised to fulfil these criteria, with an increase in NADH (oxidase) NOX2 and resultant ROS (Whitehead et al. 2010), a depletion of glutathione a ROS scavenger (Renjini et al. 2012) and dysregulation of inflammatory and autophagy pathways all having been described.

Early evidence in support of the oxidative stress theory, provided by Rando et al , indicated that *mdx* myotubes were more susceptible to oxidative stress, with diminished viability observed on exposure to lower concentrations of reactive oxygen species, relative to wild type (Rando et al. 1998). Interestingly, when the group performed the same experiments on myoblast precursor cells, no difference in *mdx* and wild-type susceptibility to ROS was observed. As dystrophin protein is expressed in wild-type myotubes, they suggested a link between dystrophin and the modulation of ROS in muscle. Further evidence in support of this putative linkage, was afforded in studies expressing minidystrophins or microdystrophins,

these being partial dystrophin transgenes bearing internal deletions, in *mdx* myotubes which demonstrated an improved resilience to reactive oxygen species (Disatnik et al. 1998).

Although the exact mechanism by which dystrophin absence causes an elevation of ROS in muscle has not yet been fully elucidated, the dysregulation of NOX2, a sarcolemma protein with mechanosensitive activation is implicated (Whitehead et al. 2010). One theory indicates that aberrant arrangement of microtubules may cause increased activation of NOX2 and subsequently increased production of ROS and calcium dysregulation (Khairallah et al. 2012). Alternatively, the localisation of nNOS to the cytosol, which is usually bound to spectrin repeats 16 and 17 on dystrophin protein, has also been implicated in the activation of NOX2 (Lai et al. 2009; Allen et al. 2016).

Regardless of the source of activation, studies of NOX2 have illustrated that its activation results in increased ROS production and its inhibition reduces ROS concentration in dystrophic muscle (Khairallah et al. 2012). Important to the pathology of DMD, reactive oxygen species are implicated in the activation of stretch-activated calcium channels causing calcium influx, protein (Hauser et al. 1995) and lipid damage (Messina et al. 2006), activation of inflammatory pathways and reduction in autophagy of damaged organelles (Allen et al. 2016). In combination, these outcomes are anticipated to lead to an increase in the muscle cell death.

1.3.3 Fibrosis and Inflammation

Fibrosis refers to the excessive deposition of extracellular matrix (ECM) proteins into tissues during regenerative processes that follow injury (Kharraz et al. 2014). In the context of DMD, chronic degeneration-regeneration cycles cause muscle fibrosis. This contributes to dystrophic muscle pathology in two main ways. Firstly, the direct replacement of skeletal muscle with non-functional fibrotic tissue serves to reduce contractility. Secondly, fibrotic tissue forms a physical barrier serving to impede blood perfusion (Blat & Blat 2015). By extension these effects also provide potential limitations for many therapeutic strategies being trialled by forming a physical barrier and reducing blood flow to the muscle tissue, thus reducing the access of therapeutics (Negroni et al. 2016).

Two principle growth factors are implicated in the causation of fibrosis, these being transforming growth factor β 1 (TGF β 1) and connective tissue growth factor (CTGF). TGF β 1 is present in a latent form in the ECM and evidence indicates it is activated both in response to muscle damage and by invading cells of the immune system (Horiguchi et al. 2012; Wynn 2008). This activation normally occurs as a natural regenerative response to muscle injury, but occurs continuously or repetitively in DMD progression. Once activated TGF β 1 binds through two receptors called activin linked kinase 5 (ALK5) and TGF β type II receptor, the downstream

phosphorylation cascade causes SMAD protein mediated changes in gene expression (Massague et al. 2005). The net result is stimulation of fibroblastic production of profibrogenic proteins and the downregulation of tissue inhibitors or metalloproteases (TIMP), which culminates in excessive deposition of extracellular matrix proteins in muscle (Kharraz et al. 2014). Observations on the administration of TGF β 1 in both *in vitro* and *in vivo* systems show that TGF β 1 induces further TGF β 1 expression in other cells in a positive cycle, and promotes the production of connective tissue (Brandan et al. 2008; Li et al. 2004). Importantly, it was also demonstrated that transfected cells expressing TGF β 1 that were then intramuscularly injected into SCID mice, differentiated to form myofibroblastic cells. Interestingly, this could be reversed upon sequestration of TGF β 1 with decorin, which directly binds to the growth factor interfering with downstream activation processes (Li et al. 2004). The second factor involved in fibrosis, CTGF, is induced by TGF β 1, and is also present in the ECM. CTGF was shown to be more potent than TGF β 1 in inducing the production of connective proteins in fibroblasts (Frazier et al. 1996; Kharraz et al. 2014).

In addition to the activation of fibrosis, the infiltration of immune cells such as macrophages, T-cells and neutrophils are implicated in heightening muscle lesions and dystrophic pathology (Evans et al. 2009). Interestingly, inflammatory signalling could be a major influence on the balance of muscle degeneration and regeneration processes. Macrophages as a case in point are present in *mdx* muscle in two distinct populations, M1 and M2, these being pro- and anti-inflammatory respectively (Shin et al. 2013). Interestingly, treatment of mice with IFN- γ , an inflammatory cytokine, was found to heighten muscle damage and deactivate M2 macrophages (Villalta, Deng, et al. 2011); contrastingly treatment with the anti-inflammatory cytokine, IL-10, promoted regeneration (Villalta, Rinaldi, et al. 2011). Thus the modulation of macrophages could be a potential therapeutic strategy in DMD.

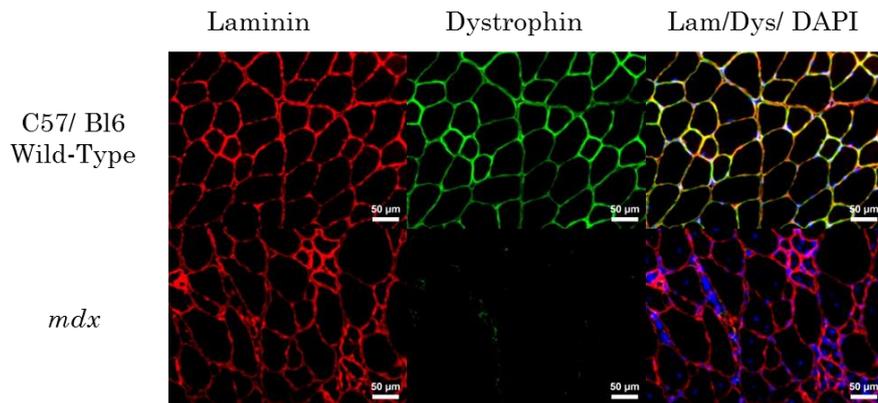
Furthermore, nuclear factor- κ B (NF- κ B) proteins, a family of transcription factors, have been implicated to heighten dystrophic pathology in muscle. The NF- κ B family comprises five members: RelA (p65), Rel-B, c-Rel, p100/p50 and p52, which can undergo homo- or heterodimerisation depending upon the source of stimulation. Under normal conditions, NF- κ B is inactivated in the cytosol by association with I κ B α , which masks the nuclear localisation signal (NLS) (Karin 1999). In the presence of pro-inflammatory signals, such as interleukin-1, I κ B kinase causes I κ B α and NF- κ B to dissociate liberating the NF- κ B and enabling translocation to the nucleus and modification of gene expression (Karin 1999). Importantly, NF- κ B was implicated to worsen muscle pathology in three main ways: i) augmentation of the ubiquitin-proteasome system, which would promote the degradation of muscle proteins; ii) enhancement of expression of pro-inflammatory mediators such as cytokines and chemokines

which heighten muscle damage and loss; iii) reduced myofibre regeneration (for a review see (Li et al. 2008). Studies have shown that deletion of a single allele of the NF- κ B p65 subunit reduces infiltration of macrophages and associated fibrosis (Acharyya et al. 2007). Consistent with this, inhibition of NF- κ B was also associated with improved regenerative capacity of muscle derived stem-cells post transplantation (Lu et al. 2012). Thus the modulation of this factor is currently under investigation with the outlook of slowing clinical progression in DMD patients (Mah et al. 2014).

1.3.4 Multiple Combined Pathogenic Mechanisms:

The above theories highlight that the pathogenesis of DMD is a multifaceted process and the summative effects of impaired calcium homeostasis, ROS dysregulation and inflammation provides a continual cycle of muscle damage and subsequent muscle necrosis. This muscle damage stimulates satellite cells, denoted as adult muscle stem cells, to activate and promote muscle regeneration (Lipton & Schultz 1979). Importantly, after chronic cycles of muscle degeneration and regeneration the response can become insufficient to address the muscle fibrosis and necrosis resulting in muscle wasting (Boldrin et al. 2015). The reasons for this are still a source of debate: original theories suggested that satellite cells may become exhausted with reduced regenerative propensity (Morgan & Zammit 2010) or depleted (Heslop et al. 2000). However, later observations that satellite cell numbers were elevated in DMD patients, and that satellite cells derived from both adult and young *mdx* mice have the same regenerative capacity on engraftment placed these theories under scrutiny (Kottlors & Kirschner 2010; Boldrin et al. 2015). Current theories suggest that the dystrophic muscle environment, such as chronic inflammation and fibrosis negatively impact upon satellite cell regeneration, reducing their ability to effectively regenerate muscle following injury (Boldrin et al. 2015). A flow chart linking the pathogenesis can be seen in Figure 1-1.

A



B

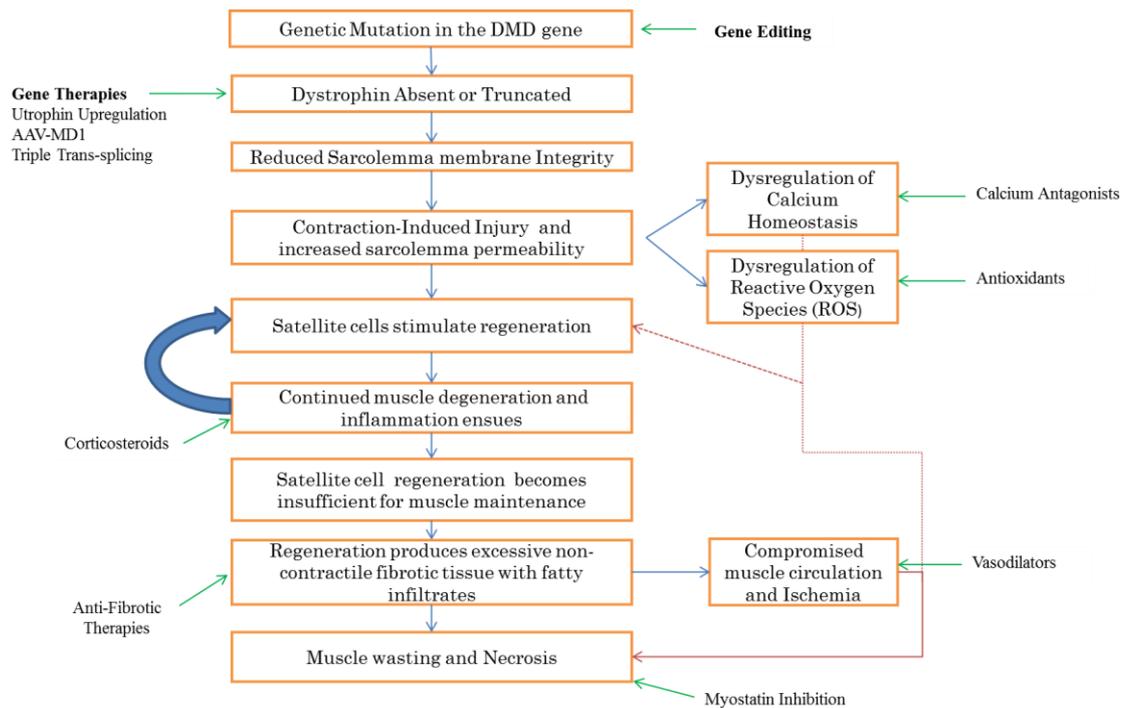


Figure 1-1 Multifaceted pathogenesis of DMD and routes to therapeutic intervention.

A) Immunohistochemistry staining from transverse 10 μ M, muscle sections from C57 Black 6 wild type and H2kb *mdx* Tibialis Anterior. Importantly, laminin is stained in red, dystrophin in green and nuclei in blue with DAPI. Images were taken at 200x magnification and kindly provided by Dr. Ngoc Lu-Nguyen. Importantly they showcase hallmarks of dystrophic muscle, with *mdx* showing inconsistent fibre size, an absence of dystrophin stain and central nucleation a marker of regeneration. B) A schematic flow-chart to illustrate DMD pathogenesis and proposed links between dystrophin absence and muscle wasting. Notably, proposed interventions and at what stage they would treat the pathology is also shown, denoted by green arrows. These include: gene-editing (Poplewell et al. 2013), gene therapy (Le Guiner et al. 2017; Koo et al. 2014; Tinsley et al. 2011), calcium antagonists (Phillips & Quinlivan 2008), antioxidants (Stern et al. 1982), vasodilators (Martin et al. 2012), myostatin Inhibition (Lu-Nguyen et al. 2015), anti-fibrosis and inflammatory therapies (Traynor 2017), and corticosteroids (Mercuri & Muntoni 2013).

1.4 DMD Diagnosis, Clinical Signifiers and Disease Progression and Treatment:

Importantly, DMD patients are born with a genetic mutation in the DMD gene; in two thirds of cases there is a familial history and in one third the mutation occurs spontaneously (Bladen et al. 2015). However, DMD often goes undiagnosed until 2 - 5 years of age, where delayed locomotor milestones, walking with an unsteady gait or delayed attainment of cognitive milestones present as the initial symptoms (Emery 2002; Bushby et al. 2010a; Allen et al. 2016).

Further indicators of disease, leading to diagnosis include: (i) the 'Gower manoeuvre', in which during rising from the floor, a child with DMD uses his arms to walk up his thighs, pushing up off of them to extend his hip and trunk muscles to attain a standing position (Emery 2002); (ii) blood samples taken for examination of creatine kinase (CK) levels, which can be elevated 100-fold in DMD patients (Zatz et al. 1991); (iii) muscle biopsy taken and histological analysis performed with an antibody for dystrophin, to assess presence or absence of the protein and finally genetic screening in an attempt to identify the mutation type (Bushby et al. 2010a).

Notably, the CK levels are a non-specific clinical marker, indicating general muscle damage in the case of DMD. This stated, CK can be elevated in response to exercise, alcohol and cardiac conditions (Allen et al. 2016). Thus additional forms of confirmation including muscle biopsy are necessitated. During childhood, ambulation and mobility of the child will continue to decline and the use of a wheelchair normally occurs between 7-12 years of age (Allen et al. 2016). As muscle degeneration becomes more prominent in teenage years and up to the 20s, severe respiratory and cardiac involvement arises. Respiratory support is provided in the form of antibiotic treatments for respiratory infections, Surgical intervention for spinal curvature, manual and assisted cough techniques, non-invasive daytime and nocturnal ventilation and as a final resort, a tracheostomy is undertaken (Birnkrant et al. 2010).

Cardiac involvement normally manifests in DMD as cardiomyopathy or arrhythmia (Bushby et al. 2010b; Fayssoil et al. 2017). The current multidisciplinary management of DMD, and advancements in care in terms of corticosteroids use and non-invasive ventilation has served to improve life expectancy in DMD patient to 20-30 years of age (Bushby et al. 2010b; Allen et al. 2016). Currently cardiomyopathy is the leading cause of mortality in DMD patients and requires further investigation (Bushby et al. 2010b; Allen et al. 2016).

Corticosteroids are currently the mainstay treatment regime of DMD (Bushby et al. 2010a; NICE 2016). These are used to dampen the inflammatory responses in muscle resulting from

continual muscle damage and cycles of degeneration and regeneration, serving to slow the clinical progression of the disease. Whilst adherence to a corticosteroid programme, such as prednisolone serves to maintain muscle strength and function, prolonging ambulation, long term administration is associated with a multitude of undesirable side effects. These include: stunted growth, bone demineralisation, increased mood-swings and weight gain (NICE 2016). This side-effects may independently require management including, dietetic advice to address bone fragility and psychosocial support. Various dosing regimens are utilised in an attempt to mitigate these side effects (Bushby et al. 2010a).

1.5 Gene Augmentation and Gene-Editing

Gene Therapy is defined as 'a set of strategies that modify the expression of an individual's genes or repair abnormal genes' (ASGCT 2014). This definition is reflective of the subdivision of gene therapy into two broad approaches: firstly, gene augmentation, which aims to deliver a new functional copy or variant of a gene, to enhance protein expression in place of the endogenous mutated gene; and secondly, gene editing, which aims to directly modify the genome for the purpose of gene repair (ASGCT 2014). Both have the common goal of ameliorating disease-causing genetic mutations.

Gene augmentation therapies utilise non-viral or viral systems to deliver a functional copy of the gene (Nayerossadat et al. 2012). Viral systems are the more efficient platform used for the delivery of genetic material (Nayerossadat et al. 2012). Notable examples include Adenoviruses, these viruses have large transgene capacity of up to 37kb particularly if the gutted format is used (Gonçalves & de Vries 2006; DelloRusso et al. 2002). This makes them the most suited for large transgenes, however, some serotypes are associated with immunogenic side effects upon administration, which has served to limit their use (Machitani et al. 2011; Alonso-Padilla et al. 2016; Teigler et al. 2014). Adeno-associated viruses (AAV), this small parvovirus can encode a DNA transgene of approximately 4.7kb (Wu et al. 2010a). They are favoured for use in the field of gene therapy on the basis of not being associated with any diseases in humans, dependence on another virus for replication and their low pathogenicity relative to Adenoviruses (Buller et al. 1981; Rose & Koczot 1972; Nayerossadat et al. 2012). Lentiviruses, this subclass of the retroviruses has a natural ability to integrate into non-dividing cells (Nayerossadat et al. 2012). The integration proficient and deficient variants, of this system in combination with their approximately 9kb transgene capacity make them an attractive viral vector for use in the DMD field (Meng et al. 2016). In contrast non-viral methods, with slightly lower efficiencies comprise chemical methods such as cationic liposomes and polymers and physical methods such as electroporation and magnetofection (Nayerossadat et al. 2012).

Importantly, with the exception of the gutted Adenoviral vectors, the 11.2kb dystrophin cDNA theoretically exceeds the packaging capacity of many viral vectors. In an effort to utilise the limited packaging capacity lentiviral and AAV systems mini and microdystrophin cDNA constructs were generated. These constructs encoded shortened but functional dystrophin proteins, with internal deletions across the central rod domain. The microdystrophin constructs, are diverse but encode the minimal necessary domains for dystrophin protein functionality. The main MD1 variant used throughout this investigation R4-24 Δ CT (nomenclature to indicate the deletion of spectrin repeats and a truncated C-terminal) was previously described by (Athanasopoulos et al. 2011; Le Guiner et al. 2017). Finally, a minidystrophin harbouring the well characterised BMD Δ 45-55 mutation was also used throughout this investigation (Taglia et al. 2015). Notably, during the production of this manuscript the insertion of full length dystrophin into a lentiviral template was achieved (Counsell et al. 2017)

Notably, the first clear demonstration of the clinical benefit afforded by gene therapy was achieved by retroviral vector treatment of severe combined immunodeficiency (SCID) (Hacein-Bey-Abina et al. 2002; Hacein-Bey-Abina et al. 2010). Since then there have been several successes in the field, including Glybera, an AAV vector engineered to express lipoprotein lipase, used in the treatment of lipoprotein lipase deficiency. Glybera was the first gene therapy to gain EU approval, although recently withdrawn on the basis of high associated cost and limited use of treatment (Ylä-Herttuala 2012; Regalado 2016). Strimvelis is another registered gene therapy product in which CD34+ cells transduced with retroviral vector to express human adenosine deaminase (ADA) are administered intravenously to treat ADA deficiency (Aiuti et al. 2017). These advances are encouraging, in addition other gene therapies using non-viral delivery systems such liposomes encoding the cystic fibrosis transmembrane conductance regulator (*CFTR*) also appear promising (Alton et al. 2016; Alton et al. 2015).

Therapeutic Gene-editing approaches are also being developed. Sangamo are currently undertaking a pilot study to assess the safety and efficacy of CR5 Modified CD34+ Hematopoietic Progenitor Cells in HIV patients (NCT02500849). This followed promising evidence obtained from a small study of 12 patients, that were given an infusion of Zinc Finger Nuclease (ZFN) modified autologous CD34+ T- cells. Interestingly T-cells, were significantly elevated per cubic millimetre of serum from 448 pre-infusion to 1517 one week later. Furthermore, T-cells post treated and were shown to persist longer and modulate viremia even during treatment interruption (Tebas et al. 2014). Secondly, *in vivo* administration

of ZFNs has very recently been announced in the treatment of Hunter's syndrome, the results of which will be eagerly awaited by the field (Grens, 2017).

1.6 Experimental Therapeutic Strategies for the Treatment of DMD:

1.6.1 Gene Augmentation Approaches:

Following the identification of the genetic causation of Duchenne muscular dystrophy 30 years ago, it was quickly identified as a candidate for gene therapy (Koenig et al. 1987). However, the large size of the dystrophin cDNA being approximately 11.2kb largely hindered the utilisation of this approach, exceeding the packaging capacity of many viral vectors (Hoffman et al. 1987). In response to this, minidystrophin and microdystrophin genes were designed; these encoding functional albeit internally-deleted dystrophin proteins, that can be coded for by minimally sized cDNA, suitable for packaging into lentiviral (LV) and adeno associated viral vectors (AAV) respectively (Harper et al. 2002). The basis of this approach was based upon two central premises: the modular nature of the dystrophin protein, permissive to the introduction of deletions, and the identification of a Becker patient, who appeared to be largely asymptomatic, despite 46% of the DMD gene being deleted, producing a highly truncated albeit functional dystrophin protein (Harper et al. 2002; England et al. 1990).

The earliest microdystrophin designs aimed to include the minimal necessary constituents for dystrophin function (Harper et al. 2002). This stated, they have continued to demonstrate promising findings in mouse and canine preclinical models; both in terms of dystrophin restoration and improved muscle strength (Foster et al. 2008; Bostick et al. 2008; Pichavant et al. 2010; Koo et al. 2011; Le Guiner et al. 2017). One phase I clinical trial was undertaken to examine the therapeutic effects of a particular microdystrophin construction (R2+H3-R22- Δ CT) described by Xiao Xiao, in 6 DMD patients (Mendell et al. 2010). Notably minimal and transient expression of the microdystrophin protein was observed, alongside immune responses to dystrophin in 4 of 6 patients (Mendell et al. 2010). This study was performed using a CMV promoter to control microdystrophin expression, and it is speculated that the use of a muscle restrictive promoter and reducing the immunogenicity of AAV capsids could all improve upon the outcomes observed (Mendell et al. 2012; Rodino-Klapac et al. 2013). It is important to realise that this microdystrophin construct is distinct from the format utilised in this study (R4-24 Δ CT). Recently, studies have focused upon the inclusion of more protein domains and altered architecture of both mini and microdystrophins to improve the functionality of resultant proteins (Reza et al. 2016; Hakim et al. 2017).

An alternative approach to overcoming the large transgene size is triple trans-splicing in which the dystrophin transgene is divided into three sequential fragments, packaging into AAVs and

co-transduced (Koo et al. 2014; Lai et al. 2005), the premise being that dystrophin will reconstitute upon the association of these three vectors. Although promising, the current systems show low efficiency and are still in the preclinical stages of development.

1.6.2 Utrophin Modulation Strategies

The utrophin gene is a homologue of the DMD gene; the resultant protein produced from this locus exhibits similar sequence, structural and binding properties to dystrophin (Love et al. 1993; Love et al. 1989). Early studies highlighted that utrophin was expressed at the sarcolemma of regenerating myofibres following muscle injury, and is upregulated in *mdx* mice during regeneration (Helliwell et al. 1992). Furthermore, the expression of full-length utrophin in transgenic dystrophin-deficient *mdx* mice was sufficient to prevent the development of muscular dystrophy in an expression dependent manner (Tinsley et al. 1998). A second related study, in a canine X-linked disease model highlighted that delivery of a mini-utrophin by an adenoviral vector was able to compensate for the absence of dystrophin. It was speculated that utrophin and mini-utrophin expression allowed for the restoration and correct formation of the DAPC, as indicated by higher levels of sarcolemmal β -dystroglycan and β -sarcoglycan (Cerletti et al. 2003). These observations suggested that utrophin protein could act as a functional surrogate for dystrophin, enabling the reconstitution of the structural link between intracellular F-actin and the DAPC. The reconstitution of such a link was speculated to improve structural integrity of the muscle and ameliorate dystrophic pathology, and led to investigations to identify drugs that would upregulate utrophin in skeletal muscle.

In light of the potential benefits of utrophin upregulation, an exhaustive screen was performed to identify compounds with the ability to selectively upregulate utrophin, and SMTC1100 was selected (Tinsley et al. 2011). Upon daily dosing of this compound in 4 week old *mdx* mice, an age when muscle characterised by necrosis and degeneration, a 2 fold increase in utrophin RNA was observed, and a subsequent increase in utrophin protein synthesis. Functionally, SMTC1100 treatment translated to a 75% reduction in serum CK levels, protection against forelimb exercise-induced weakness and improved resistance to fatigue in *mdx* mice. Furthermore an additive effect was observed when this compound was used in combination with the corticosteroid prednisolone (Tinsley et al. 2011).

These findings were sufficient to enable SMTC1100 to progress to a Phase I human clinical trial assessing the drug for safety and tolerability in healthy volunteers (Tinsley et al. 2015). This comprised of two main investigations, a single administration dose escalation, and a twice daily administration regimen. Promisingly, the drug was found to be safe and well tolerated, the only adverse side effect being stool discolouration attributed to unabsorbed drug passing

through the gastrointestinal tract. Furthermore, the drug was rapidly absorbed and produced serum levels comparable to those required to produce a 50% upregulation of utrophin *in vitro*. This stated, a high level of variation in serum concentration of the drug was observed despite similar levels of absorption. Investigators posited that the patient's diet maybe an influential factor in drug persistence and clearance, recommending the drug should be administered with food, preferably with a high fat content in future studies.

Similar trends persisted in a later clinical trial performed in paediatric patients with DMD where despite patients being instructed to eat 10 minutes prior to SMTC1100 administration, a large variability in drug serum concentration was still observed. Intriguingly, the SMTC1100 plasma concentration was more rapidly reduced over 11 days of treatment in DMD patients relative to healthy volunteers falling 56-62% relative to the 23-40% previously reported. However a number of biochemical markers indicative of muscle damage declined over the course of the study, which was interpreted as a positive outcome (Ricotti et al. 2016). These findings have subsequently prompted further on-going clinical trials. Firstly, the examination of SMTC1100 in paediatric patients with DMD following a balanced diet (Clinical Trial Number: NCT02383511). Interim reports indicate this trial has been completed with a new formulation of SMTC1100 that can attain the same serum concentration as the previous formulation at 1/10th of the dose, but formal results are currently awaited (Summit Therapeutics Plc 2016; Guiraud, Squire, et al. 2015). In the meantime, this trial has been extended to a 48 week Phase 2 clinical trial (Clinical Trial number: NCT02858362).

1.6.3 Nonsense Mutation Read-Through Therapies:

It is estimated that approximately 10% of DMD cases are caused by the presence of a nonsense mutation (Bladen et al. 2015; Welch et al. 2007). This causes the premature termination of translation, and the resultant mRNA is subject to degradation through the nonsense mediated decay pathway. This prompted the examination of compounds that could enable selective read-through of nonsense codons (Welch et al. 2007).

Early observations indicated aminoglycoside-based antibiotics could promote read-through of premature stop codons, but concerns related to drug toxicity hindered further therapeutic development (Malik et al. 2010). However these findings prompted high-through screening of 800,000 novel drug compounds to be undertaken to identify high efficacy read-through compounds with lessened toxicity (Welch et al. 2007). The results of this screen identified the compound PTC124 (Trade names Ataluren /Translana). Notably, this compound shows no structural similarity to aminoglycosides but is able to elicit selective read-through of premature termination codons at relatively low doses (Welch et al. 2007).

Importantly, the initial screening highlighted that Ataluren was able to promote dystrophin production in primary myoblast cultures from DMD patients and *mdx* mice (Welch et al. 2007). Furthermore a combined regime of oral administration and intraperitoneal injections in *mdx*, showed a reduction in CK levels, a restoration of dystrophin to 25% wild-type levels and improved EDL muscle resilience to eccentric contraction (Welch et al. 2007). These preclinical findings were enough to support two subsequent Phase I trials in 62 mixed gender healthy volunteers (Hirawat et al. 2007). Importantly, single dosing regimens of 100, 150 and 200mg/kg and repeated doses of 50mg/kg 12 hours apart were safe and well tolerated. The findings that the drug specifically targeted premature stop codons and had no evidence of accumulation or toxicity prompted phase II clinical trials in DMD patients (Hirawat et al. 2007), and eventually suggested potential therapeutic benefit of Ataluren (Finkel et al. 2013; Bushby et al. 2014).

In the first phase IIa proof of concept trial, all patient myoblasts derived from biopsies, and cultured in the presence of Ataluren, showed evidence of dystrophin restoration. Furthermore, on administration of Ataluren, 23 of 38 patients (61%) showed evidence of dystrophin restoration in pre and post-treatment comparisons. The mean dystrophin restoration was estimated to be 11% and there was a concomitant decline in serum CK levels. Over this 28 day trial, the treatment was well tolerated (Finkel et al. 2013). A longer 48 week Phase IIb randomised, double blind trial, mirrored the beneficial findings of the former with the use of functional endpoints. Promisingly, patients with a mean baseline 6 minute walk test distance (6MWD) of 350m \pm 97.6m enrolled in a 40mg/kg/day dose regime of Ataluren showed a mean improvement of 31.3m relative to the placebo group. Although not statistically significant, this prolonged ambulation and improved muscle function was deemed meaningful and advocated Ataluren's progression to a Phase III clinical trial (Bushby et al. 2014).

Although the mean 31.3m improvement in the walk test in the Phase IIb trial was not statistically significant; two independent studies of the 6MWT, a primary end point assessing locomotion, deemed the finding as meaningful (McDonald et al. 2010; McDonald et al. 2013). This was instrumental to Ataluren becoming the first drug to gain marketing and approval in Europe (Haas et al. 2015). The Phase III clinical trial continues to develop upon this trend of success, on the division of groups on the basis of the 6MW test (<300m, 300-400m and 400m+). Notably, significant improvement was observed in the 300-400m 6MW Ataluren-treated group relative to the placebo, with a mean 42m difference in walking distance. In the other groups a slower more stabilised rate of decline was observed. The cumulative efforts of these studies demonstrate that Ataluren is promising treatment for DMD patients with a

nonsense mutation, providing conservative benefit or stabilisation of disease progression. As a consequence an additional long term trial is planned (NCT03179631).

1.6.4 Exon Skipping by Antisense Oligonucleotides

The rationale of antisense oligonucleotide (AON) exon-skipping is to convert an out-of-frame DMD mutation to an in-frame BMD mutation at the transcriptomic level. Importantly, exons adjacent to deletion mutations in DMD are generally not in-phase with each other, meaning when they are transcribed, the mRNA will not form complete wild-type codons. The resultant out-of-frame mRNA sequence will quickly generate a downstream stop codon, and be subject to nonsense-mediated decay. AONs are short synthetic nucleic acid stretches, with modified backbone chemistries that bind to and sterically hinder the splicing of a target exon within the dystrophin pre-mRNA (Aartsma-Rus et al. 2017). The net result is the exclusion of a target exon, the splicing of in-phase exons and restoration of the downstream open reading frame. For example, deletion of DMD exons 48-50, results in an out-of-frame mutation, as exon splicing being 47-51 are not in-phase. Targeting exon 51 with an exon skipping AON, to promote its exclusion from the transcript would result in exon 47-52 splicing. These exons are in-phase and would result in the restoration of the dystrophin open reading frame, enabling translation of a functional dystrophin protein harbouring an internal deletion (Jarmin et al. 2014; Verhaart & Aartsma-Rus 2012; Aartsma-Rus et al. 2017).

Due to it having the highest therapeutic applicability with 13-15% of DMD patients being eligible for treatment, AON therapy targeted to exon 51 is the most advanced in development. Two distinct AON chemistries, have been developed to target exon 51 (Aartsma-Rus et al. 2009). The first product was a 2' O-methyl phosphorothiorate (2'OMePS) AON, Drisapersen (Goemans et al. 2011; van Deutekom et al. 2007; Voit et al. 2014), and the second product a phosphorodiamidate morpholino oligomer (PMO) Eteplirsen (Mendell et al. 2013). [For a review of a range of AON chemistries see (Chan et al. 2006)]. These AONs were subject to a plethora of clinical trials and recurrent trends observed included variable dystrophin restoration observed across patients enrolled and marginal but largely non-significant improvements in the 6 minute walk test in patients (Cirak et al. 2011; Cirak et al. 2012; Goemans et al. 2011; Goemans et al. 2016; Kinali et al. 2009; Mendell et al. 2013; Mendell et al. 2016; van Deutekom et al. 2007; Voit et al. 2014; Mendell et al. 2017). In the case of Drisapersen, a recent long term open-label extension study indicated that the treatment was well tolerated with the most common adverse events reported being injection site reactions. Finally, this study also indicated that for a small cohort of patients this treatment appeared particularly promising as natural history comparisons indicated a 100m decline in comparable untreated patients (Goemans et al. 2016). Recently and rather controversially Eteplirsen

gained accelerated FDA approval, pending further assessments due in 2021 (Aartsma-Rus & Krieg 2017). The approval proved controversial on the basis that despite dystrophin positive myofibres appearing elevated from 9.4 – 22.6% post treatment; the overall increases in dystrophin protein determined by western blot showed a marginal mean increase from 0.28-0.93%, which was not observed in all patients (Aartsma-Rus & Krieg 2017). This finding suggests that whilst myofibres may be dystrophin positive they express dystrophin at very low amounts. In combination these studies emphasise that more research is warranted to assess whether such small changes in dystrophin protein could serve to stabilise disease progression.

Eteplirsen gaining accelerated FDA approval and being under evaluation by the EMA, represents an exciting advancement for Exon51 skipping, however a number of hurdles still remain for this therapeutic strategy. Whilst, AONs have been developed for all possible therapeutic exon skips in the DMD gene, distinct AON sequences and chemistries constitute separate medicinal compounds under current EU legislature (Aartsma-rus et al. 2005; Wilton et al. 2007; Aartsma-Rus et al. 2017). As a consequence each compound will require individual clinical testing as exemplified by AONs designed to Exons 45 and 53 (Clinical Trial: NCT02310906) (Aartsma-Rus et al. 2017). This format of clinical development could prove challenging to the field, particularly in the context of AONs with low patient applicability of less than 1%. Furthermore, not all patients would be eligible for this form of treatment with 64% of patients potentially eligible for single exon skip therapy, 79% for double exon skipping, and 90% for multi-exon skipping (Verhaart & Aartsma-Rus 2012). The screening of all reagents for multi-exon skipping would likely prove very labour intensive. Moreover, multiple exon skips could theoretically be hindered by variable efficacy of AONs administered. Finally, whilst this strategy could ameliorate some of the morbidity associated with DMD, attaining exon skipping in the heart still remains challenging for which tricyclo AONs and peptide-lined PMOs are currently being investigated (Relizani et al. 2017; Betts et al. 2012). Promisingly, *in vivo* delivery of tricyclo AONs targeted to the 5' splice site of intron 23 achieved exon 23 skipping and restoration of dystrophin to 38% wild type levels in the *mdx* heart. This stated, the doses of tricyclo AONs used (200mg/kg) are considerably higher than the therapeutic ranges used for Drisapersen(6mg/kg) or Eteplirsen (50mg/kg) (Relizani et al. 2017; Goemans et al. 2016; Mendell et al. 2013).

1.6.5 Myostatin Inhibition Studies:

Myostatin, also referred to as growth and differentiation factor 8 (GDF8), is a member of the TGF- β superfamily, but is considered unique due to its muscle specific expression (Schuelke et al. 2004). Physiologically, GDF8 is transcribed as an immature precursor protein, this undergoes proteolytic cleavage to release a 24kDa homodimer, which remains associated with

its pro-peptide in a non-covalent manner (Latres et al. 2015). This homodimeric form of myostatin persists in serum, complexed to proteins such as follistatin and growth and differentiation factor associated serum protein (GASP-1) that maintain it in an inactive form (Latres et al. 2015). Finally, a BMP1/tolloid serum metalloprotease disrupts bonding between the pro-peptide and mature peptide, and liberates mature active GDF8 (Lee et al. 2010).

The mature GDF8 is able to interact with activin receptor IIB (ActRIIB) for which it exhibits high affinity, and leads to Smad 2/3 becoming phosphorylated, entering the nucleus and ultimately modulating downstream gene expression (McPherron et al. 1997; Lee 2004). The net result of this process is the negative regulation of muscle mass. As muscle wasting is prominent in a range of muscular dystrophies and atrophic neuromuscular diseases and conditions, inactivation or inhibition of myostatin or subsequent myostatin signalling was identified as a potential therapeutic target (Thomas et al. 2000). Evidence in support of this premise was afforded by the hypertrophy present in a range of naturally-occurring and transgenic animal models where loss-of-function mutations are present in the myostatin gene, including one human precedent of myostatin knock-out (McPherron & Lee 1997; Chen & Lee 2016; Schuelke et al. 2004). Furthermore, it was demonstrated that *mdx* mice harbouring a myostatin knock-out not only showed hypertrophy but an increase in strength (Wagner et al. 2002).

As a consequence of this observation many strategies have been undertaken in exploration of myostatin inhibition in DMD. In preclinical models, AONs have been used to induce out-of-frame skipping in myostatin mRNA. Studies combining AONs to restore the dystrophin reading frame and induce out-of-frame skipping in myostatin, demonstrated that a reduction of myostatin had an additive effect on dystrophin restoration (Lu-Nguyen et al. 2015). In light of Eteplirsen gaining FDA approval, this is one strategy that could be used to improve the efficacy of AON therapies.

An alternative strategy for myostatin inhibition is use of therapeutic myostatin antibodies, the premise being these would bind to myostatin protein preventing its downstream signalling. The first myostatin antibody tested, Myo29, demonstrated suitable safety margins for intravenous delivery to patients with muscular dystrophies. However, the antibody failed to meet clinical endpoints, based upon both physical and biological assessments leading to its discontinuation (Wagner et al. 2008). However, post-hoc analyses indicated that the antibody dose and serum concentration in the study may have been insufficient to elicit an effect, due to blood clearance being higher in patients than first anticipated (Singh et al. 2016). Nonetheless two further antibody constructs are currently in phase II trials for DMD these

being PF-06252616, (NCT02310763) and BMS-986089, (NCT02515669) respectively. Promisingly, both showed high affinity to myostatin and evidence of muscle hypertrophy in healthy volunteers in phase I trials results of Phase II in DMD patients are anticipated in 2020 (Bhattacharya et al. 2017; Spinazzola & Kunkel 2016). Finally, continuing with the trend, myostatin soluble receptor moieties have been tested in preclinical models and in clinical trial. The trial of one formulation, ACE-031, was halted due to bleeding and dilation of blood vessels, but a new formulation, ACE-081, is now in clinical trial for facioscapulohumeral muscular dystrophy (FSHD: NCT02927080), and a phase 1/2a follistatin gene therapy has been recently completed (Campbell et al. 2017; Mendell et al. 2015) Both studies aim to prevent myostatin binding to its endogenous receptor and eliciting negative regulation of muscle mass.

1.7 Gene Editing Background and the DNA Damage Response

The mammalian genome is subject to continual assault from both endogenous sources such damage caused by DNA metabolism and reactive oxygen species (Klungland et al. 1999; Kirkwood 2005), and environmental sources such as exposure to DNA-damaging chemicals or radiation (Harman 1956). Strikingly, it is estimated that a single cell could experience up to 10^5 spontaneous DNA lesions per day (Hoeijmakers 2009). In the context of DNA as a genetic blueprint, the maintenance of eukaryotic genome integrity is vital, ensuring the correct transmission of genetic material to progeny and the prevention of disease causing mutations (Ciccina & Elledge 2010).

In response to this threat eukaryotes have evolved the DNA damage response (DDR) system, a multifaceted signal transduction pathway to identify and resolve DNA lesions, safeguarding the genome against deleterious mutations (Ciccina & Elledge 2010). Importantly, the DDR comprises many distinct DNA repair pathways, the activation of which is under stringent regulation and is influenced by the nature of the DNA lesion, whether it is a single stranded break (SSB) or a double stranded break (DSB) (Heyer et al. 2010; Bothmer et al. 2017; Ciccina & Elledge 2010). This DNA lesion specificity is perhaps anticipated on the basis that erroneous repair could cause fragmentation, non-disjunction or deleterious mutations (De Jager et al. 2001). Importantly, gene-editing or genome editing, aims to exploit these natural mechanisms of DNA repair, to effect a change in the genome that corrects the disease causing mutation.

Four main pathways that are routinely utilised in the field of gene editing including:

1. Canonical Non-Homologous End Joining (NHEJ),
2. Alternative NHEJ (alt-NHEJ) also referred to as Microhomology End Joining (MMEJ),
3. Single Strand Annealing (SSA)
4. Homology Directed Repair (HDR).

The current studies in this thesis will largely focus on canonical NHEJ in the assessment of gene modification at a defined genomic locus, and HDR in attempts to facilitate the integration of dystrophin transgene cDNAs within intron 1 of the DMD gene, with the aim of restoring spatial and temporal expression of dystrophin in DMD cells. Thus these two DNA repair pathways will be discussed in great detail.

1.7.1 The Canonical Non-Homologous End Joining Pathway (NHEJ)

The NHEJ pathway is the main mechanism of DSB repair in proliferating mammalian cells (Mao et al. 2008; Velic et al. 2015). This is because, relative to HDR which is constrained to late G2 and S-phases of the cell cycle, when a sister chromatid is present to function as a repair template, the NHEJ pathway occurs throughout the cell cycle (Mao et al. 2008; Heyer et al. 2010). Notably HDR and NHEJ can be thought of as working antagonistically, as during the activation of HDR in G2 and S phases, NHEJ is transiently suppressed (Heyer et al. 2010; Velic et al. 2015).

At the simplest mechanistic level, the NHEJ pathway can be considered as the binding of terminal DNA ends liberated by the DSB and their direct ligation (Hoeijmakers 2009) (Figure 1-2). This process is achieved by four core protein mediators: Ku70/ Ku80 heterodimeric complex; DNA dependent kinases (DNA-PKcs); Ligase IV (LIG4); and X-Ray Cross-Complementing protein 4 (XRCC4) (Ciccia & Elledge 2010; Decottignies 2013; Weterings & Chen 2008).

The Ku70/Ku80 heterodimeric complex is a circular protein with a central pore, which enables its loading onto exposed DNA termini. Once bound this binding of Ku70/Ku80 serves to sterically hinder and prevent the binding of alternative sensors of such as PARP1 which would promote DNA end processing. Furthermore, it enables loading of DNA-Pkcs, DNA dependent phosphokinases, which act to further stabilise DNA ends and prevent processing by auto-phosphorylation. Ultimately this serves to recruit LIG4 and XRCC4, which enable end ligation with XLF4, which functions as a stimulatory factor. If the DNA termini are not compatible for direct ligation, then Artemis, a substrate for DNA-PKcs, becomes phosphorylated. Once phosphorylated, Artemis functions as an exo- and endo-nuclease, promoting DNA-end processing until the DNA ends are suitable for ligation (Velic et al. 2015; Ciccia & Elledge 2010).

Inherent in this process is the potential for subtle sequence alterations at the junction of the DSB, these are known as micro-insertions or -deletions (InDels). Importantly, following delivery of customisable endonucleases, InDels are used frequently in gene- editing studies as a proxy for the presence of DSBs (Qiu et al. 2004; Liu et al. 2016a). This stated, InDels can serve to

underestimate the extent of DNA cleavage occurring as the DNA can be directly ligated without the requirement for further end processing and without the emergence of InDels.

1.7.2 The Homology Directed Repair Pathway:

In contrast to the NHEJ DNA repair pathway, which is considered to repair DSB with relatively low fidelity, the HDR pathway enables precise correction. Crucially, in this form of repair the DSB termini are processed to form 3' ssDNA overhangs that can invade isogenic sequences promoting the repair of DNA from a genetic template (Velic et al. 2015; Decottignies 2013; Jasin & Rothstein 2013). See (Figure 1-2)

Importantly, the process of HDR is tightly regulated by phosphorylation signalling cascades involving many protein mediators. Initially, DNA damage is sensed by the MRN protein complex, which comprises three protein components, RAD50, MRE11, and Nibrin (NBS1). The Rad50 protein, a member of structural maintenance of chromosomes (SMC) family, is implicated in tethering DNA ends resulting from DSBs, in close proximity (De Jager et al. 2001; Velic et al. 2015). In addition, the Rad50-associated ATPase and coiled coil domains enable interaction with MRE11, a protein with a range of functions including DNA endonuclease, exonuclease and unwinding activities (Assenmacher & Hopfner 2004; Williams et al. 2007). This complex serves to stabilise the free DNA ends, initiate end processing and association with the final component of the MRN complex, the NBS1 protein. NBS1 recruits the ataxia-telangiectasia mutated (ATM) protein to the DSB, interacting with ATM via its C-terminal domain.

The ATM protein becomes activated by the MRN complex and further signal transduction occurs via phosphorylation of CtBP-interacting protein (CtIP), and the breast and ovarian cancer type 1 susceptibility protein (BRCA1) (Paull 2015). These protein mediators serve to promote DNA end resection, and the generation of 3' ssDNA overhangs. Notably, if this form of repair occurs in the G1 phase of cell cycle, then DNA end processing is constrained by the absence of the BRCA1 protein. In this instance, limited resection and 3' ssDNA overhangs are generated by the action of CtIP alone, which forms the basis of the alternative NHEJ DNA repair / MMEJ repair (Ciccio & Elledge 2010; Nakamae et al. 2017). In contrast at the G2 phase of the cell cycle, BRCA1 protein is present and interacts with the CtIP protein, inactivating it after initial DNA resection has occurred (Huen et al. 2010). This primes the DNA end for more extensive resection, as mediated by Exo1 and BLM following ATM phosphorylation (Ciccio & Elledge 2010).

The exposed 3' ssDNA overhang becomes coated in and stabilised by replication protein A (RPA), this enables the binding of RAD51 or RAD52. If the 3' ssDNA overhangs harbour

repetitive sequences, the break could be resolved by the 3' overhangs annealing, overhangs being resected and ligated by a RAD52 and an enzymatic complex of XPF/ ERCC1. This process is called single strand annealing (Ciccio & Elledge 2010).

In contrast, the full HDR pathway is characterised by a 3' ssDNA strand invasion of an exogenous isogenic sequence, a process initiated by a Rad51 heterogenic filamentous protein complex binding and stabilising of the overhang (Jasin & Rothstein 2013). Interestingly, many protein mediators are also implicated in this process, the two main ones being casein kinase 1 (CK1) and RRCA2. CK1 serves to phosphorylate Rad51 and aid in its recruitment to sites of DNA damage, where it subsequently displaces Rad51 (San Filippo et al. 2008). BRCA2 directly associates with Rad51, and one theory indicates this could serve to modulate this DNA repair pathway, with the temporal BRCA2 expression constraining it to late G2, S-phases of the cell cycle (Moynahan et al. 2001). The result of these interactions is the formation of an intermediate D loop- structure (West 2003). Importantly if the second DNA end remains uncaptured, the invading 3 ssDNA can be extended by DNA polymerases, dissolution of the D-loop structure occurs as mediated by regulation telomere elongation helicase 1 (RTEL1) and the newly synthesised strand anneals to the second DNA end, a repair process referred to as synthesis dependent strand annealing (SDSA) (Ciccio & Elledge 2010)

For completion of conventional HDR, the 3' ssDNA needs to invade the exogenous repair template, the second strand needs to be captured and form a double holliday junction (dHJ) (Jasin & Rothstein 2013). Importantly, this genomic structure can also be resolved in numerous ways including dHj dissolution and genomic cross-over or non-crossover events. These processes are mediated by further protein complexes BLM/TOPOIII, GEN1, MUS81/ EME1 or SLX1/SLX4 (Ciccio et al. 2008; Fekairi et al. 2009; Ip et al. 2008) which promote the remodelling and/or subsequent cleavage of DNA ends in the pursuit of DNA repair. Notably, a cross-over event refers to the exchange of markers of genomic material flanked by the dHJ and in the context of gene therapy can be used to confer precise disease ameliorating DNA modifications (Musunuru, 2017).

Importantly, this presentation of the HDR mechanism showcases the complexity of this DNA repair pathway. The multiple exit points preceding the desired genomic cross-over and the dependence on a plethora of mediatory proteins are salient considerations in the inherent low efficiency of HDR. Furthermore, an awareness of mediatory proteins provides potential drug targets to modulate the activity of DNA repair pathways and improve HDR efficiency.

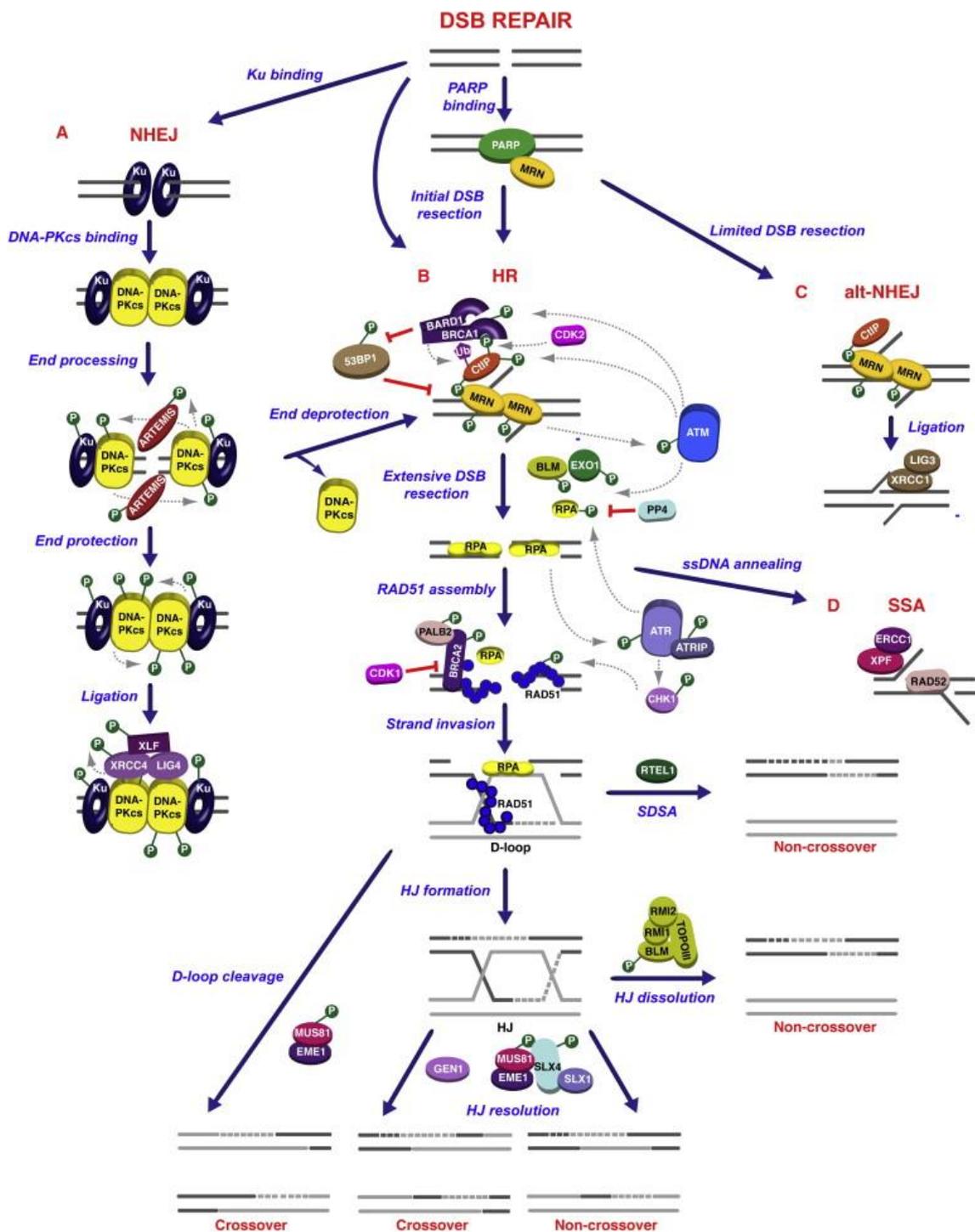


Figure 1-2 A schematic diagram showing the various DNA repair pathways, employed in the resolution of a DSB: Taken from (Ciccia & Elledge 2010).

A) The NHEJ pathway, the Ku70/ Ku80 heterodimeric complex quickly associated with DNA termini, preventing end processing. The association with DNA Pkcs causes a phosphorylation cascade that resolves DNA breaks by direct ligation with ligase IV or by subsequent processing with Artemis and ligase IV. B) The HDR pathway (sometimes referred to as Homologous Recombination or HR pathway), PARP out competes Ku70/ Ku80 binding allowing the MRN complex: MRN11, RAD50, and NBS1 to be recruited. This in turn forms MRN/CtIP/BRCA1 complex notably, the constituents of this complex are only available at late G2/ S phases of the cell cycle and initiate end resection. More extensive end resection is achieved by Exo1 and BLM, to achieve a 3' ssDNA overhang that becomes coated in RPA. Eventually this is displaced by RAD51, a filamentous protein that mediates strand invasion as regulated by BRCA2. D-loop

structures are formed following strand invasion, these can be resolved by cleavage mediated by MUS81/EME1 or subject to dissolution with the RTEL1, this form of repair SDSA can result in cross-over or non-crossover events. If the second strand is captured, then a double holliday junction is formed; this structure in turn can be subject to dissolution by BLM/TOPOIII complex or subject to crossover or non-crossover events with by the nucleases GEN1 and SLX1/SLX4, which associate with MUS81/EME1. C) The alternative NHEJ pathway occurs as a result of a CtIP partial end resection process independent of BRCA1, this generates short overhangs. D) The single stranded annealing pathway, this occurs after partial end resection when the 3'ssDNA can be coated in Rad52 and the DNA ends directly ligated together.

1.8 Gene Editing Background: Customisable Nucleases and the Induction of a DSB:

A major milestone to the field of gene editing was the observation by Maria Jasin et al in 1998 was that the HDR DNA repair pathway could be stimulated more than 100-fold following the induction of a DSB at the target locus (Liang et al. 1998). Crucially, for the exploitation of this finding in the correction of disease causing mutations, endonucleases were required that could be engineering to introduce precise DSBs at user-defined genomic locations. Multiple variants of these so-called 'customisable endonucleases' have since been engineered or discovered including meganucleases (MGN), zinc finger nucleases (ZFNs), transcriptional activator like effector nucleases (TALENs) and Clustered Regularly Interspaced Palindromic Repeat/Cas9 nucleases (CRISPR/Cas9). The advent of these nucleases have served to advance reverse genetics through the generation of mouse models with conditional or permanent gene knock-out (Collins et al. 2007). In a broader sense they have also been explored in the correction of disease causing mutations (Prakash et al. 2016).

1.8.1 Meganucleases

Meganucleases (MGN) represent the first generation of customisable endonucleases, derived from homing endonucleases, with large DNA recognition sites spanning 12-45bps (Epinat et al. 2003; Thierry & Dujon 1992) (Figure 1-3). Initially the 5 families of homing endonucleases, LAGLIDADG, GIY-YIG, HNH, His-Cys box and PD-(D/E)XK (Orlowski et al. 2007; Zhao et al. 2007), found to be naturally occurring in phages, bacteria, archea and some eukaryotes, were examined for their application to gene targeting. However this was hindered by the low repertoire of 300 naturally occurring proteins and the reliance on homing endonuclease sites being present within the mammalian genome. As a consequence focus shifted to engineering the protein structure of the homing endonuclease (Epinat et al. 2003).

From three dimensional structural modelling of LAGLIDADG family proteins it was identified that two alpha-helices confer the main propensity to cleave DNA. Either side of these helices are four stranded beta sheets which provide an additional DNA binding interface (Epinat et al. 2003). Interestingly however, it was noted that both domains in part contribute to DNA binding and cleavage activity of the MGN (Moure et al. 2003). On this basis, many strategies were proposed for expanding the DNA targeting repertoire of MGNs, including computational structure-based designs, and domain swap engineering (Chevalier et al. 2002; Ashworth et al. 2010). Notably, the design of MGNs to target distinct genomic locations proved challenging, often resulting in diminished efficacy which was attributed to the non-modular nature of the protein, with binding and catalytic activity not appearing functionally distinct (Stoddard 2014;

Guha et al. 2017). As a consequence of these limitations further gene-editing platforms were developed.

1.8.2 Zinc Finger Nucleases

The first generation of artificial nucleases using this platform fuses the CH2 Zinc Finger DNA binding modules in tandem to a type IIS nonspecific DNA cleavage domain of FokI. Together these domains form a ZFN monomer (Kim et al. 1996; Tupler et al. 2001; Li et al. 1992) (Figure 1-3). Individual zinc finger protein modules identify 3 nucleotides of genomic sequence, and 3-6 of these modules are combined sequentially within each monomer (Kim et al. 1996; Pavletich & Pabo 1991). Thus each monomer is involved in binding 9 -18 nucleotides of genomic DNA. Importantly the FokI nuclease, originally isolated from *Flavobacterium okeanoicoites*, = is tethered to the termini of the ZF monomer and require obligatory dimerization in order to cleave DNA; the net result being staggered DSB in DNA (Li et al. 1992; Bitinaite et al. 1998). As a consequence monomers are situated either side of a genomic target site, designated left and right ZFNs, with a 5-7 nucleotide spacer region between them (Kim & Kim 2014a). The obligatory dimerization, with two arms and a spacer significantly improve the specificity of this gene editing system (Kim & Kim 2014a).

This ZFN platform has been used broadly by Sangamo Biosciences and shows promise in the development of gene editing therapies for HIV and Hunter's syndrome (Tebas et al. 2014; Grens, 2017). This stated, construction of Zinc Finger nucleases continues to be challenging with those generated from zinc finger module libraries showing evidence of low efficacy and cytotoxicity due to off-target effects (Ramirez et al. 2008; Cornu et al. 2008). Whilst production processes have improved to try and select for high functioning zinc finger nucleases and reduce off-target effects, the production of ZFN is costly and has limited the breadth and scale of their use (Miller et al. 2007; Szczepek et al. 2007; Sander et al. 2011; Gupta et al. 2012; Bhakta et al. 2013).

1.8.3 Transcriptional Affector Like Effector Nucleases (TALENs)

This gene editing platform fuses transcription activator-like effectors (TALEs), a family of proteins from pathogenic plant bacteria to the type-IIS FokI nuclease sub-units to form the TALENs. in this manner they are structurally akin to ZFNs (Miller et al. 2011; Zhang et al. 2011). (Figure 1-3) In nature, TALE proteins are employed by pathogenic bacterium such as *Xanthomonas oryzae* during the infection process to subvert plant host transcription machinery to produce a cellular environment favourable to survival (Boch 2011; Kay et al. 2007; Sugio et al. 2007). Structurally, TALEs comprise an N-terminal domain , a central repeat region and a C-terminal domain encoding a nuclear localisation signal (NLS) and an acidic

activation domain (Mussolino et al. 2011; Kay et al. 2007; Römer et al. 2007). Distinct from ZFNs, TALENs are composed of 17.5 tandem repeats/monomers each of 35 amino acids, with one monomer recognising a single nucleotide in the major groove of DNA (Mak et al. 2012; Deng et al. 2012). Furthermore from experimental studies, it was found that deletion of amino acids at positions 11 -14 ablated the activity of a TALE monomer (Römer et al. 2007). In this manner, it was identified that amino acids at R12 and 13 within the central repeat region determined the specificity of each domain; these repeats were then identified as Repeat Variable Di-residues (RVDs) (Boch et al. 2009; Moscou & Bogdanove 2009). The four main RVDs, ranked in their experimentally determined hierarchy of binding are Asn-Glycine (NG), His-Asp (HD), Asn-Asn (NN) and Asn-Ile (NI), which recognise thymine, cytosine, guanine and adenine respectively (Meckler et al. 2013; Streubel et al. 2012).

This 1:1 ratio of RVD to nucleotide makes computational design of TALENs relatively straightforward (Kim & Kim 2014a). This stated, the protein engineering required to construct a TALEN remains a challenging and time consuming process, despite many construction methods including golden gate cloning, solid phase assembly and ligation independent cloning (Cermak et al. 2011; Reyon et al. 2012; Schmid-Burgk et al. 2012; Briggs et al. 2012). Crucially, this gene editing platform has extensive and diverse targeting capabilities, with the only pre-requisite of conventional targeting being a thymine at the 5' end of the target sequence (Kim & Kim 2014a). This only stands to improve with the production of TALENs that can dispense with this criterion able to bind with alternative bases at the 5' end (Lamb et al. 2013). Similar to ZFNs, the main limitation of this system is challenging and relatively costly construction and screening required.

1.8.4 CRISPR/Cas9

The newest advent to the gene-editing repertoire is CRISPR/Cas9, this utilises two components a RNA guide to direct DNA binding and the Cas9 nuclease to cleave a genomic site (Jinek et al. 2012) (Figure 1-3). The novel RNA to DNA binding interface underpinning this platform is distinctive from its predecessors, making this system both simple to construct and versatile in application (Kim & Kim 2014b). First discovered in bacterial and archaeal genomes almost 30 years ago, it has since been characterised to confer adaptive immunity to prokaryotes from foreign nucleic acids (Ishino et al. 1987; Mojica et al. 2000). In the endogenous system, the initial exposure of bacteria to phage DNA is resolved by non-specific cleavage of the phage genome. Subsequently small fragments 20 nucleotides in length, referred to as protospacers, are integrated into the bacterial genome between adjacent flanking repeats. Upon re-exposure to the same phage, after this 'immunisation' process, the 20 nucleotide sequence is transcribed and processed to form CRISPR RNA (crRNA). This crRNA then associates with trans-

activating CRISPR RNA (tracrRNA), forming a complete guide that can recruit the Cas9 nuclease. The crRNA confers DNA binding specificity to the nuclease, introducing targeted DSBs into the phage genome, neutralising the immunological threat posed to the bacteria (Horvath & Barrangou 2010; Barrangou et al. 2007; Makarova et al. 2006).

Notably, examination of this system across prokaryotes classified CRISPR into three main classes, Type I, II and III. Interestingly type I and III use multiprotein complexes, while type II utilises a single protein Cas9 (Makarova et al. 2011). On the basis of these observations, focus was placed upon re-purposing the type II *Streptococcus Pyrogenes* Cas9 system for the purpose of genomic engineering in mammalian cells. In pursuit of this, co-expression of crRNA and tracrRNA was attempted, alongside a humanised Cas9 protein (Jinek et al. 2013; Mali, Esvelt, et al. 2013; Cong et al. 2013). The low efficiency of this three component system attributed to limitations with delivery, promoted a single chimeric guide fusing crRNA and tracrRNA via a hairpin loop to be engineered downstream of a U6 RNA III polymerase promoter (Jinek et al. 2013). Whilst this structure has been subsequently revised (Chen et al. 2013; Dang et al. 2015), it served to simplify the constituents of guide delivery, with the 20 nucleotide crRNA region being interchangeable. Oligonucleotides could be designed, self-annealed and cloned into U6 expression cassettes to confer new DNA binding specificity to the Cas9 nuclease (Ran et al. 2013).

On association of crRNA and tracrRNA or the newly developed format of a chimeric sgRNA, the Cas9 protein undergoes conformational change providing a central pore for the loading of the Cas9 protein onto DNA at the defined genomic locus (Nishimasu et al. 2014; Jinek et al. 2014). The summation of this process is the correct positioning of two catalytic domains within the Cas9 protein, the HNH and RuvC, which cleave complementary and non-complementary strands of the DNA respectively (Nishimasu et al. 2014). This process results in a blunt cut 3 nucleotides upstream of the Protospacer Adjacent Motif (PAM), a recognition motif that binds to the C-terminal of the Cas9 facilitating protein-DNA association and subsequent DNA cleavage (Anders et al. 2014; Nishimasu et al. 2014; Hsu et al. 2013b; Mali et al. 2013). Notably, the PAM for *Streptococcus Pyrogenes* is 5' NGG or 3' NCC, where N denotes any nucleic acid base, but alternative PAM recognition sites are utilised by bacterial orthologues of Cas9 (Ran et al. 2015; Zetsche et al. 2015; Hou et al. 2013). Importantly, the propensity for the Cas9 system to cleave a genome site is governed by the presence of PAM throughout the genome. In addition to this PAM feature, RNA guides must initiate with a 5' guanine whether this is native sequence or appended nucleotide due to the nature of the U6 promoter (Ma et al. 2014; Gagnon et al. 2014). Since the initial demonstration of its use, the Cas9 protein been engineered to include a range of alterations, with the outlook to improve its specificity or

broaden its function. These include, inactivation of a single catalytic domain to produce nickases that introduce a single stranded break (SSB) or DNA lesion, which when paired in a dimeric fashion can cause a DSB (Ran et al. 2013). Also there has been the modification of internal amino acids responsible for DNA binding to produce enhanced SpCas9, a nuclease with greater specificity and reduced off-target effects (Slaymaker et al. 2016). Advancements in the CRISPR/Cas9 field remains dynamic and on-going, and a detailed current review is provided by (Tycko et al. 2016).

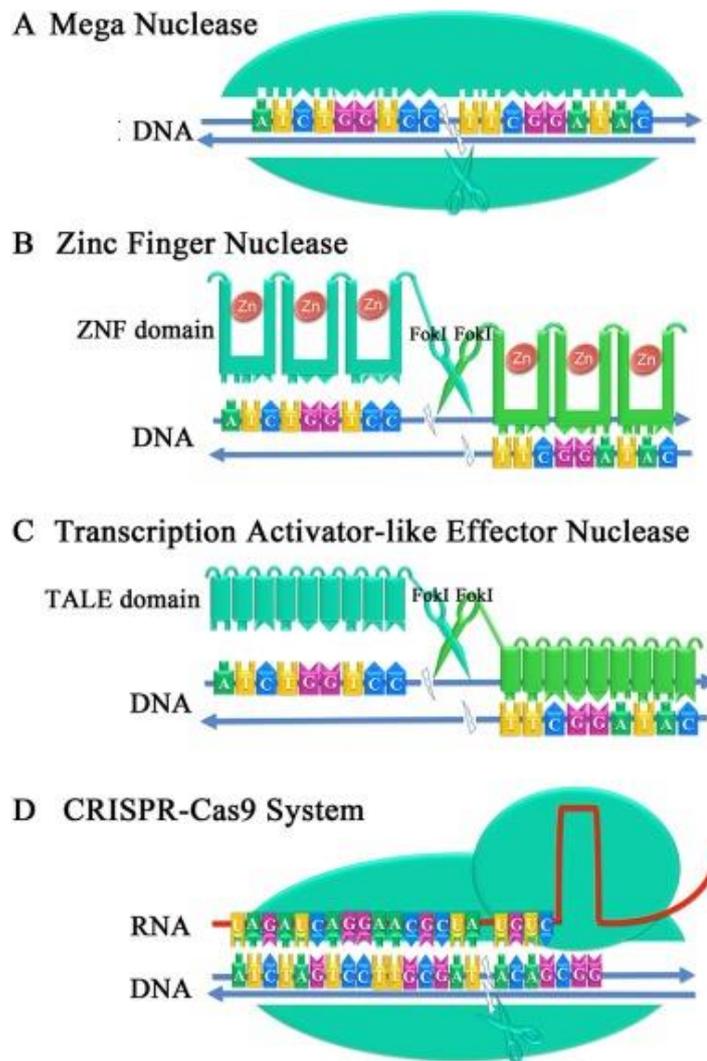


Figure 1-3 A schematic diagram of each of the four main classes of Endonuclease used in gene-editing. Taken from (Gao et al. 2018)

A) Meganuclease (MGN): this is engineering from the naturally occurring homing endonucleases with a recognition site spanning 12-45bps. Notably at a crystal level this comprises of alpha helices and beta sheets, which are contiguous. This served to prove challenging to protein engineering as the binding and cleavage functions did not appear distinct or modular. B) Zinc Finger Nuclease (ZFN), this comprising of CH2 Zinc finger DNA binding domains/monomers, 3-6 of which are attached in tandem prior to the fusion to FokI nuclease at the carboxyl terminus. A single domain or monomer forms an interaction with three nucleotides. Importantly, the left and right zinc finger nuclease are spaced 5-7 nucleotides apart to enable nonspecific FokI type II DNA endonuclease to dimerise and form staggered DNA cuts. C) Transcriptional Activator Like Effector Nucleases (TALENs), in which TALE domains, are attached in tandem prior to the fusion of the FokI nuclease at the carboxyl end. Importantly, 1 TALE monomer identifies 1 Nucleotide of DNA. The FokI nuclease dimerises in a space to produce a double stranded break, akin to ZFN. D) CRISPR/Cas9 comprised of two main components an RNA guide 20 nucleotides in the SpCas9 system and a Cas9 nuclease. The RNA guide confers DNA binding specificity, by Watson-Crick base pairing. The RNA: DNA interface causes conformational changes in the Cas9 nuclease, bringing the RuvC and HNH catalytic domains together upstream of a Protospacer Adjacent Motif recognition site in the genome. This orchestrates DNA cleavage.

1.9 Current Gene-Editing Attempts in the Context of DMD:

In the context of human genetic disease, and as distinct from gene-augmentation, genome editing involves the direct modification of chromosomal DNA with the prospect of providing permanent repair of disease-causing mutations. In the example of DMD this could overcome current life-long adherence to therapeutics, required with corticosteroids and other strategies under development such as Ezutromid, Translana and immunosuppressive drugs (a concomitant therapy with allograft transplantation). Furthermore, theoretically requiring only a single administration, it could overcome immunological complications associated with viral vector gene augmentation strategies, such as AAV-based microdystrophin gene transfer, or full-length dystrophin cDNA gene transfer via AAV-based triple-transplicing mechanisms. As a consequence of these perceived advantages, numerous gene editing therapeutic strategies are being developed for DMD, but are currently confined to *in vitro* and pre-clinical models. This stated, the recent human clinical trials of gene editing in patients with HIV infection, by Sangamo emphasise the therapeutic promise of this evolving field and by extension the approaches under development for the treatment of DMD.

Early demonstration of gene editing approaches for DMD include reading frame restoration by utilising low fidelity of NHEJ DNA repair and the resultant INDELS at a DSB junction to correct the dystrophin reading frame (Figure 1-4). The first demonstration of this was afforded by Ousterout et al (2013a), in which TALENs were directed to an DMD exon 51 sequence, thus generating a DMD myoblast clone bearing INDELS anticipated to restore the dystrophin reading frame (Ousterout et al. 2013a). Notably, this type of therapeutic strategy would be eligible to 13% of patients. Being akin to AON-mediated exon skipping, therapeutic development would face similar challenges requiring efficacious guide designs to a range of target exons and in a clinical setting requiring many individual therapeutic approvals (Aartsma-Rus et al. 2017). Moreover, as INDEL profiles at a DSB are variable, being either additions or deletions of one or more bases, this approach is likely to repair the reading frame in only a third of events (Ousterout et al. 2013a). This stated, recent evidence presented by Overbeek suggests that INDELS resulting from Cas9 cleavage events results in a defined non-random profile retained across replicates. In light of this, the design of CRISPR guides with INDEL profiles biased to the restoration of the dystrophin reading frame could be envisaged (van Overbeek et al. 2016).

Developing upon this approach, NHEJ mediated chromosomal DNA deletion was next developed, in a strategy which multiplexes or uses multiple guides to flank an exon or exons that render the *DMD* gene out of frame (Figure 1-5). The cleavage at both sites subsequently results in the removal of the flanked exons and the restoration of reading frame, to generate

BMD-type dystrophins. Demonstrations of this strategy include deletions of exon 51 (Ousterout et al. 2015), exon 2 duplication (Lattanzi et al. 2017), exon 44 (Li et al. 2015) and exons 45-55 (Ousterout et al. 2013b; Maggio et al. 2016) in human myoblast cultures or iPSCs in (Li et al. 2015). These various gene-editing mediated deletions would have applicability to ~13%, ~15%, ~8% and ~62% of DMD patients respectively. Strikingly, with the exception of Ousterout who used ZFNs, all of these studies used CRISPR/Cas9, thus giving a demonstration of the relative ease with which the Cas9 system can be multiplexed (Cong et al. 2013; Mali, Esvelt, et al. 2013). The final two studies by Ousterout 2013b and Maggio 2016, represent milestones in the field of gene editing not only focusing upon increasing patient applicability but also packaging *Streptococcus Pyrogenes* Cas9 (SpCas9) into lentiviral and adenoviral vectors respectively. This trend towards vectorisation has continued with expression of SpCas9 into a dual AAV vector system, and a smaller bacterial orthologue SaCas9 into single AAV vectors (Wright et al. 2015; Ran et al. 2015). Thus the CRISPR/Cas9 system was quickly transitioned to *in vivo* mouse models of DMD with multiple demonstrations of exon 23 removal in *mdx* mice, and a single exon 53 removal in *mdx4^{cv}* mice (Bengtsson et al. 2017). This approach appears to be favoured in the field currently due to the high efficiency of the NHEJ pathway, however, the inclusion of two guides brings with it inherent risk of increased off-target effects. Moreover, both guides need to cleave at high and comparable efficacy to provide a high occurrence of the desired deletion.

The final type of these gene editing strategies aiming to modify and correct chromosomal sequences is reliant upon HDR which whilst less efficient than NHEJ, orchestrates more precise correction of the genome (Heyer et al. 2010). Furthermore, this HDR-mediated process can substitute or restore genetic material lost in the endogenous genetic mutation. The simplest of these strategies uses single stranded oligonucleotides (ssODNs) as the repair template to correct subtle point mutations within the genome (Figure 1-4). A notable example of this was recently achieved in *mdx* zygotes using an ssODN to introduce a point mutation into the endogenous nonsense mutation in exon 23 and thus inducing mosaicism in mice and 2%-100% dystrophin expression restoration. This type of approach could be used to address point mutations which occur in 20% of DMD patients. Alternatively, exogenous repair templates encoding a range of replacement cDNA sequences flanked by isogenic sequences upstream and downstream of the DSB in the genome could be designed to introduce larger sequence elements (Figure 1-5). Demonstrations of this include insertion of exon 44 in iPSCs (Li et al. 2015), exons 45-52 in immortalised patient myoblasts harbouring for this deletion (Poplewell et al. 2013) and exon 53 in *mdx4c* (Bengtsson et al. 2017). Research by Poplewell et al (2013) was important to this field of therapeutics as the first demonstration; using patient myoblasts

harbouring a major deletion that restoration of the dystrophin reading frame could be achieved by a MGN mediated knock-in of a 4.2kb cDNA encoding exons 45-52.

The investigations presented in this thesis aim to assess the feasibility of a CRISPR/Cas9 mediated knock-in of an exon sequence based on the microdystrophin cDNA (MD1; Le Guiner et al, 2017) at the 5' end of the DMD gene. Notably, the exogenous repair template is promoter less and would utilise the endogenous full length dystrophin Dp427m promoter to drive expression. This work has potential therapeutic applicability to 95% of patients. Moreover, a novel exogenous repair template was constructed with the propensity to be adapted for other exonic cDNA insertions notably of $\Delta 45-55$ minidystrophin and full-length dystrophin. This approach could provide permanent correction of diverse disease-causing mutations across the DMD gene, providing a correction strategy for a patient cohort in which there is currently an unmet clinical need.

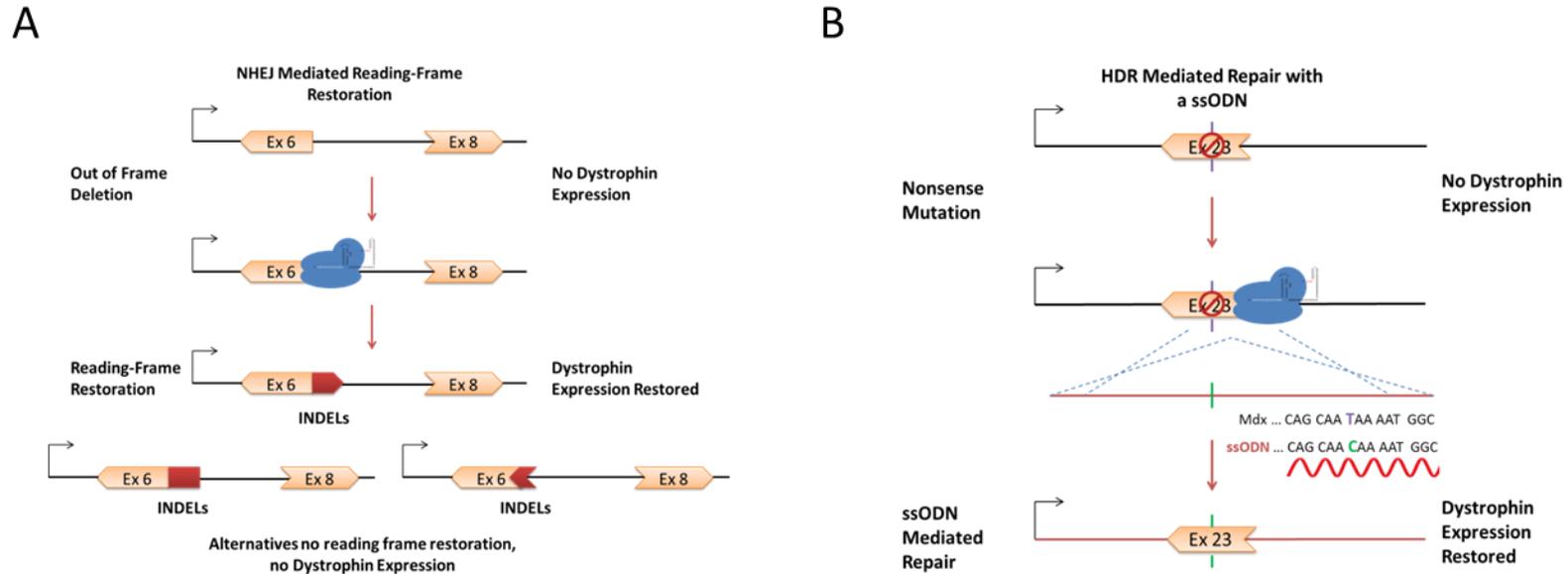


Figure 1-4 Diagrams highlighting NHEJ-reading frame restoration and ssODN integration gene-editing strategies that could be employed to fix a variety of DMD mutations.

A) Utilising NHEJ to overcome the deletion of Exon 7 and restore of the dystrophin reading frame. In the absence of exon 7, exons 6 and 8 are ‘out of phase’ with one another and the resultant mRNA is incorrect and degraded, resulting in the absence of dystrophin. Following the introduction of a DSB by Cas9, the DNA can be repaired by the efficient but error prone NHEJ pathway. This causes random small insertions or deletions in the local DNA, which in some instances can restore the reading frame and by extension dystrophin expression by ‘chance’. B) HDR mediated repair of point mutations using single stranded oligo nucleotide. In this instance a nonsense mutation in Exon 23, causes premature termination of dystrophin transcription. The subsequent transcript then undergoes decay and no full length dystrophin is produced. By cleaving DNA in the nearby area and delivering a ssODN with subtle mutations this can be corrected. Once broken HDR can occur in which the ssODN can ‘invade’ the DNA displacing the mutated sequence and integrating in the correct nucleotide. This stops the formation of the premature stop codon allowing full length dystrophin to be transcribed (Long et al. 2014). Taken from Moore et al. 2015.

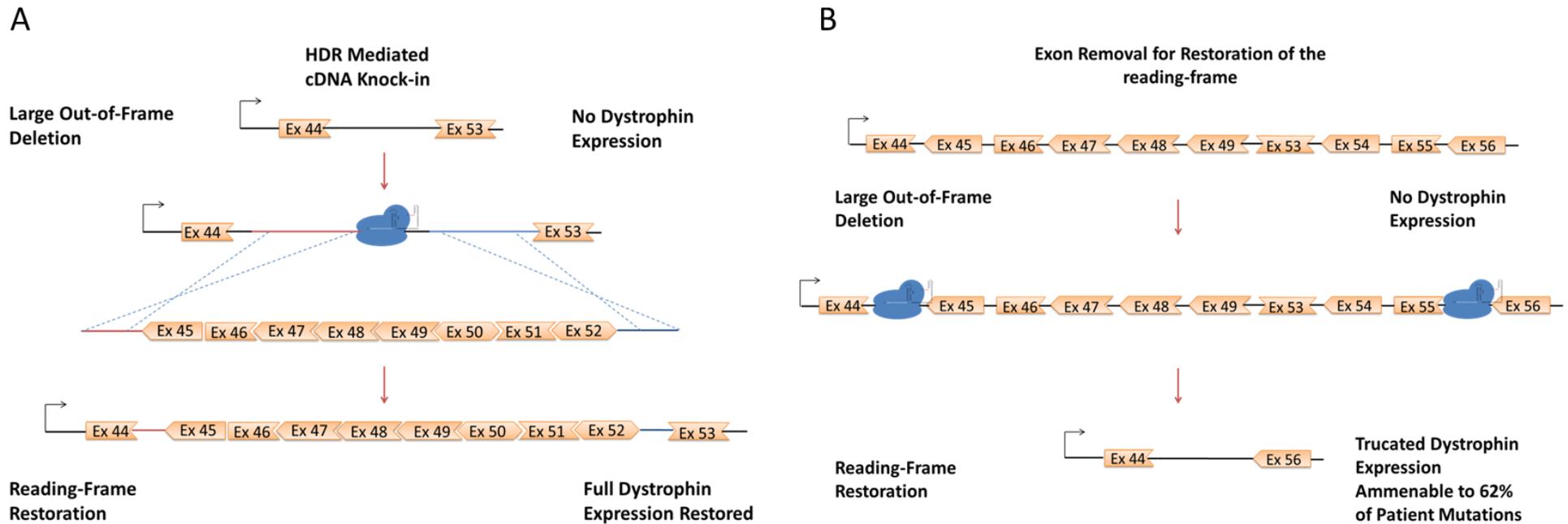


Figure 1-5 Schematic diagrams highlighting HDR-reading frame restoration by cDNA exogenous repair template integration and NHEJ-mediated deletion gene-editing strategies that could be employed to fix a variety of DMD mutations.

A) HDR Mediated repair utilising a cDNA repair template, can be used to insert pieces of the DMD gene that have been lost. In this instance production of a DSB and a cDNA donor containing exons 45 -52 can restore the reading frame and full-length dystrophin expression (Poplewell et al. 2013). B) Multiplexing CRISPR guides can be used to remove exons in the restoration of the reading-frame and the expression of truncated dystrophin. The deletions of exons 50-52 in this example can be overcome by cutting out exons between Intron 44 and 55. The removal of the exons included within this region enables the restoration of the reading frame, with exon 44 and 56 being spliced together. This removes a mutational hotspot accounting for 62% of patient mutations and enables the production of a shortened dystrophin with some function that could provide therapeutic benefit (Kabadi et al. 2014; Maggio et al. 2016). Taken from Moore et al. 2015.

1.10 Project Aims:

Duchenne Muscular Dystrophy (DMD) is a complex neuromuscular disease with multisystem manifestations including muscular wasting, impaired cognitive functioning, reduced gastrointestinal functioning, cardiac conduction defects and respiratory impairment (Mercuri & Muntoni 2013; Bushby et al. 2010a). Whilst clinical management requires a multidisciplinary team, care is palliative treating only the secondary manifestations of the disease rather than the underlying genetic aetiology. Furthermore the only current treatment option is corticosteroids. Whilst administration of corticosteroids dampens muscle inflammation, prolonged use is contraindicated due to side effects including: bone demineralisation, weight gain and behavioural changes (Mercuri & Muntoni 2013).

There is a clear unmet medical requirement for improved treatment options which has spurred the development of pharmacological, cell-based and gene therapies. Importantly, this has resulted in a number of DMD therapeutic strategies advancing to clinical trial, examples include: Ataluren (TM Translana), exon skipping (Drisapersen and Eteplirsen/ExonDys51) and SMTC1100 (Ezutromid). Notably, Ataluren has gained conditional marketing authorisation and drug approval in Europe, and Eteplirsen/ExonDys51 FDA approval. These treatment strategies are mutation specific with application limited to specific patient cohorts. Ataluren treats nonsense mutations, thought to account for 13-15% of DMD cases (Bushby et al. 2014). Exon skipping of exon 51, by Eteplirsen could confer benefit to ~13% of patients (Aartsma-Rus et al. 2009). Furthermore, 64% of patients would be eligible for single exon skip antisense reagents, 79% for double exon skip and 90% for multi-exon skipping therapies (Verhaart & Aartsma-Rus 2012). Notably, this is not applicable to all patients and gaining approval of AONs on an individual basis represents a significant hurdle to surmount, particularly in exons skips with very low patient applicability.

Awareness of the limited number of DMD patients amenable to mutation specific treatments has prompted the scientific community to consider more universal approaches to achieve the amelioration of dystrophic phenotypes. Two strategies that exemplify this are AAV-microdystrophin gene augmentation and utrophin upregulation by SMTC1100 (Ezutromid). Importantly, microdystrophin gene augmentation has shown promising results when trialled in a range of pre-clinical models of DMD (Foster et al. 2008; Koo et al. 2011; Le Guiner et al. 2017); although one clinical trial has been completed that failed to show suitable efficacy, revisions to the study design have been highlighted by the investigators that would likely improve clinical outcomes (Mendell et al. 2012). In contrast SMTC1100 has advanced through safety and efficacy phase I trials and is currently progressing to phase II (NCT02858362) (Tinsley et al. 2015). Whilst promising, these approaches have inherent limitations. In the case

of AAV- microdystrophin, repeated administration may likely be required, increasing the risk of adverse immunological side effects. Furthermore, being an internally truncated form of dystrophin, these recombinant microdystrophins may not retain all of the necessary functionality of wild type dystrophin. In contrast, studies using SMTC1100 indicate that attaining optimal serum concentrations of the drug remains challenging and such a therapy would require life-long adherence to drug administration. Furthermore, despite structural similarity to dystrophin, utrophin has a distinct expression profile, is unlikely to interact directly with microtubules like dystrophin or localise nNOS correctly (Belanto et al. 2014; Li et al. 2010). In light of this, utrophin upregulation may prove compensatory but is unlikely to address all dystrophic pathology.

The research described in this thesis, was based upon a novel gene-editing approach in which intron 1 of the DMD gene was targeted. This approach hinged upon the successful introduction of a DSB by a genome editing designer endonucleases and co-delivery of an exogenous repair template. Notably, the promoter less exogenous repair template employed in this study was constructed to enable cloning of microdystrophin (MD1; le Guiner et al 2017), minidystrophin (deleted for exons 45-55: $\Delta 45-55$) and full length dystrophin cDNA variants, from dystrophin exon 2 onwards. In this manner, the HDR repair pathway will be exploited to enable integration of these cDNA elements into DMD intron 1 as a large artificial exon downstream of the native exon 1 DMD gene sequences. This targeting of the exogenous repair template to intron 1 has the advantage of utilising the endogenous full-length dystrophin promoter, Dp427m, to drive expression. Furthermore being close to the start of the DMD gene, it provides a near universal and permanent therapeutic strategy to potentially overcome most DMD mutations. In order to contribute towards this goal our aims were defined as follows:

- Customisable endonucleases were designed to intron 1 of the DMD gene (Chapter 3). Intron 1 was identified as most desirable as it would maximise patient applicability to 95% of patients.
- Efficacy of endonuclease designs were assessed using mismatch detection assays (Chapter 4).
- Assessment of the effects of cDNA coding sequence optimisation on full length, $\Delta 45-55$ and MD1 protein expression was undertaken and an appropriate exogenous repair template designed (Chapter 5).
- Finally, these reagents were combined and tested in both HEK293T and immortalised DMD patient myoblasts.

These aims were designed to provide a proof-of-principle demonstration of this near universal genome editing strategy to counter DMD mutations, assessing whether targeting DMD intron 1 and utilising the endogenous full-length promoter is a feasible strategy for future *in vivo* development and testing.

Chapter 2 Materials and Methods

2.1 General Laboratory Reagents

All general chemical reagents were purchased from Sigma, Invitrogen, BDH or VWR (unless stated otherwise) with standard chemical purity graded as AnalaR (Analytical Reagents for analysis applications): As listed below. All of the reagents were dissolved in double distilled H₂O (ddH₂O), unless stated otherwise. The solutions required for tissue culture were autoclaved at 121°C for 15 min, except for the solutions containing protein, detergents or glucose which required passing through a 0.22 µm filter (Falcon). All solutions were stored at room temperature unless indicated otherwise.

Reagent :	Manufacturer:
Acetic acid (CH ₃ COOH)	VWR
Agarose	Invitrogen
Ammonium acetate AcO(NH ₄) ₂	BDH
Bovine Serum Albumin	Invitrogen
Boric acid	Sigma
CaCl ₂	Sigma
Citric acid	Sigma
Dimethyl sulphoxide (DMSO)	Sigma
EDTA	Sigma
Ethanol (EtOH)	VWR
Glucose	Sigma
Hydrochloric acid (HCl)	BDH
Liquid broth (LB)	Sigma
Methanol	VWR
Magnesium chloride (MgCl ₂)	Sigma
Magnesium sulphate (MgSO ₄)	BDH
NaOH	Sigma
Na ₂ HPO ₄	Sigma
NP-40	Fluka Biochemica
Paraformaldehyde (PFA)	Sigma
Potassium chloride (KCl)	Sigma
Phosphate buffered saline (PBS) pH 7.3	Gibco
Sodium chloride (NaCl)	Sigma
Sodium dodecylsulphate (SDS)	Sigma
Sodium bicarbonate (NaHCO ₃)	Sigma
Sucrose	Sigma
Tri-sodium citrate	Sigma
Trizma Base	Sigma
Trizma hydrochloride (HCl)	Sigma

2.2 Bioinformatics Assessments of the 5' End of the *DMD* gene

2.2.1 Common Intron 1 Sequence Identification

2.2.1.1 *Dystrophin Isoforms*

- Dp427c NM_000109.3
- Dp427m NM_00406.2
- Dp427p1 NM_004009.3
- Dp427p2 NM_004010.3

2.2.1.2 *Genomic Builds*

- Homo Sapien Chromosome X, GRCh37.p13 Primary Assembly (Accession NC_000023.10)
- C57BL/6J Chromosome X, GRCm38.p1 C57BL/6J Primary Assembly (Accession NC_000086.7)

The first exon of all full length isoforms and exon 2 of Dp427m where they converge was downloaded in standard FASTA format. This was then aligned to the genomic builds using NCBI BLAST, accessed here: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. The genomic co-ordinates were then used to identify and download a genomic region between Dp427p2 and Exon 2 for both human and mouse genomes.

2.2.2 Rigorous Local Sequence Alignment:

The human and mouse common intron 1 sequences and the subsequent 345bp region of interest identified, were both aligned using EMBOSS matcher from the EMBL-EBI bioinformatics suite (McWilliam et al. 2013). Under default parameters: BLOSUM62 Matrix, Gap Open 14, Gap extend 4, Alternative matches 1, Output pair.

2.2.3 Endonuclease Design

Sequences identified as candidates for the design of endonucleases, were subject to analysis for TALEN and guide design using TALEN Hit Software: (Collectis Bioresearch 2011) and CRISPR MIT (Optimised CRISPR Design 2013). The only selection made was for the human or mouse genome builds on the CRISPR MIT tool otherwise, all parameters were default.

2.3 Bacterial Cultures

2.3.1 Materials

- Ampicillin (Sigma): prepared as 1000x stock in ddH₂O at 50 mg/ml concentration, filtered through 0.22 µm filter (Falcon) and stored at -20°C.
- Kanamycin (Sigma): prepared as 200x stock in ddH₂O at 20 mg/ml concentration, filtered through 0.22 µm filter (Falcon) and stored at -20°C.
- Agar (Fisher Scientific)
- Glycerol (Sigma)
- Lysogeny Broth (LB) (Sigma).
- LB agar: 3g agar was added to 4g LB solution made up to 200ml with ddH₂O. The LB-agar solution was incubated at 60°C for at least an hour and subsequently cooled

before the addition of desired antibiotic. ~25ml of LB-agar was added into 100 mm petri dish in proximity to the Bunsen burner. Once solidified the plates were sealed in cling film and stored at 4°C for no longer than 1 month.

- LB SOC medium: 100 µl of 1 M MgSO₄ and 20 µl 1 M glucose (0.22µm filtered) were added to 10 ml LB medium.
- TOP 10 (*E.coli*) competent cells (Invitrogen/ NEB)

2.3.2 Heat Shock Transformation of Bacteria:

A plasmid concentration of between 100-200ng of DNA contained within 5µl, 1/10th of the volume of bacterial competent cells was added to 50µl of competent cells, flicked gently to agitate the suspension and incubated at 4°C on ice for 30 minutes. The bacterial suspension was then shocked at 42°C for 30 seconds, and allowed to recover at 4°C on ice for 2 minutes. Subsequently, 250µl of antibiotic-free LB SOC media, an enriched media, was added to the bacterial suspension. This was then incubated at either 32°C or 37°C degrees for 1 hour shaking at 200rpm. This is then plated at alternative volumes 50 or 100µl, in an attempt to attain correct bacterial densities and incubated overnight on agar plates with appropriate antibiotic selection.

2.3.3 Production of Starter Cultures and Glycerol Stocks

If bacterial colonies were present, after overnight incubation at 37°C, they were transferred to 5ml LB in a universal tube and supplemented with a corresponding antibiotic. This was performed with either a sterile loop or a Gilson pipette. Once a starter culture had been inoculated, this was incubated at 37°C for 8-16 hours, shaking continuously at 200rpm.

Post incubation 750µl of Bacterial suspension was mixed with 250µl of 80% sterile glycerol; the resultant glycerol stock was stored at -80°C. The remaining 4.25ml of bacterial suspension could either be used to inoculate a larger Maxi or Mega-prep culture or centrifuged at 3000rpm for 15minutes (Beckmann Centrifuge). If centrifuged, the bacterial pellet was subject to DNA extraction using the Qiagen Mini-Prep kit or stored for later extraction at -20°C.

2.3.4 Mini, Maxi and Mega Preparations:

To produce small amounts of plasmid DNA, for the purposes of restriction-digest mapping, a bacterial pellet from the centrifugation of the starter culture was processed using the Qiagen Miniprep Kit. Larger and purer DNA preparations were produced by Maxi and Mega prep cultures: Maxi 500µl of starter culture inoculated into 250ml of LB supplemented with antibiotic, Mega 1000µl of starter culture inoculated into 500ml of LB with antibiotic. These bacterial suspensions were grown for 16 hours prior to being centrifuged at 3000rpm for 45 minutes (Beckmann coulter) and processed with Qiagen Maxi or Mega prep kits respectively.

2.3.5 Picking Colonies for Colony PCR:

If colonies were to be selected for screening by colony PCR, the colony in question would be picked and inoculated into 10µl of LB media. This was performed for 8 colonies using a labelled 8-tube PCR strip, and each suspension mixed. Then using a multiple pipette boy, 5µl of all samples was aliquoted into fresh tubes to ensure samples were produced in duplicate. One group was then be boiled at 98°C for 2 minutes to lyse bacteria and provide a crude DNA suspension prior to being stored at 4°C in the PCR thermocycler. The second group was stored at 4°C and used to produce a starter culture if the colony was identified as consistent.

2.4 Plasmids

All plasmids for this investigation were either Maxi or Mega prepped, eluted in water or TE buffer and working stocks diluted to approximately 1µg/µl prior to storage at -20°C. It should be noted that in all cases ampicillin resistance genes are within the plasmid backbone, thus this was used in bacterial culture methods, during plasmid production.

2.4.1 Gene- Editing Plasmids

2.4.1.1 U6 Guide Plasmid:

The plasmid was kindly gifted by Dr. Taeyoung Koo, from Professor Jin Soo Kim's lab. This comprises of a U6 RNA polymerase III promoter upstream of a guide cloning site, consisting of a 677bp spacer region flanked at the 5' and 3' end with asymmetric BsaI restriction sites. Immediately downstream of the 3' BsaI restriction site is the invariant tracrRNA sequence. Upon cloning of a 20 nucleotide 'crRNA' or spacer region, between the BsaI sites a full chimeric single guide RNA will be formed. This plasmid was used as the backbone for the production of all U6-driven guide sequences in this investigation.

2.4.1.2 HC-Cas9 Plasmid

The plasmid was kindly gifted by Dr. Taeyoung Koo, from Professor Jin Soo Kim's lab. This contains human codon optimised *Streptococcus Pyrogenes* Cas9 downstream of Cytomegalovirus (CMV) and a T7 RNA polymerase promoter (T7). Finally, the 3' end of Cas9 has a HA tag and nuclear localisation signal (NLS), prior to terminating in a bovine growth hormone polyA. This plasmid was used for the T7 RNA polymerase and was used for Cas9 In Vitro transcription experiments.

2.4.1.3 CBh-Cas9 (Px165)

This plasmid was ordered from Addgene (Plasmid Number #48137), where it had kindly been donated by the Zhang lab group (Ran et al. 2013). It contains human sequence-optimised *Streptococcus Pyrogenes* Cas9 (SpCas9), under the control of the Chicken Beta Actin (CBh) promoter. This is a modified CAG promoter with validated robust expression in a range of mammalian cells. Importantly, the SpCas9 transgene is 3x flag-tagged at the 5' end and

contains (NLSs) at both the 5' and 3' termini, prior to terminating in a bovine growth hormone poly A site. This construct requires dual transfection alongside a U6-driven guide plasmid to mediate gene-modification.

2.4.1.4 U6- Guide- CBh-Cas9-2A-GFP (Px458)

This plasmid was ordered from Addgene (Plasmid Number #48138), where it had kindly been donated by the Zhang lab group (Ran et al. 2013). This is a sister plasmid to the Px165. Distinct from the Px165 it contains a U6 RNA III polymerase, with BbsI two asymmetric restriction sites spaced 22bps apart, immediately downstream of this is the invariant tracrRNA sequence. Upon cloning of a 20 nucleotide 'crRNA' or spacer region, between the BsaI sites a full chimeric single guide RNA will be formed. Similarly to the Px165 downstream of this site the plasmid comprised the same CBh promoter, followed by 3x flag tags and NLSs at both the 5' and 3' ends. However, immediately after this is a 2A viral peptide linker, followed by an eGFP and a bovine growth hormone polyA. Thus this plasmid is particularly useful as an All-in-one Cas9 delivery system. Furthermore the bicistronic Cas9 and eGFP expression proved useful in allowing fluorescence to function as a proxy for transient transfection efficiency.

2.4.1.5 pet-MBP-Cas9

Custom Pet plasmid was used in Paris with Dr. Jean Paul Concordert and Dr. Anne De-Cian to transform Rosetta BL2 competent bacterial cells and induced Cas9 protein production. It comprises of an N-terminal terminal hexahistidine maltose binding protein (His-MBP), sequence optimised SpCas9 and a C-terminal Histidine tag. This plasmid was modified from a plasmid on Addgene (plasmid number #39312) kindly donated by Doudna's lab (Jinek et al. 2012). Importantly, additional nuclear localisation signals were sub-cloned at 5' and 3' ends of the Cas9 transgene, in an effort to enhance protein production.

2.4.1.6 pBIGT-VGEJ-AAV

This plasmid was used in Paris with Dr. Jean Paul Concordert and Dr. Anne De-Cian during an in vitro cleavage assay with Cas9 protein. The main constituent of interest to this investigation was a unique XmnI site, allowing for the 5.8kb plasmid to be linearised and a recognition site for a published GFP T2 guide (Mali, Yang, et al. 2013).

2.4.1.7 pCI-GFP

This plasmid was kindly gifted by Dr. Susan Jarmin. It comprises of a Cytomegalovirus (CMV) promoter, which drives enhanced Green fluorescent protein (eGFP). This terminates on an SV40 PolyA signal. This resultant green fluorescent signal was used as a proxy for transient transfection efficiency.

2.4.2 Dystrophin Plasmids

Investigations using dystrophin were concerned with the examination of sequence optimised and native dystrophin constructs. In the former the sequence has been optimised to enhance dystrophin transcription. As such three main dystrophin variants were examined in both sequence optimised and native formats including: Full length dystrophin, $\Delta 45-55$ minidystrophin with the in-frame deletions of Exons 45-55 and finally MD1 R4-23 Δ CT. The latter, MD1 contains the minimal necessary constituents for functional dystrophin, with deletions of spectrin-like repeats R4-23 and the C-terminal domain.

Native variants – The cDNA of dystrophin variants is that physiologically present /naturally occurring. This was constructed from consensus sequence.

Optimised Variants – The cDNA of dystrophin variants has been subject to the GeneArt ‘sliding window’ algorithm to optimise codon usage to species, avoid sequence motifs detrimental to transcription in order to boost expression.

Crucial to understanding these constructs, the experiments were undertaken in two plasmid backbones: pCI and pAAV. The full length and $\Delta 45-55$ minidystrophin variants were tested in both plasmid backbones respectively. In contrast, the MD1 was tested in pAAV in alone.

2.4.2.1 pCI Backbone:

The standard pCI backbone contains a cytomegalovirus (CMV) promoter which was used to drive expression of full length and $\Delta 45-55$ dystrophin transgenes respectively. The individual constructs are tabulated below:

Native and Sequence Optimised Dystrophin Variants, in a standard pCI-Backbone driven by a CMV promoter		
Full Length	CMV- Nat hDys	CMV-HuDys- SO
$\Delta 45-55$	CMV- Nat hDys $\Delta 45-55$	CMV-HuDys $\Delta 45-55$ – SO
MD1	N/A	N/A

2.4.2.2 pAAV Backbone

The standard pAAV backbone was used for the expression of dystrophin cDNA variants. It contains, a muscle restrictive Spc5-12 promoter, upstream of dystrophin cDNA variants, an SV40 PolyA, an ampicillin cassette and two inverted terminal repeats necessary for packaging into AAVs.

Native and Sequence Optimised Dystrophin Variants, in a standard pCI-Backbone driven by a Spc5-12 Muscle restrictive promoter		
Full Length	Spc5-12- Nat hDys	Spc5-12 -HuDys- SO
$\Delta 45-55$	Spc5-12 - Nat hDys $\Delta 45-55$	Spc5-12 -HuDys $\Delta 45-55$ - SO
MD1	Spc5-12- Nat- hMD1	Spc5-12 – HuMD1 –SO

All aforementioned dystrophin native and sequence optimised variants were cloned and kindly gifted to me by Dr. Susan Jarmin.

2.4.3 Repair Template Plasmids

2.4.3.1 *pUC57-hINT1-RT*

This plasmid was synthesised by GenScript, in accordance with the exogenous repair template sequence provided, with avoidance of the unique restriction sites as requested. This is a promoter-less plasmid comprising: a 1kb left arm of homology, a β -globin splice acceptor, sequence optimised dystrophin exons 2-6, a Cyan Fluorescent Protein (CFP) with the initiating methionine residue removed, flanked between 5' FseI and 3' NotI restriction sites, a mutated Woodchuck Hepatitis Virus Post-transcriptional regulatory element (mWPRE) and polyA, a Sv40-Zeocin-Sv40 polyA and a 1kb right arm of homology. The sequence for this plasmid was constructed throughout the course of the investigation and formed the foundation of the exogenous repair template. The FseI and NotI, restriction sites enabled directional cloning of all of the sequence optimised dystrophin variants.

2.4.3.2 *pUC57-hINT1-MD1-RT (MD1 RT)*

This is the daughter plasmid of Puc57-hINT1-RT. The only distinction is the CFP between the FseI and NotI restriction sites has been removed and MD1 Exon 6 onwards has been sub-cloned into the construct. As a consequence, MD1 Δ Ex1 was reconstituted downstream of the β -globin splice acceptor. This is used for all of the integration-based experiments in this investigation.

2.4.3.3 *Lentiviral ISce-IT-hINT1-RT*

This plasmid was kindly gifted by Prof. Rafael Yanez. It encodes the same variants as the Puc57-hINT1-RT between the 1kb left and right arms of homology. This aforementioned sequence was sub-cloned in reverse orientation between AgeI and KpnI restriction sites within the lentiviral backbone; this was to avoid any aberrant splicing events that could occur during viral synthesis. The lentiviral backbone contained: the Gag gene, Primer binding site (PBS), Rev Responsive element, a central polypurine tract and long terminal repeats; all sequences necessary for lentiviral production.

2.5 Computer Software used throughout Thesis:

- Vector NTI / Snap Gene – to visualise and analyse plasmids of interest.
- GraphPad Prism – for graphical presentation and statistical analysis of samples.

2.6 Descriptive Statistical analysis:

All graphical portrayals of the datasets use mean as a measure of the central tendency, and standard error of the mean (SEM) as a measure of deviation within the samples. Where

appropriate, a Shapiro-Wilks test was used to establish the normal distribution of datasets in question, the null hypothesis being that the data input follows a normal or Gaussian distribution. Thus in instances where the p-value derived is greater than the alpha value ($p > 0.05$) the true null hypothesis cannot be rejected and the data is deemed normally distributed. In instances of a normal distribution the data sets were subject to a parametric statistical analysis.

2.6.1 Inferential Data Analysis

Comparison of two groups: a comparison of two groups was performed if by unpaired t-test if the datasets adhered to a normal distribution.

Comparison of multiple groups: If the conditions of normality were met between multiple conditions. Then a one-way or two-way ANOVA was performed.

2.7 Molecular Sub-Cloning Methods

2.7.1.1 Materials

- Restriction enzymes and buffers (All NEB products).
- 50x TAE: 242 g Tris Base, 57.1 ml Glacial Acetic Acid, 100 ml of 0.5 M EDTA pH 8.0 were brought to a total volume of 1 L with ddH₂O.
- UltraPure Agarose (Invitrogen)
- 1000x SYBR Safe DNA gel stain (Invitrogen).
- 5x Loading buffer (BioLine).
- DNA molecular weight markers, Hyperladders I, IV and V (BioLine):
- Qiaquick gel extraction kit (Qiagen).
- Qiaquick PCR Purification Kit (Qiagen)
- High Salt Buffer (Roche)
- T4 DNA ligase (NEB/ Promega/ Fermentas).

2.7.1.2 Endonuclease Restriction Digest

To confirm the identity of plasmids produced, restriction digests were designed using SnapGene and restriction digest conditions checked on the NEB website (<https://nebcloner.neb.com/#!/redigest>). Once designed a 1µg aliquot was subject to restriction digest mapping using a variety of single and multiple cutters. Restriction digest reactions comprised of: 1x Cutsmart Buffer (NEB), 1µl of Enzyme (~10U, adjusted if necessary), 1µg of plasmid DNA preparation and ddH₂O to attain a 20µl volume. This was then incubated at 37°C for 3 hours to ensure digestion had continued to completion prior to the restriction enzyme being inactivated at 80°C for 20 minutes. The latter step being optional as inactivation is not suitable for all enzymes.

2.7.1.3 Gel Electrophoresis

To permit examination of plasmid digestion profiles, the samples were loaded on agarose gels. Agarose percentages of 0.5%, 1% and 2% (mass: volume) in 1x TAE buffer, were routinely used in the separation of DNA, with selection being influenced by the anticipated band fragments sizes. Following microwaving and agitating the solution to ensure agarose had completely dissolved, it was cooled slightly and inoculated with SYBR[®] Safe DNA stain to attain a final concentration of 0.01%. This was then poured in a 10cm dish and allowed to set. An 8 μ l aliquot of DNA samples were then mixed with 2 μ ls of 5x loading dye, and loaded in to the wells, alongside 5 μ l of a hyperladder of appropriate range to determine DNA size. The samples were then electrophoresed at 50v for 1 hour, and at 70v for sequential 20 minute intervals until DNA fragments had separated adequately for determination of size. The DNA fragments were then visualised under U.V as permitted by the SYBR[®] Safe, the sizes determined by comparison to hyperladder were then compared with those anticipated from SnapGene.

2.7.1.4 Gel and PCR Purification:

Desirable DNA fragments for sub-cloning either in the form of the destination vector or the insert, required purification to ensure the maximal success for sub-cloning projects. If the products were separated on an agarose gel, the fragment was excised and solubilised using Qiaquick Gel-extraction protocols in accordance with manufacturer's protocols, to attain DNA of interest. In contrast, if the DNA was amplified, it was subject to Qiaquick PCR purification, in accordance with manufacturer's protocols.

2.7.1.5 T4 DNA Ligation

T4 DNA ligase facilitates the phosphodiester linkage formation between an insert and a destination vector. Molar ratios of 1:5 and 1:10 of destination vector: insert are used, in which the destination vector is used at a concentration of either 50 or 100ng. Standard reactions comprised: 2 μ l of 10x ligase buffer and 1 μ l of T4 ligase, and ddH₂O to attain a constant total reaction volume of 20 μ l. Ligations were commonly incubated at 22°C for 1 hour and then 16°C overnight, heat inactivated at 65°C for 10 minutes. If the concentration of destination vector and insert permitted, this 20 μ l reaction was reduced proportionately to a 10 μ l volume, to improve ligation efficiency.

2.8 Guide Sub-Cloning Methods

2.8.1 Background

In order to produce U6-driven guides that could be co-transfected alongside a plasmid expressing Cas9 nuclease, oligonucleotides were ordered from Eurofins MWG or IDT, annealed and ligated into either the U6 guide or the Px458 plasmids. Overhangs 5' CACCG' were appended to the forward 20 nucleotide guide oligo and 5'AAAC' appended to the reverse guide oligo. Notably, a 'C' appended at 3' end of the 20 nucleotide guide sequence, to permit the addition of the 5' Guanine. In these investigations a GX²⁰ guide format was utilised.

2.8.2 Human Primers

Primer Name	Primer Sequence 5'-3'
hINT G1 F:	CACCGAGACTAAGAACCAGATGCTA
hINT G1 R:	AAACTAGCATCTGGTTCTTAGTCTC
hINT1 G4 F:	CACCGTCTGTCTAATTACATTATAA
hINT1 G4 R:	AAACTTATAATGTAATTAGACAGAC

2.8.3 Mouse Primers

Primer Name	Primer Sequence 5'-3'
mINT G1 F:	CACCGAGACTAAGGACCAGATGCTA
mINT G1 R:	AAACTAGCATCTGGCCTTAGTCTC
mINT1 G2 F:	CACCGTAATGCAATTAGACAGACTA
mINT1 G2 R:	AAACTAGTCTGTCTAATTGCATTAC
mINT1 G3 F:	CACCGTCTGTCTAATTGCATTATAA
mINT1 G3 R:	AAACTTATAATGCAATTAGACAGAC
mINT1 G4 F:	CACCGTATGAAGCATGGTATTATTT
mINT1 G4 R:	AAACAAATAATACCATGCTTCATAC
mINT1 G5 F:	CACCGTATTTTGCTTCCTTAGCATC
mINT1 G5 R:	AAACGATGCTAAGGAAGCAAATAC

2.8.4 Preparation of the U6 Guide Plasmid

Standard Bsal digests were performed on 2µg of U6- guide plasmid in triplicate; digests were undertaken for 2 hours at 37°C. This was performed alongside 1µg internal controls of uncut (enzyme substituted with water), linearized with NotI enzyme (cuts once) and 1µg Bsal digest. Post digestion enzymes were inactivated. Then 10µl of internal controls and the full 20µl of Bsal digest was resolved on a 0.8% agarose gel. This was visualised by the white light as opposed to U.V to prevent degradation. The 2464bp, corresponding to the U6 vector with the removal of the 677bp spacer region, was subject to Qiaquick gel extraction for the 3 triplicates. If the 1st sample had low yield, below 10ng/µl the resultant mixtures were pooled.

2.8.5 Self- Annealing Oligonucleotides

A reaction mixture of oligonucleotides was produced comprised of: 5µl Roche high Salt (H buffer), 10µl of forward and reverse primers at a concentration of 100pmol/µl and 75µl of ddH₂O. This was then subject to the 'Surveyor hybridisation programme'. See section 1.15.

2.8.6 Ligation

This was then subject to a 10µl ligation, comprising: 1µl T4 ligation buffer, 1µl of plasmid backbone, 1µl of annealed oligos, 1µl T4 ligase and 6µl of ddH2O.

2.8.7 U6 Guide Colony PCR

2.8.7.1 Materials

- 2x PCR Mastermix (GeneSys Ltd)
- Primers

Primer Name	Primer Sequence 5'-3'	Amplicon Size (Bp)
U6 Colony PCR F	CATATGCTTACCGTAACTTGAAAG	787bps (with U6 guide plasmid spacer)
U6 Colony PCR R	CGGACTAGCCTTATTTAACTTGC	144bp (if a 20 nucleotide guide was cloned in).

A 1µl aliquot of crude DNA extract produced from boiling a bacterial colony in 5µl of LB (see section 1.3.5), was added to a 2x Mastermix PCR (GeneSys Ltd). This was produced in accordance with the standard PCR method (section: 15.1.1).

The cycling conditions: Denaturation 98°C 2 minutes, **35 cycles** – (98°C 30s, 60°C 30s, 72°C 45s), **Final Extension** 72°C 10 minutes and held at 4°C.

This resultant amplicon was then resolved on a 2% agarose gel alongside hyperladder IV.

2.8.7.2 U6 Guide plasmid Sequencing

If colonies screened, indicated the insertion of self-annealed primers by colony PCR, bacterial starter cultures were produced and mini preps undertaken. The resultant DNA preparation was diluted to 5ng/µl for a 15µl volume and supplemented with 2µl 10pmol/µl [10µM] U6 sequencing primer 5' 'GACTATCATATGCTTACCGT'3'. This was then submitted to Eurofins for overnight sequencing, to fully confirm the identity and presence of the guide.

2.9 Sub-cloning Microdystrophin into the pUC57-hINT1-RT.

2.9.1 Preparation of plasmids

A 2µg concentration of both Puc57-hINT1-RT and Spc512-HuMD1-SO plasmids were double digested in 4 repetitions, with FseI and NotI restriction enzymes. The reaction mixtures were incubated at 37°C for 2 hours prior to heat inactivation at 80°C for 20 minutes. This was performed alongside 1µg internal controls consisting of: Uncut (enzyme substituted with water), linearised with SpeI-HF enzyme (this cuts once) and a double digest of FseI and NotI. Samples were resolved on 1% agarose gel against hyperladder I.

Pu57-hINT1-RT: Destination vector backbone: 6638bps

Spc512-HuMD1-SO: Provides the MD1 exon 6 onwards insert: 3213bps

These DNA fragments were visualised by white light as opposed to U.V, to prevent degradation. The above fragments were extracted by Qiaquick gel extraction and eluted in 30µl. If the 1st sample was observed to be of low yield below 10ng/µl, samples were pooled. The resultant mixtures were ligated at a variety of ratios, see section 1.6.1.5.

2.9.2 Microdystrophin Colony PCR

Primer Name	Primer Sequence 5'-3'	Amplicon Size (Bp)
HuDys CO Seq 1	GGGCCTGATCTGGAACATCATC	221bps (If MD1 is present)
nTTS HuDys F1 R2	CGAACAGGTCGGGTCTGTGGC	No amplification if MD1 was not successfully cloned.

These primers span the FseI 5' restriction site junction of the pUC57 hINT Repair template, initiating in the repair template backbone and extending into MD1 exon 6 onwards sequence. The amplicon will only be present if MD1 has been cloned successfully.

Colony PCR mixture was produced as described (see section: 15.1.1)

The cycling conditions: Denaturation 98°C 2 minutes, **35 cycles** – (98°C 30s, 65°C 30s, 72°C 45s), **Final Extension** 72°C 10 minutes and held at 4°C.

2.10 Tissue Culture

2.10.1 General Tissue Culture Reagents:

- **Sterile PBS:** 1 PBS Tablet (Gibco) dissolved in 500ml of ddH₂O. This was either autoclaved or filtered with a 0.22µm filter, with Class II Lamina flow hood.
- **Trypan blue** (Gibco)
- **Trypsin (10x)**- 0.5% trypsin in PBS (Sigma)
- **T175cm³, T75 cm³ Tissue Culture Flasks** (Corning)
- **Tissue Grade DMSO** (Sigma)
- **Sterile Glass Stripettes 5, 10 and 25ml (Starlabs)**
- **15ml and 25ml Falcon Tubes** (Corning)
- **Mr Freezie**
- **Screw Top Cryovials** (Corning)

2.10.1.1 Growth Media for Culturing HEK293T

- **Dulbecco's Modified Eagle's Medium (DMEM)** with 4.5 g/l glucose and glutamax (Gibco).
- **10% Heat Inactivated Foetal Calf Serum** (PAA / Gibco). Vol. 50ml. Heat inactivation of FCS is achieved by heating the 500ml bottle to 56°C for 30 minutes.
- **Penicillin-streptomycin** 100 Units/ml (Sigma) Vol. 5ml

2.10.1.2 Growth Media for Culturing Immortalised Patient myoblasts (Δ45-52)

- **Skeletal Muscle Cell Growth Media** (Promocell)
- **Skeletal Muscle Cell Growth Supplement Mix** (Promocell)
- **20% Heat Inactivated Foetal Calf Serum** (PAA / Gibco) Vol. 100ml

- **1.4 % L-Glutamine** (Sigma) Vol. 7ml
- **5% Gentamycin** (Sigma) Vol. 25ml

2.10.1.3 Growth Media for Culturing H2kb.mdx

- **Dulbecco's Modified Eagle's Medium (DMEM)** with 4.5 g/l glucose and glutamax (Gibco)
- **20% Heat Inactivated Foetal Calf Serum (PAA / Gibco)**. Heat inactivation of FCS is achieved by heating the 500ml bottle to 56°C for 30 minutes.
- **0.5% Chick Embryo Extract** (Sera Laboratories International) Vol. 2.5ml
- **0.2% Gamma Interferon** (Peprotech) Vol. 1ml
- **Penicillin-streptomycin** 100 Units/ml (Sigma) Vol. 5ml

2.10.1.4 Cell Lines:

Human Embryonic Kidney 293T (HEK293T) –Cells were obtained from ATCC. These were expanded and stored in liquid nitrogen. HEK293T used for the experimenters were usually in the range of passage 6 -16.

Cas9 expressing Human Embryonic Kidney 293T (HEK293T.Cas9) – Cells were attained from Gencopeia. This cell line harbours the stable integration of the bi-cistronic CBh-Cas9-Ef1a-eGFP cassette, at the AAVSI locus

Immortalised Patient Myoblasts Δ45-52 – These human myoblast cells were derived from patient biopsies bearing deletions of exons 45-52 inclusive. These cells were transduced with retroviral vectors expressing telomerase and cyclin-dependent kinase 4 (CDK4), to immortalise the cells (Mamchaoui et al. 2011). Upon successful transduction and stable integration, these vectors conferred the cells both puromycin and neomycin resistance. These myoblasts with greater proliferative capacity than conventional myoblasts provide a suitable in-vitro dystrophic model for the investigation of therapeutic approaches. These were kindly gifted to the lab by Dr. Vincent Mouley and maintained in a passage range of passage 9 – 17.

H2kb mdx: This mouse cell line was derived from the production of a transgenic mouse model, by crossing the H2kb-tsA58 background, with an *mdx* mouse (Morgan et al. 1994). The H2kb-tsA58 encodes a variant of the large T-antigen tsA58 downstream of a H2kb promoter, which upon induction with interferon gamma at 33°C serves to immortalise the cells. The *mdx* mouse contains a nonsense mutation in exon 23, serving to make the mouse dystrophin negative and exhibit a dystrophic pathology(Sicinski et al. 1989). Thus this immortalised dystrophic cell line represents an in-vitro mouse model for the investigation of therapeutic approaches. These cells were gifted by Dr Mike Antoniou, maintained in the passage of 13-19.

2.10.1.5 Thawing Cells

Cells were taken from liquid nitrogen, incubated in the 37°C water bath and thawed quickly. They were partially submerged in 37°C water and visually inspected until a 'rice size grain of ice' was visible normally after 2 -3 minutes.

HEK293T and H2kb.mdx: Once almost thawed the 1ml of cell suspension was aliquoted into 9mls of an appropriate pre-warmed growth media. This was then centrifuged at 1,000 rpm for 5 minutes (Beckmann, coulter at room temperature). The supernatant aspirated, the cell pellet re-suspended and transferred to a T175cm³ flask to make a total media volume of 25ml. Cells are then incubated: HEK293T - 37°C 5% CO², H2kb mdx - 33°C 5% CO².

Immortalised Patient myoblasts: Post thawing patient myoblasts are too fragile for centrifugation; rather the 1ml is applied directly to 24ml of warmed skeletal growth medium and transferred to a T175cm³ flask. This is then incubated for 24 hours and media is replaced the following day. Importantly, this process removes DMSO from the freezing solution, which may reduce cell viability.

2.10.1.6 Culturing Cell Lines

Cells were monitored daily until 70-80% confluency was attained after approximately 2-3 days of growth. The media was aspirated, the cells washed with 1X PBS, prior to the of 5ml addition of pre-warmed 1x Trypsin (*5x Trypsin for patient myoblasts*). This was then incubated at 37°C for approximately 2 minutes to liberate the cells from the flask surface, then neutralised with 10mls of appropriate growth media, and centrifuged in a 50ml falcon tube at 1,000 rpm for 5 minutes. Cells were re-suspended in an appropriate volume of fresh media, cells were then counted using a haemocytometer. This was repeated 3 times to attain an accurate cell number then cells were seeded at a density of 8x10⁵ per 25ml in a T175 flask.

2.10.1.7 Freezing Cell Lines

A 10% DMSO freezing solution was prepared for cells, comprised of: 9ml heat inactivated Fetal Calf Serum and 1ml of DMSO. A defined cell density of 1x10⁶ was centrifuged at 1000rpm for 5 minutes (Beckmann coulter) and pelleted in a sterile eppendorf. This pellet was re-suspended in 1ml of freezing solution and transferred to a screw-top cryovial. Once all cryovials prepared, these were placed in a Mr freezie, with a controlled cooling system or filled with absolute Isopropanol. This allowed the cells to gradually cool to -80°C by leaving overnight in a -80 freezer, prior to being transferred to liquid nitrogen for long term storage.

2.11 Transient Transfection Protocols

2.11.1 Cell Densities for 6 well Plate

Cell Line	Cell Density	Notes
HEK293T	5×10^5	70-80% confluency in 24 hours Careful media change undertaken, if evidence of de-attachment was observed. Cells were re-seeded.
HEK293T.Cas9	3×10^5	70-80% Confluency in 24 hours
Immortalised Patient Myoblasts Δ 45-55	8.3×10^4	60-70% Confluency in 48 hours. Low seeding was to avoid the occurrence of differentiation. Visually inspect cells to ensure this. Media change necessitated at 6 hours.
H2kb.mdx	1.25×10^5	70-80% Confluency is attained in 24 hours. The six well plates were coated with ECM. ECM Preparation: 9mls DMEM and 1ml of ECM. A 350 μ l aliquot of this suspension was applied per well. This was then incubated at 37°C for 1 hour. Then aspirated.

2.11.2 Transfection Reagents and Ratios:

Cell Line	Transfection Reagent and Ratio Used
HEK293T	Viafect 5:1 Lipofectamine Reagents: Messenger Max 3:1 CRISPR Max 3:1
HEK293T.Cas9	Viafect 5:1 Lipofectamine Reagents: Messenger Max 3:1
Immortalised Patient Myoblasts Δ 45-55	Viafect 3:1 Lipofectamine Reagents: 3000 3:1 Messenger Max 1.5:1
H2kb.mdx	Viafect 3:1

2.11.3 Viafect Transient Transfection

- **Viafect Transient Transfection Reagent** (Promega)
- **Serum Free Dulbecco's Modified Eagle's Medium (DMEM)** (Gibco)

Cells were seeded on a six well plate at Day 0 to attain 70-80% confluency by Day 1 or 2 as noted above. Once 70-80% confluency was reached, media was changed 1 hour prior to transient transfection. A master mix containing an experimentally defined transfection reagent volume (μ l): DNA mass (μ g) ratio was prepared. A 200 μ l volume transfection mixture was prepared for a single well, on a 6 well plate. The volume was scaled up proportionality for the number of repetitions required, with an additional 200 μ l volume in excess of that required prepared to account for pipetting error that may occur.

In the production of a single mastermix, a calculated volume of serum free DMEM at room temperature was pipetted into an eppendorf. Shortly, following this a defined DNA mass was added to the DMEM e.g. 4µg and the DNA-DMEM suspension agitated. This was incubated at room temperature for 5 minutes as per the manufacture's protocol. Then a defined volume of Viafect transfection reagent was added drop-wise with continual agitation of the suspension then incubated at room temperature for 15 minutes. Post incubation, the transient transfection mixture was added to the well in a dropwise circular motion to ensure maximum cell coverage.

Due to the non-toxic nature of Viafect transfection reagent a media change was not necessitated. Cells were routinely incubated for 48 hours if harvested for DNA and 72 hours if harvested for protein.

2.11.4 Lipofectamine Based Transient Transfection:

2.11.4.1 Lipofectamine Reagents

- Lipofectamine 3000 (Invitrogen)
- Lipofectamine mRNA Messenger Max (Invitrogen)
- Lipofectamine CRISPR Max (Invitrogen)
- Opti-Mem (Gibco)
- GFP mRNA positive control (Trilink)
- Cas9 mRNA (Trilink)

For lipofectamine based transfections, two separate stocks were produced in accordance with the manufacturer's protocol: i) A set mass of DNA or RNA was diluted in serum free DMEM with a transfection additive e.g. p3000 at 2µl/µg of DNA and ii) A volume of Lipofectamine transfection reagent required to produce the appropriate transfection reagent (µl): DNA (µg) ratio and was diluted in serum free DMEM. The appropriate ratios being either: 1.5:1 or 3:1, the latter being used more routinely. Both stocks were prepared to a final volume of 125µl per well, and the master mix was scaled for further repetitions and adjusted to include half a transfection volume in excess to account for pipetting error. The diluted DNA or RNA sample was then added to the transfection reagent mixture and incubated for 10 minutes at room temperature. This was then added in a circular dropwise fashion to the well. For patient myoblasts, which appeared susceptible to lipofectamine 3000 toxicity, the media was changed 6 hours post transfection.

2.11.4.2 Production of Ribonucleotide Proteins and Additional Considerations for Protein Transfections.

Prior to the transient transfection of Cas9 protein, the production of ribonucleotide protein (RNP) complex was required; this being the Cas9 protein complexed to an appropriate mRNA guide. In pursuit of this an appropriate concentration of guide mRNA was aliquoted to a fresh

ependorf tube and heated to 90°C for 2 minute. This was then subject to a 10 second pulse centrifugation of max speed and cooled for 5 minutes at 4°C. The Cas9 protein was then added to guide and it is incubated at room temperature for 10 minutes.

This step was performed once lipofectamine CRISPRMax mixtures had been prepared to ensure that the Cas9 RNP complex was not left for prolonged periods at room temperature. In addition, Lipofectamine CRISPRMax solutions were made using Opti-MEM; this was to prevent the Cas9 protein precipitating out of solution. Transfections with the CRISPRMax used at a 3:1 transfection reagent (µl) to Cas9 nuclease (µg) ratio. In this instance Cas9: guide complexes were transfected at either a 2:1 or 4:1 ratio: referring to 2.5µg Cas9: 1.25µg guide RNA and 2.5µg Cas9: 625µg of guide RNA per well respectively. Transfections were performed in 6 well plates when cells had achieved approximately 50% confluency (in line with the 30-70% indicated by the manufacturer's protocol), and DNA was harvested 24 hours later for assessment with mismatch detection assays.

2.12 Flow Cytometric Analysis

2.12.1 Materials

- 4% Paraformaldehyde (PFA) (Sigma)
- FACs Beads (BD Falcon)
- FACs Buffer: 500ml of PBS, 2ml of 0.5M EDTA and 5ml 10% NaN³
- 5ml round bottom snap cap FACs Tubes (BD Falcon)
- FACs Canto II (BD Biosciences)
- Clean and Rise Solutions (BD Falcon)
- FlowJo Software (BD Falcon)

2.12.2 Analysis of GFP Expression in Cell Culture:

Flow cytometric analysis was routinely used throughout this investigation to determine a subpopulation of cells expressing Green Fluorescent Protein (GFP). This was used as a proxy for transient transfection efficiency, when using GFP tagged plasmids: pCI-eGFP or U6-Guide-CBh-Cas9-2A-GFP (Px458).

In order to assess the proportion of transiently transfected cells, cell cultures were pelleted (see section 1.12.2). This process was performed using 5ml round bottom FACs snap tubes, rather than 15ml falcon tubes. Once the cells were pelleted, the media was removed and the cells were subject to a 500µl 1x PBS wash. Following which, 400µl of 2% PFA was applied. The cells were agitated to re-suspend and stored at 4°C in the cold room wrapped in foil for no longer than 1 week prior to analysis.

The FACs machine was calibrated with the FACs beads, in which two drops of FACs beads was suspended in 600µl of FACs buffer. The samples were analysed using the FACs Canto II, using

three graphical plots FSC/SSC (Live population), FSC-A/FSC-H (Single cells) and a FITC Histogram (GFP positive populations). Initially, mock samples were analysed to determine and refine gating voltages. Once a single dense population was centrally positioned on the FSC/SSC plot, this was gated to determine the number of live cells (gate p1). This population was then presented in the FSC-A/FSC-H graph, if the population was dense and approximately at a 45 angle these were gated to determine the number of single cells (gate p2). Finally, this sub-population of cells was displayed on an FITC histogram: cell count/FITC log. The voltages were adjusted so the peak population of the mock sample was positioned above 10^2 , as is convention with FACs analysis. Once gating was achieved, all samples were analysed on a medium flow rate, using 10,000 events in the P2 gate as an end point for each analysis.

2.12.3 Analysis of Flow Cytometric Data

Once the analysis had been performed, the outputs were exported and analysed with the use of FlowJo software. The samples would be gated once again; on the FITC histogram, mock samples would be gated to give approximately 1% GFP positive cells. If the proportion of cells in the FITC histogram gating was higher than this baseline in subsequent samples, they were deemed GFP positive. This was used to determine a percentage of GFP positive cells per transfection.

2.13 DNA/RNA Extraction and Quantification from Cell Culture:

2.13.1 Materials:

- 1X Sterile PBS
- 1X Trypsin
- Centrifuge
- Qiagen DNAeasy Kit
- Qiagen Qias shredder
- Qiagen RNeasy Kit
- Nanodrop ND-1000 spectrophotometer. (Labtech)

2.13.2 Cell Pelleting for Extraction

Cells were washed in 500 μ l of room temperature sterile 1X PBS and removed from 6 well plates with 500 μ l of 1X pre-warmed trypsin. This was incubated for 2 minutes at 37°C at 5% CO₂, the trypsin was then pipetted gently over the surface of the well to remove maximal numbers of cells. This trypsin was then neutralised with 2ml of 10% DMEM. The number of cells was quantified and a suitable suspension in the range of 1 – 5x10⁶ cells was then transferred to a 15ml falcon tube and centrifuged at 0.6g for 5 minutes. The cell pellet was washed with 2ml of 1x PBS and centrifuged once more. In the short-term if extraction was undertaken that day the cell pellet was stored at 4°C. If extraction was pursued at a later date, the pellet was stored at -80°C.

2.13.3 DNA Extraction

Pelleted cells within the suitable range were subject to Qiagen DNAeasy extraction, in accordance with the manufacturer's protocol. Resultant DNA was eluted in AE buffer: a 200 μ l was routinely used for HEK293T cell pellets and 100 μ l for Δ 45-55 Immortalised Patient Myoblast cell pellets.

2.13.4 RNA Extraction:

Pelleted cells within the suitable range were re-suspended in 350 μ l of RLT buffer and subject to Qiagen Qias shredder; this served to homogenise and purify the crude mRNA lysate. The resultant eluate was then subject to an RNA extraction using a Qiagen RNAeasy kit.

2.13.5 Nucleic Acid Quantification

A 1 μ l aliquot of DNA or RNA preparation was pipetted onto the nanodrop. The amount of light absorbed by this solution at a wavelength of 260nm is converted to estimates of DNA or RNA concentration and purity. The ratios of 230:260 and 260:280 were used as provisional indicators of DNA purity; if ratios obtained were in the range of 1.8-2.0 and 2.0-2.2 then the DNA sample was deemed pure. Notably, pure RNA samples generally absorb slightly higher at 2.0 and 2.2. If uncertainty about the purity of the sample persisted, then a 2 μ l aliquot was resolved by gel electrophoresis.

2.14 Protein Extraction and Quantification:

2.14.1 Materials

- **Sterile PBS:** 1 PBS Tablet (Gibco) dissolved in 500ml of ddH₂O. This was either autoclaved or filtered with a 0.22 μ M filter, with Class II Lamina flow hood.
- **PAPBNI RIPA Buffer:** NaCl 0.15M, HEPES 0.05M, NP-40.1%, Sodium Deoxycholate (SOC) 0.5%, SDS 0.10%, EDTA 0.01M, Protease Inhibitor tablet 1 in 50ml (Roche). This was aliquoted into 5mls, and stored at -20.
- **Cells Scrappers** (Invitrogen)
- **Eppendorfs**
- **Benchtop Microcentrifuge**
- **DC Assay Protein kit:** Reagent A, S and B (BioRad).
- **2 μ g BSA Standard** (Invitrogen)
- **96 Well plate** (Corning)
- **96 Well plate reader** (Genbank)

2.14.2 Protein Extraction:

Cells were washed with 500 μ l ice cold PBS, prior to the addition of 100 μ l PABPN1 RIPA Buffer. The cells were then scrapped down to the bottom of the well, with the plate held at a 45° angle, before being incubated at 4°C on ice for 5 minutes. The sample was then transferred to pre-chilled and labelled eppendorfs, prior to being vortexed every 30 seconds for a further 15 minutes. The resultant protein lysates were then centrifuged at 13,000 rpm for 15 minutes at 4°C, to allow cell debris to pellet. The supernatant was then transferred to a fresh pre-chilled and labelled 0.5ml screw top tube, and stored at -20°C.

2.14.3 Protein Quantification:

A defined dilution series of BSA in the protein extraction PABPN1 Buffer was prepared providing concentrations ranging from 0- 2 μ g of BSA respectively as shown below:

Conc (μ g)	2	1.8	1.5	1.2	1	0.8	0.6	0.4	0.2	0
RIPA buffer μ l	0	4	10	16	20	24	28	32	36	40
BSA μ l	40	36	30	24	20	16	12	8	4	0

These protein standards were loaded at a volume of 0.5 μ l, alongside extracted protein samples at a volume of 5 μ l in a 96 well plate. All samples were loaded in triplicate to assert the accuracy/ ensure reliability of the resulting absorption readings obtained.

In the fume hood, 25 μ l Reagent A +S (1ml of A to 20 μ l of S) was added to each of the samples from a low to high concentration wherever possible; this being a precautionary measure to prevent contamination. Subsequently, 200 μ l of Reagent B is applied and the resultant mixture is agitated and incubated for 15 minutes at room temperature. A colorimetric analysis is undertaken at 750nm using the (Gen) 96 well plate reader. The absorption readings were then

used to calculate an average protein concentration of the three samples from the standard curve.

2.15 Western Blotting

2.15.1 Materials

- NuPage 10X Reducing Agent (ThermoFisher)
- NuPage 4x Loading Dye Sample (ThermoFisher)
- NuPage 3-8% Tris Acetate precast gradient gels (ThermoFisher)
- NuPage Antioxidant (ThermoFisher)
- Prestained HiMark Ladder (Life technologies)
- NuPage 3-8% Tris Acetate Running Buffer (ThermoFisher)
- NuPage 20X Transfer Buffer (ThermoFisher)
- Absolute Methanol (VWR)
- X-Cell sure lock mini cell and Tank (Thermo Fisher)
- 0.45 μ M Nitrocellulose membrane (GE Healthcare)
- Ponceau Stain (ThermoFisher)
- Filter paper
- Marvel Milk Powder
- Tween 20 Detergent (Sigma)
- ECL solution 1 and 2 (Promega)
- Amersham Hyperfilm 18cm x 24cm (GE Healthcare)

2.15.2 Antibodies

2.15.2.1 Primary Antibodies

Antibody name	Raised in	Dilution Used	Binds to
6C5 (Dr. Glenn Morris)	Mouse	1 in 100	17a.a C-Terminal
MannEx 1011C (Dr. Glenn Morris)	Mouse	1 in 100	Hinge/Spectrin repeats (Exon 10-11)
Cas9 (Diagenode)	Mouse	1 in 5000	Cas9
Alpha-Tubulin (Abcam 40774)	Rabbit	1 in 2500	Alpha Tubulin Subunit

2.15.2.2 Secondary Antibodies

Antibody name	Dilution Used
Goat α Mouse HRP (Dako)	1 in 2000
Goat α Rabbit HRP (Dako)	1 in 2000
Goat α Mouse (Green Fluorescence) (LI-COR)	1 in 10000
Donkey α Rabbit (Red Fluorescence) (LI-COR)	1 in 10000

2.15.3 Sample Preparation

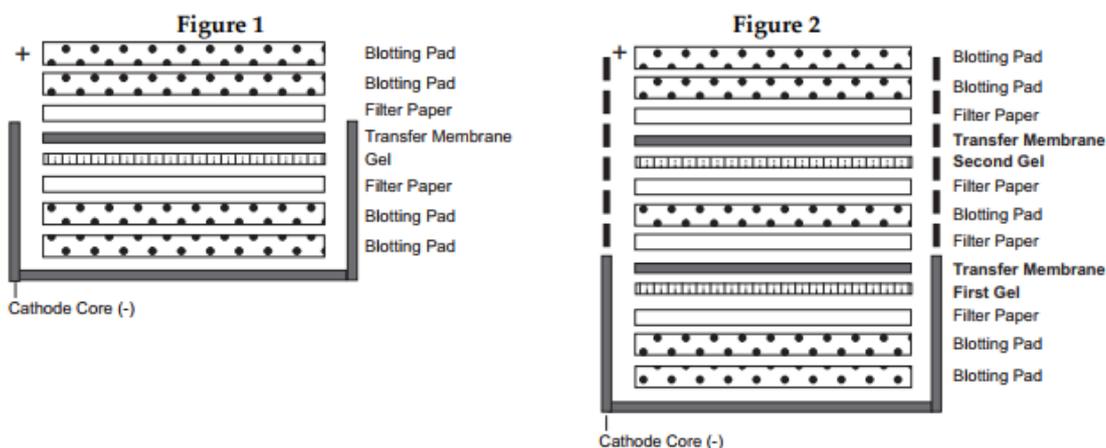
Samples of total protein lysate were produced in a 4x master mix, this was to allow repetitions with antibodies if required. Typical a (4x) 40µl master stock would contain: 200µg total protein (Dystrophin screening), 50µg total protein (Cas9 screening). Samples were then prepared in a 1.5ml screw top tube with: 4µl Reducing Agent, 10µl of Loading Sample Dye and the remaining volume is supplemented with ddH₂O. The samples alongside a positive control, either dystrophin extracted from muscle or from a previous positive transfection were denatured by heating to 70°C for 10 minutes. Samples were then cooled on ice.

2.15.4 Gel Preparation and Electrophoresis

A 3-8% Tris-Acetate precast gradient gel was used to resolve dystrophin and cas9 proteins. In preparation of the gel, the comb was removed and wells washed with ddH₂O. In addition, a white adhesive strip sealing the foot of the gel was removed. Gels were then placed in the X-Cell Blot tanks vertically. Then a 10µl sample of the 4x master stock of each protein sample was loaded alongside, a pre-stained Hi-Mark ladder. The surrounding tank was filled approximately 1cm from the top with 1X Tris-Acetate buffer and 500µl of antioxidant was applied immediately prior to the initiation of electrophoresis. The gel was run for approximately 1 hour and 15 minutes at 150V, in accordance with the Nupage technical guide. The blue loading dye reaching the 'foot' of the gel and the ladder separation were parameters by which sufficient separation was assessed. During this time blotting pads were soaked in 1x transfer buffer supplemented with 10% methanol and 1ml antioxidant. Filter paper and 0.45µM nitrocellulose membranes were cut to correct size for the transfer.

2.15.5 Electro-transfer to Nitrocellulose Membrane.

Upon suitable separation of the ladder and by extension the proteins, the Nupage Electro-transfer cassette was prepared in accordance with the protocol: Diagram included below (Taken from NuPage Technical guide 2013):



Once pre-soaked blotting pads were applied to the bottom of the electro-transfer cassette, the 3-8% Tris acetate gels cases were 'cracked' open to liberate the polyacrylamide gels. The top of the gel, above the top band of the Hi-Mark ladder was removed and disregarded. The remainder of the gel was then floated using the buffer to be situated above the filter paper, lifted out of the transfer buffer and transferred to the cassette. Once performed, the 0.45µm nitrocellulose membrane was submerged in buffer and placed on top. This was then rolled across the surface of the gel using a plastic roller to ensure tight contact throughout. A filter paper and a blotting pad placed on top. The whole cassette was kept wet during this time. If a second gel was present then the process was repeated. When completed this was placed in the X-Cell tank. The top of the electro-transfer cassette was refilled with buffer and the surrounding area filled with cold ddH₂O. The proteins were then transferred for 2 hours at 30V.

2.15.6 Post-Transfer Nitrocellulose Membrane Checks and Blocking

Following the two hour transfer, the membrane is stained with 1X Ponceau. This stains all proteins across the lanes and is used to ensure that the transfer was complete and successful. The stain was then washed off with 0.1% PBS-T, washing at 5 minute intervals until no stain was visible.

The nitrocellulose membrane was then blocked with 5% (w/v) Marvel milk in 0.1% (v/v) PBS-Tween for 1 hour at room temperature, to prevent non-specific binding. Once the membrane was blocked, the membrane was incubated at 4°C shaking overnight in a suitable dilution of primary antibody.

2.15.7 Visualisation of the Nitrocellulose Membrane: The ECL Method

The nitrocellulose membrane was then subject to 4 washes with 0.1% PBS-T for 5 minutes and incubated with a conjugated HRP antibody diluted to 1 in 2000 in 5% milk, made with 0.1% PBS-T at room temperature for 1 hour, whilst gently shaking at 60 rpms.

Post incubation the membrane was washed a further 4x for 5 minutes in 0.1% PBS-T. Once clean, 3mls of ECL solution 1 and 2 were mixed. Immediately on mixing the solution was applied to the nitrocellulose membrane. This was incubated for 1 minute at room temperature, blotted dry and transferred to an X-ray cassette for development. In order to visualise the blot, the film was placed into contact with the membrane for defined timed intervals, e.g. 30 seconds, 1 minute and then placed through the developer, until a blot was observed. Cas9 protein western blots were developed for 3 seconds.

2.15.8 Visualisation of the Nitrocellulose Membrane: The Odyssey Method

This method also requires 4 washes for 5 minutes. In this instance however, the secondary antibody is conjugated to a fluorescent label that requires a 1 in 10,000 dilution prior to the 1 hour incubation. This was then scanned at 700nm and 800nm channels respectively. Ordinarily, the protein of interest would be present within the 800nm channel, such as dystrophin and the loading control which is by nature more abundant within the 700nm, this being alpha tubulin.

2.16 Quantification of Dystrophin

2.16.1 Software

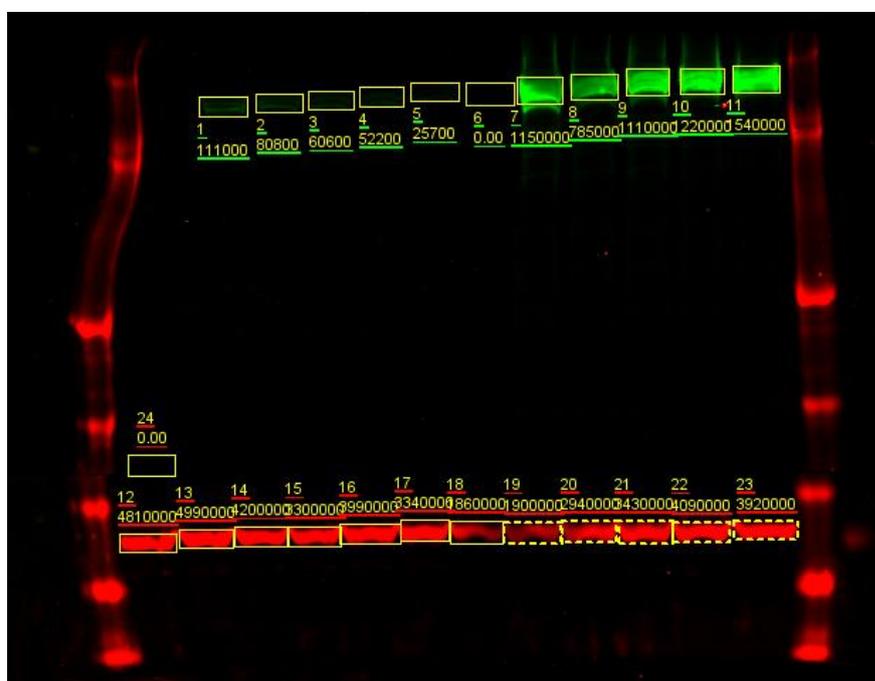
- Image Studios Version 4 (Li-Cor)

2.16.2 Methods

The nitrocellulose membrane was then visualised with Image Studios Version 4. Bands were automatically identified using the software, and adjusted to be tight to the band in question in individual 800nm and 700nm channels. User defined noise values were subtracted away from the band in question and used to attain intensity values, for both dystrophin (800nm) and α -tubulin (700nm). An image is included for reference below. The values were normalised to tubulin:

$$\frac{\text{Dystrophin Fluorescence Intensity}}{\alpha - \text{Tubulin Fluorescence Intensity}} = \text{Normalised Values}$$

This was performed for each lane and the native and optimised dystrophin (n=5) and a mean value was attained. Once mean values were attained the Sequence optimised dys/tub ratio was normalised to the native dys/tub, to establish a fold difference in expression.



2.17 Polymerase Chain Reactions (PCR)

A standard three stage polymerase chain reaction: Denaturation, Primer annealing and Extension, was undertaken throughout the course of this investigation. The general materials and methods are outlined here. Specific amplification reactions will be outlined in their respective sections.

2.17.1 Materials

- 2X PCR Master mix (GeneSys Ltd)
- GoTaq G2 Flex (Promega)
- Hotstart Phusion Flex (NEB)
- Thermocycler (VWR)
- Primers (Eurofins MWG /IDT)
- dNTP mix 100mM (Bioline)

2.17.1.1 General Polymerase Chain Reaction Methodology:

All primers were bought from Eurofins MWG and or IDT and diluted to 100pmol/ μ l and subject to a further 1 in 10 dilution to attain a stock [10 μ M], prior to being used in standard PCR reactions. All PCR amplification reactions were supplemented with 250ng genomic DNA or 100ng of plasmid DNA. In addition, the annealing temperatures were determined experimentally for individual primer pairs using temperature gradients and the extension times were calculated in accordance with the polymerase manufacturer's protocol.

2.17.1.2 Colony PCR with the 2x PCR Mastermix (GeneSys Ltd)

This master mix contains all necessary constituents for a PCR: Magnesium, free nucleotides, Taq polymerase, Red Buffer/loading dye. All constituents are in propriety amounts and require only a 2-fold dilution, to attain suitable concentrations for amplification. A standard PCR reaction comprises: 12.5 μ l master mix, 1 μ l of forward and reverse primer and 9.5 μ l of ddH₂O to attain a final volume of 25 μ l.

2.17.1.3 Amplification Using GoTaq G2 Flex

This polymerase had relatively low fidelity, and was used for general amplification procedures. The standard PCR reaction comprised: 1.5mM Magnesium Chloride, 0.2 μ M forward and reverse primers, 0.2mM dNTP mix, and 0.05 u/ μ l Gotaq G2 Flexi polymerase. This equated to 1.5ml Magnesium chloride, 0.5 μ l forward and reverse primers, 0.5 μ l dNTP mix and 0.25 μ l of GoTaq G2 flex in a 25 μ l reaction.

2.17.1.4 Amplification Using HotStart Phusion Flex

This polymerase has approximately 50x higher fidelity than Taq polymerase, thus it was used for amplification for mismatch detection assays. The 5x high fidelity buffer was already supplemented with 1.5mM Magnesium. Thus the PCR reaction comprised: 0.5 μ M of forward and reverse primers, 0.2mM free dNTPs and 0.02u/ μ l of Hotstart phusion polymerase. These

concentrations were often achieved in a 50µl total reaction volume, these reactions were produced in triplicate, allowing for purification and enrichment processes to be undertaken.

2.18 Mismatch Detection Assays:

Mismatch detection assays were used to analyse amplicons spanning the target site for the presence of small Insertions or Deletions (InDels). The presence of InDels indicates that the NHEJ DNA pathway has been active and thus can be used as a proxy for gene modification. As such DNA extracted from treated cell cultures, progress through 4 stages:

- Asymmetric amplification of the genomic locus of interest.
- Hybridisation of the resultant heterogeneous amplification, containing: homogeneous W.T amplicons – absence of InDels and heterogeneous modified amplicons – InDels are present – heteroduplexes are present.
- Incubated with Surveyor / T7E1 enzymes to cleave heteroduplexes.
- Resolve samples by agarose gel electrophoresis or Poly Acrylamide Gel Electrophoresis (PAGE)

2.18.1 Materials

- Surveyor Nuclease 100 reaction Kit (Transgenomics)
- T7 Endonuclease I (NEB)

2.18.2 Human Intron 1 High Fidelity PCR:

2.18.2.1 Primers for Assessing Guides 1,2 and 4

Primer Name	Primer Sequence 5'-3'	Amplicon Size (Bp)
hINT1 1960F	CTTAACCATTTTGAGGGGTAAGTTT	512bps
hINT1 2471R	CTTGATCTTTGAATGAAGTTGGTCT	

The Human Intron 1 amplification was performed in triplicate 50µl reactions with HotStart Phusion Flex (see section 15.1.4).

The cycling conditions: Denaturation 98°C 30 s, **35 cycles** – (98°C 10s, 63°C 30s, 72°C 24s), **Final Extension** 72°C 10 minutes and held at 4°C.

2.18.2.2 Primers for Assessing Guide 3

Primer Name	Primer Sequence 5'-3'	Amplicon Size (Bp)
hINT1 1960F	CTTAACCATTTTGAGGGGTAAGTTT	582bps
hINT1 2541R	CTTTGTTGAGTACTGCACTTCCAGTC	

The cycling conditions: Denaturation 98°C 30 s, **35 cycles** – (98°C 10s, 63°C 30s, 72°C 24s), **Final Extension** 72°C 10 minutes and held at 4°C.

2.18.3 Mouse Intron 1 High Fidelity PCR:

2.18.3.1 Primers for Assessing Mouse Guides 1 -5

Primer Name	Primer Sequence 5'-3'	Amplicon Size (Bp)
hINT1 1950F	CTTAACCATTTTGAGGGGTAAGTTT	556bps
hINT1 2775R	CTTGATCTTTGAATGAAGTTGGTCT	

The cycling conditions: Denaturation 98°C 30 s, **35 cycles** – (98°C 10s, 60°C 30s, 72°C 24s), **Final Extension** 72°C 10 minutes and held at 4°C.

2.18.4 Surveyor Assay

Cell cultures were transiently transfected with U6-driven guides and Cas9 nuclease, and incubated for 48 hours prior to a DNA extraction. The DNA samples were amplified from both treated samples and cells incubated in the presence of transfection reagent alone, in a 50 μ l PCR reaction using primers asymmetric with regard to the target site. The PCR amplicons were then hybridised using the following transgenomic 'surveyor hybridisation programme' protocol:

95 °C for 10 min then 95 °C to 85 °C (-2.0 °C/s)

85 °C for 1 min then 85 °C to 75 °C (-0.3 °C/s)

75 °C for 1 min then 75 °C to 65 °C (-0.3 °C/s)

65 °C for 1 min then 65 °C to 55 °C (-0.3 °C/s)

55 °C for 1 min then 55 °C to 45 °C (-0.3 °C/s)

45 °C for 1 min then 45 °C to 35 °C (-0.3 °C/s)

35 °C for 1 min then 35 °C to 25 °C (-0.3 °C/s)

25 °C for 1 min and finally hold at 4 °C.

Once hybridised 6 μ l, 12 μ l and 18 μ l aliquots of samples were pipetted into PCR tubes and supplemented with 1 μ l enhancer and 1 or 2 μ l of surveyor nuclease. These were then incubated at 42°C for 1 hour and stopped with 0.9, 1.5 and 2.2 μ l of the stop solution respectively.

2.18.5 The T7 Endonuclease I

Amplicons of transiently transfect samples as described in 1.16.2, were more routinely assayed with T7 Endonuclease I. This is an alternative to the Surveyor enzyme with the propensity to cleave heteroduplexes formed upon hybridisation.

The human *DMD* intron 1 PCR was undertaken in triplicate and the three resultant 50 μ l reactions pooled. The total 150 μ l PCR volume was then subject to Qiaquick PCR purification, in accordance with the manufacturer's protocol and eluted in 50 μ l of ddH₂O. The resulting concentration of the amplicon was quantified with the use of the nanodrop.

Once quantified a standard digestion in a PCR tube comprised: 2 μ l of NEB Buffer 2, 400ng of sample amplicon and ddH₂O to attain a final volume of 19 μ l. Samples were then subject to the 'surveyor hybridisation' programme.

Upon completion of the hybridisation programme, 1 μ l of T7 Endonuclease I was added and the reaction was incubated at 37°C for 30 minutes. This was then stopped with 2.5 μ l of 1.5M EDTA.

2.19 Indel Quantification Approaches

2.19.1 Gel Based Quantification Methods:

2.19.1.1 Materials

- 10% TBE Polyacrylamide gels (ThermoFisher)
- 1x TBE buffer. 89 mM Trizma base, 89 mM boric acid, 2 mM EDTA pH 8.0.

The total volume of digested sample (or 17µl volume remaining after the removal of 3µl for Agilent analysis) was either resolved by standard agarose gel electrophoresis (see section 1.6.1.3), or by PAGE gel electrophoresis.

In the case of precast 10% TBE polyacrylamide gels; the gel was prepared, by removing the comb and washing wells with ddH₂O. In addition, a white adhesive strip sealing the foot of the gel was removed. This was then carefully loaded with T7E1 digested amplicons. The gel was then resolved at 200V until the loading dye reached the foot of the gel after 30- 45minutes.

Once resolved the gel would be extracted from the pre-cast cassette and incubated with 1x TBE buffer supplemented with 0.02% SYBR Safe, in the dark for 1 hour. This was then rinsed with 1X TBE and visualised under U.V. If successful, the image captured would then be analysed with the use of Image J software.

2.19.1.2 Image J analysis

Gel images of samples resolved following analysis with the mismatch detection assay, were opened with Image J software. This software uses boxes positioned to identify bands, using the number of pixels to provide a histogram plot. To allow for a more accurate determination of gene modification boxes were drawn the full length of the lane. The resultant histogram showed peaks for uncut amplicon, cleavage product 1 and 2, as well as for any residual noise. Lines were drawn under the base of each peak and gene modification was calculated using the following formula:

$$\% \text{ Gene Modification} = (1 - (1 - \text{Fraction cleaved})^{0.5}) * 100$$

Equation 1: An equation used to determine gene modification at a particular locus.

Where the fraction cleaved is calculated by:

$$\text{Fraction cleaved} = \frac{\text{Total Cleaved products (Indicating Heteroduplex Density)}}{\text{Total Products: The sum of the unedited amplicon and cleaved products}}$$

Equation 2: An equation used to determine the fraction cleaved in order to determine the percentage of gene modification.

This calculation is performed to account for the fact that any in-situ hybridisation resulting in heteroduplex formation, comprised of one wild type strand and one mutated strand. The above system of quantification was previously presented by (Guschin et al. 2010), it is this method of quantification this study followed.

2.19.2 Agilent Bioanalyser Analysis

- Agilent Bioanalyser
- DNA High Sensitivity Chips

2.19.2.1 Preparation of DNA Dye-Gel Mixture

To prepare Gel-Dye mix for the Agilent chip, the DNA dye concentrate and DNA gel matrix were allowed to equilibrate to room temperature for 30 minutes. The DNA dye was subject to a 10s pulse vortex at maximum speed using a bench-top microcentrifuge. A 25 μ l volume of the DNA dye mix was then aliquoted into the DNA gel matrix vial, and vortexed. This was then spin-filtered from 15 minutes at 6,000rpm. If this DNA Dye-gel mix was to be stored, it would be wrapped in foil to reduce light exposure and stored at 4°C.

2.19.2.2 Loading the Agilent Chip

If using pre-prepared DNA dye-gel mix, this was allowed to equilibrate at room temperature for 30 minutes. Once at room temperature, the high sensitivity DNA chip was placed in the priming station and 9 μ l of this mixture was then pipetted into the bottom –right well labelled with a white G encased in a black circle. A fresh syringe was placed in the priming station, which was set at the lowest priming position and 60 seconds was put on a timer. This was then depressed for 60 seconds, until it was held in place by the metal clip; allowing the DNA-gel mix to travel through the capillaries of the chip. Once the 60 seconds had passed, the metal clip was released, and the plunger left for 5 seconds before being pulled back to the 1ml position.

The primer station was then opened and 9 μ l of the DNA Dye-gel mixture pipetted in the middle and top right wells labelled with black Gs . Following this 1 μ l of the yellow capped ladder was placed into the bottom right of the chip which has a ladder symbol. Then 5 μ l of DNA marker was pipetted into all remaining wells. Then 1 μ l T7 Endonuclease I enzyme mismatch assayed samples were loaded into the wells. No wells were left empty, if there were insufficient samples the well was loaded with 1 μ l ddH₂O. The chip was then vortexed for 60 seconds at 2400rpm and read within 5 minutes.



2.19.2.3 Running Samples

The chip was then removed carefully, from the vortex and placed in the Agilent receptacle. The lid was carefully closed, and a High Sensitivity DNA analysis was selected from the menu. The programme was started and progressed in an automated fashion. Once the analysis was complete electrodes were cleaned with 350µl of ddH₂O in a designated chip and air dried for 10 seconds. The results for each sample was shown in the form of a histogram; the area under each peak identified (uncut, cleaved product 1 and 2) being proportional to concentration of DNA fragment present. The histograms were therefore analysed and exported alongside pseudo-gel images.

2.19.3 Tracking of Indels by DEcomposition (TIDE)

Finally, computational software was used to analyse the presence of InDels in an amplicon. This was reliant upon sequencing of amplicons using 2µl of hINT1 1960F primer (as described in section 1.7.4.2). The resultant ABI or SCF files from Eurofins MWG were downloaded for wild-type and treated samples. Multiple mock and treated sequence traces were aligned using the TIDE web tool, and the mean decomposition of the trace detected was used as a proxy for gene modification (Brinkman et al. 2014).

This alignment was performed under default parameters of: Alignment window left boundary 100bps upstream of break point and right -10 bps. Decomposition window left boundary max INDEL and 5bps downstream from breakpoint, right boundary 5bps upstream of shortest amplicon. InDel size range 10bps. p- Threshold value: p<0.001.

2.20 In Vitro Transcription of mRNA guides

In order to produce single chimeric RNA guides with the ability to associate with Cas9 and mediate DNA cleavage, the following primers were ordered from IDT, USA.

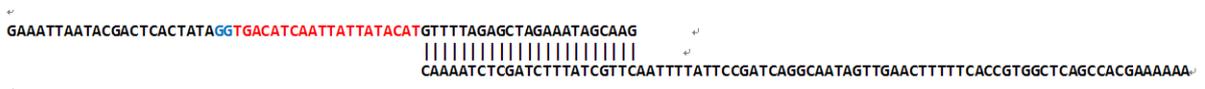
Forward Extended DNA Primer Guide 1:

GAAATTAATACGACTCACTATAGG (AGACTAAGAACCGATGCTA) GTT TTA GAG CTA GAA ATA
GCA AG

Reverse Invariant Extended DNA Primer:

AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTA GCCTTATTTTAA CTT GCT
ATT TCT AGC TCT AAA AC

Figure 1: Hybridisation of Extended DNA primers:



The forward extended DNA primer is guide specific, whereas the reverse is common to all RNA guide production; hence these were ordered as 4nM and 20nM customer ultramer primers with standard desalting. For subsequent reactions, they were re-suspended in DEPC treated water to a concentration 100pmol/μl, in 40 and 200μl respectively.

The primers are subject to a standard 3 step PCR, where a 5' region of the tracrRNA proximal to the crRNA (20nt guide without PAM) hybridises to one another. During the amplification process a linear double stranded PCR amplicon with a functional T7 promoter is produced, this is highlighted in blue. This amplicon can subsequently be processed for In vitro RNA transcription.

2.20.1 PCR Amplification and Purification of Extended Guide Oligonucleotides:

2.20.1.1 Materials

- Hot Start Phusion Flex (NEB)
- DEPC Treated Water
- 10mM dNTP (Bioline)
- Forward and Reverse oligonucleotides (IDT)
- Qiaquick PCR purification (Qiagen)

In order to produce a double stranded DNA amplicon, the following 50μl reaction mixture was produced in triplicate: 36μl DEPC treated water, 5μl HotStart Phusion Buffer, 10mM dNTP 2.5μl, 0.5μl of forward and reverse oligonucleotides at F 100pmol/μl, and 0.5μl HotStart Phusion Flex DNA polymerase.

Denaturation 98°C 30s, **35 cycles** – (98°C 10s, 54 °C 20s, 72°C 30s), **Final Extension** 72°C 5 minutes and held at 4°C

Post amplification the 50μl PCR products were pooled into an Eppendorf, giving a total volume of 150μl. This was then purified using a QiaQuick PCR purification kit, in line with manufacturer's protocols. The final elution volume of 50μl DEPC water was then nanodropped and to ensure the concentration was in the range of 150 – 200ng/μl.

2.21 In Vitro RNA Transcription of Guides and Clean-up:

2.21.1.1 Materials

- 100mM free NTPs [ATP, CTP, GTP and UTP] (NEB, UK)
- T7 RNA Polymerase (NEB)
- Murine RNase inhibitor (NEB)
- DNase (NEB or Quiagen)

Once the PCR amplicon has been purified, the following reaction should be set up at room temperature: 18.5μl DEPC-treated water, 28μl 50mM MgCl₂, 4μl of 100mM each NTP [ATP,

CTP, GTP and UTP], 10µl T7RNA buffer (NEB,UK), 10µl T7RNAP, 2.5µl RNase Inhibitor, 15µl PCR amplicon [100ng/µl]. This would then be incubated at 37°C for 16 hours, prior to being maintained at 4°C until the sample can be processed.

The RNA would then be DNase treated with 10µl 10x DNase Buffer and 1µl of DNase (NEB, UK) at 37°C for 30 minutes.

Optional isopropanol precipitation step: the reaction mixture was treated with 2.5x isopropanol absolute and incubate at -20°C and then centrifuge at 13,000 rpm for 10 minutes, aspirate the supernatant and allow to air dry. This was omitted in general production of mRNA.

Then 500µl of PB Buffer was added, the RNA pellet was re-suspended and work progressed to the QIAquick PCR Purification Kit, in accordance with the manufacturer's protocols. Once completed RNA was eluted in 30µl of DEPC-treated water and samples stored at -80°C.

2.21.1.2 Functional Validation of RNA Guides:

To assert that the chimeric mRNA guide produced is functional, 2.5µg of mRNA was transiently transfected into HEK293T.Cas9 cells via a 3:1 Lipofectamine Messenger Max reagent to DNA ratio (See section 1.10 and 1.10.4). Following a 48 hour incubation, DNA was harvested via the DNAeasy Blood and Tissue Kit, amplified and subjected to T7E1 mismatch detection assay (section 1.16.4).

2.22 In Vitro Transcription of Cas9

2.22.1 Materials

- Qiaquick PCR purification
- T7 Ribomax Large Scale RNA production system (Promega, UK)
- RNAeasy Mini-prep Kit (Qiagen)

Cas9 mRNA was produced using a T7 Ribomax Large Scale RNA production system. In this instance 10ug of Endofree Maxi prepared HC-Cas9 plasmid was linearized in triplicate for 10 hours at 37°C with SpeI-HF (NEB, UK).

The digest reaction comprised: 7µl DEPC-treated water, 2µl Cutsmart buffer, 1µl of Enzyme SpeI-HF and 10µl of DNA [1ug/µl]. Subsequent to this the reaction was maintained at 4°C then purified using QiaQuick PCR Purification, in accordance with the manufacturer's protocols. Once completed the digest was eluted in 50µl of DEPC-treated water.

Post purification, 5µl of mixture loaded with 1µl 5x Loading Dye onto a 1 (w/v) % agarose gel and visualised via gel electrophoresis using SyBr safe. This was to ensure the plasmid had been linearized and that no non-specific star activity had occurred. If the digest was suitable, an RNA In Vitro transcription reaction was set up consisting of: 40µl of DEPC-treated water, 20µl of T7

Transcription 5x Buffer, 5µl each of individual NTPs [100mM ATP, CTP, GTP and UTP], 10µl Linear DNA template [1ug/µl], and 10µl of T7 Enzyme Mix This was incubated at 30°C for 3.5 – 4 hours.

Following the transcription incubation the reaction was treated with RQ1 RNase-Free DNase, which was added 1 unit per µg of template DNA; this was then incubated for 15 minutes at 37°C. Following the removal of any residual DNA, the reaction mixture was then applied to an RNeasy Mini Kit. The process was carried out in line with the manufacturer's protocol. Finally, the RNA was obtained in two sequential elutions of 50µl of DEPC water.

The quality of the RNA produced was examined by Gel Electrophoresis, by loading 2µl of RNA with 2µl of 5x loading dye (Bioline) onto a 2.5% Agarose gel as outlined by the Ribomax protocol run alongside hyperladder III. Provided there was no evidence of RNA degradation, the RNA was quantified via nanodrop and frozen at -80C in aliquots to prevent repeat freeze thawing cycles.

2.23 Cas9 Protein Production

2.23.1 Materials (Work Undertaken in Paris):

- All Salts were obtained from Sigma Aldrich
- Protease Ultra 1x Inhibitor (Roche)
- FPLC Akta Purifier (GE Healthcare)
- Instant Blue Dye (Expediton)
- Tris (2-carboxyethyl)phosphine hydrochloride solution (ThermoFischer)
- Acrylamide (Euromedex)
- Protein Ultra 15 concentrator spin columns (Amicon)
- TEV protease (Made in the lab)
- HisTrap Column (GE Healthcare)
- 12-14kDa Dialysis Membrane (Spectropore)
- Sp Sepharose (GE Healthcare)
- Superdex 200 (GE Healthcare)
- Centrifuge (Hettich rotanta 460R)
- Ultracentrifuge (70TI Beckmann)

Heat shock transformation of BL21 Rosetta 2 competent *E.coli*, with 10-100ng pET-Cas9 protein encoding plasmid was undertaken by Dr. DeCian. The resultant suspension was plated onto an agar plate supplemented with 50µg/µl kanamycin and grown overnight at 37°C to obtain single colonies. Once obtained a few colonies were inoculated into a 200ml starter culture of 2XYT media, with 50µg/µl kanamycin and 2% glucose. This was incubated at 30°C overnight, prior to the OD600 being taken, a more substantial 2L culture of the 2XYT medium was inoculated with the starter culture until an initial OD600 was obtained. This was then grown at 37°C until an OD of 0.7 was achieved, this culture was then cooled in the cold room for 20 minutes induced with IPTG to a concentration of 0.2mM prior to being incubated for 3 days at 16°C.

The sequential processes of bacterial lysis and Cas9 protein purification were undertaken with the support of Dr. DeCian. After the 72hour incubation, the bacteria were spun down to produce a pellet of approximately 15g. The pellet was washed in 1x PBS and re-suspended in 80ml of lysis buffer comprising: 25mM Tris-oxaloacetate pH 8, NaCl 500mM, Imidazole 20mM, Protease Inhibitor, 1mM TCEP, filtered using a 0.22µm vacuum filter. This suspension was then sonicated for 15 minutes, using a 1s on, 1s off on ice, prior to spinning once more. The resultant lysate was then loaded onto a pre-equilibrated HisCrude FF 5ml column, overnight with the use of the Akta, and a 100ml wash is undertaken after the full lysate is loaded. Post loading the his-tagged Cas9 protein would be immobilised by its affinity for nickel, present in the column and protein contaminants would have been washed through. To liberate the Cas9 protein of interest an elution buffer with the same constituents of the lysate buffer but an increase in imidazole to 250mM was loaded to the column, the Imidazole out competes the

Cas9 for the nickel moieties thereby eluting the Cas9 protein. Normally elution occurs over 8 column volumes, 40mls was used to elute the Cas9 protein in 3ml sequential fractions from the point of elution buffer entry into the column. A 2 μ l aliquot of each of the fractions was visualised on 8% SDS-PAGE and visualized with instant blue. This combined with absorption peaks at 260nm and 280nm, was used to ascertain which column fractions the Cas9 protein was present to be pooled and enriched

The protein eluted from the crude HisTrap column was concentrated using 100kDa amicon inserts into a 50ml falcon spun at 4600rpm for 20 minutes. Anything 100kDa protein or smaller will pass through the filter into the flow through, whereas the 150kDa Cas9 protein will remain in the supernatant. In this manner the crude Cas9 protein was concentrated to 3ml. This was then subject to MBP cleavage using the TEV protease and dialysed using a 12-14kDa membrane (spectrapore) in 2L 1X dialysis buffer in a large burette. This buffer comprised: 20mM Hepes NaOH pH 7.5, 100mM KCl, 10% Glycerol and TCEP 1mM, one again all passed through a 0.22 μ m filter. The protein was then concentrated once again and then subject to ion-exchange chromatography using a 2ml SP sepharose column. This ion-exchange step is considered optional for many protein purification protocols, however, it is used to remove the MBP tag and free nucleotides DNA/RNA contamination perhaps pertinent when considering the Cas9 protein hybridises to RNA. The positive cas9 protein binds to the negatively charged beads during the loading process. The Cas9 protein is then eluted by the addition of varying concentrations of Buffer B, this being the same as A with the addition of a 10 fold increase in KCl to the concentration of 1000mM. Initially, there is a two column volume addition of elution gradient of 10-38%, 38-80% and finally 80-100%. As before 2 μ l aliquots of the eluate and the increase of the 260-280ratios was used to indicate the fractions Cas9 protein was eluted in

The previous purification gave rise to two pools of Cas9 protein, pool 1 deemed slightly purer and pool 2 which likely is still pure but requires examination for contamination. In this manner the two pools of MBP-cleaved and enriched protein were further enriched using amicon filters as previously described, initially the 400 μ l of the purer sample was loaded in two tranches to the superdex size exclusion column equilibrated using 5% Buffer B. Notably, this column does not have a loading and elution process rather all protein loaded travels at different speeds through porous beads depended upon their size. Smaller contaminants and proteins are eluted sooner. This serves to provide a homogenous suspension of 150kDa Cas9 protein. Similarly to the previous purification processes, 2 μ l of protein eluted was visualised on an 8% SDS-PAGE gel, at this stage samples were deemed pure and adjusted to 30 μ M and 90 μ M concentrations for the purer preparation denoted P108 and the second pool denoted P109. At this stage the

Cas9 protein was considered pure and concentrated to achieve a desired concentration of 4.7mg/ml, aliquoted to 20µl and snap frozen in liquid nitrogen.

2.23.2 In Vitro Cleavage Assay using Cas9 protein and In Vitro Transcribed

RNA:

Ribonucleoprotein complexes were generated by mixing in vitro transcribed T2-GFP guide RNA and Cas9 protein. The resultant RNP was generated at 5, 15, 50 and 150nM concentrations. This was then incubated with a Vgej-AAV with XmnI linearised plasmid encoding a GFP target sequence at 5nM, for either 30 minutes or 1 hour at 37°C. The resultant reaction was stopped using: 40µl RNase A (Invitrogen 20mg/ml), 50µl Sucrose 800µ/µl and 40µl Proteinase K (NEB). The cleavage products were then resolved on 0.8% agarose gel with ethidium bromide.

2.24 Zeocin Selection Methods

2.24.1 Materials

- Zeocin 8 x 1.25ml (Invitrogen)

2.24.2 Generation of a Zeocin Kill-Curve:

Following passaging, cells were plated in duplicate at a density estimated to provide between 25-35% 24 hours post seeding: HEK293T 1.7×10^5 and patient myoblasts 8.3×10^4 . In each instance, 24 hours post seeding growth media was removed and replaced with fresh media supplemented with incremental concentrations of zeocin. The standard concentration range was: 0, 50, 100, 200, 400, 600, 800 and 1000µg/ml. In order to ensure the cells were under constant selection, 1ml of the growth media was removed from each well and replaced every 3 days with 1ml of fresh growth media supplemented with an appropriate concentration of zeocin.

The seeding of cells in incremental concentrations of zeocin was denoted Day 0; microscopy images were subsequently taken at day 3, 7, 10, 14, and 21 respectively. This was to gradually examine morphological changes and any cellular de-attachment that may take place as a consequence of zeocin exposure over this time period and to assess whether cell cultures showed reduced viability in the presence of this antibiotic.

2.25 Production of Polyclonal MD1 Cell lines

HEK293T cells and Patient myoblasts were subject to a standard Viafect transient transfection of U6-Guide 1-CBh-Cas9-2A-GFP and hINT1-MD1-RT, at 2µg: 2µg or 4µg: 2µg. These were incubated at 37°C at 5% CO₂ for 48 hours, and then the cultures were examined for the presence of green fluorescence. This was used as a visual indicator of Cas9 nuclease expression.

If Green fluorescence was present, cells were passaged. To achieve this media was aspirated, each well was washed with 500µl 1x sterile PBS, before the addition of 500µl of 1x Trypsin. Once trypsin was applied to the 6 well plates, they were incubated at 37°C for 2 minutes. Post incubation the trypsin was neutralised with 2ml of appropriate growth media. The cell suspension was then transferred to a 15ml falcon tube, which was centrifuged at 1000rpm for 5 minutes, before cells were re-suspended in 3mls and counted. Following this process treated HEK293T and patient myoblasts were seeded at cell densities of 1.7×10^5 and 8.4×10^4 respectively. HEK293T samples were seeded in duplicate across two plates; this was to reduce the risk of contamination and to allow the cells to be exposed to two antibiotic doses of 600µg and 800µg respectively. This process was repeated with patient myoblasts except with doses of 800µg and 1000µg.

Seeding the cells in the presence of zeocin was classed at Day 0 of selection. To maintain selection 1ml of the growth media was removed from each well and replaced every 3 days with 1ml of fresh growth media supplemented with an appropriate concentration of zeocin. In addition cultures were the subject to microscopy imaging at 3, 7, 10, 14 and 18 days respectively. This was performed to examine visual differences in growth between the wild-type and treated cultures, in the presence of zeocin. Patient myoblasts were unable to survive this process. In contrast, selection continued for HEK293T.

By day 20, wells showed evidence of 'patchy' confluence where localised plaques had formed within wells. The cells were therefore passaged and split 1 in 2, and grown under continued zeocin selection until confluency was attained being photographed at 24 and 27 days. Once confluent these plates were split 1 in 2 again. The cells from 3 of the 4 x 6 well plates were harvested for DNA, RNA and Protein.

The cells from the fourth plate were counted: a proportion was used to generate a monoclonal culture, the remainder were placed in cryovials in a Mr Freezie at -80°C overnight and transferred to liquid nitrogen for long term storage.

2.26 Generation of HEK293T-MD1 Monoclonal Cultures

2.26.1 Materials

- Cloning Discs (Sigma; 50 discs/tube)
- Crystal Violet Powder (Sigma): 0.1g crystal violet, 80ml of ddH₂O and 20ml of absolute methanol. (0.1% solution)
- Sterile Forceps
- Media supplemented with zeocin
- 1 X Trypsin (Gibco)
- 1 x Sterile PBS (Gibco)

- 10cm Tissue Culture Dishes (Corning)
- 24 well plate (Corning)

2.26.2 Culturing Cells at Low Density to Attain Single Colonies:

To attain monoclonal cultures, the cells counted from the fourth six well plates in polyclonal cultures were serially diluted to cell densities of 10^5 and 10^4 in 10mls of cell growth media. These were then seeded into 10cm petri dishes in the presence of the concentration of zeocin antibiotic, reflective of the enriched polyclonal condition 600 μ g or 800 μ g for HEK293T/ 800 μ g or 1000 μ g for Δ 45-55 patient myoblasts. When examined at 24 hours post seeding individual cells could be seen as attached to the dish with ample distances between them due to high dilution. These were cultured until individual colonies became 1 -3mm in diameter and could be seen by eye. At this stage, microscopy images were taken at 25x magnification.

2.26.3 Identification and Picking of Individual Colonies

The 10cm petri dishes were held overhead and examined for individual colonies with sufficient spacing for picking. Once colonies were identified, cloning discs were poured into a separate clean petri dish, with 500 μ l of trypsin pipetted to the opposing edge of the dish.

Once the materials were organised, petri dishes harbouring single colonies were carefully tipped to allow the media to be carefully aspirated, and colonies washed with 2ml 1x sterile PBS. Following washing the cloning discs were picked up with sterile forceps, dipped into 1x Trypsin and then applied to the colony of interest identified. This was performed for four colonies for each petri dish and incubated for 2 minutes at 37°C.

During this 2 minute incubation, a 24 well plate was prepared with 1ml of growth media supplemented with antibiotic per well (at concentrations consistent with enriched polyclonal conditions, as described in 1.24.2). Once the incubation was completed colonies were picked off of the plate (attached to the cloning disc) and transferred into a well in the 24 well plate.

Once four colonies were picked the remaining colonies were stained by adding 1ml of 0.1% crystal violet stain. The petri dish was inverted to ensure the stain attained maximum coverage, and then excess stain was washed away carefully with water. Prior to the dish being inverted and left to air dry.

These cells in the 24 well plate were cultured and expanded until 4x 6 well plates were attained for the purposes of freezing a stock and DNA, RNA and protein extractions.

2.27 Molecular Characterisation of Colonies

2.27.1 Integration PCRs

2.27.1.1 Primers

Primer Name	Primer Sequence 5'-3'	Amplicon Size (Bp)
LAH Out F	GGAATGTACCTTCCTAATGTGATCA	1291bps
LAH Out R	AAGGTTTTCTTCTGCACGTCC	
LAH In F	TGAAGCACAGATTCTACTTTCAGAG	1139bps
LAH In R	GAAAGGCAAAGTGGATGTCAGTAAG	
RAH Out F	GCATCACAAATTTACAAATAAAGC	1277bps
RAH Out R	TGCGGTTTATTGACATTAAGACAT	
RAH In F	TTTCACTGCATTCTAGTTGTGGTTT	1171bps
RAH In R	AAATTTGCAGCATAATTGGAAGTGT	

- LAH – Left Arm of Homology, RAH – Right Arm of Homology

DNA harvests from excess cells at the 48 hour time point, polyclonal and monoclonal cultures were examined for the integration of MD1, by using primers that span from the intron 1 genome into the exogenous repair template.

2.27.1.2 Nested PCR

At the 48 hour time point it was necessary to nest the PCR with outer and inner primers spanning the intron 1 genome – exogenous repair template junction. Standard PCR reactions were set up using GoTaq G2 Flexi with the outer primer sets across left and right junctions (see section 1.15.1.3).

Cycling conditions: **Denaturation** 95°C 2minutes, **20 cycles** – (95°C 30s, 57 °C 30s, 72°C 90s), **Final Extension** 72°C 10 minutes and held at 4°C

Toward the end of the amplification cycle, an amplification using the inner intron 1 primers was prepared and supplemented with 1µl of amplification product from the first round.

Cycling conditions: **Denaturation** 95°C 2minutes, **35 cycles** – (95°C 30s, 57°C 30s, 72°C 80s), **Final Extension** 72°C 10 minutes and held at 4°C.

2.27.1.3 Single Round PCR

For the later more enriched polyclonal and monoclonal populations a single round GoTaq G2 flexi PCR using only the outer PCR primer primers across the intron 1 genome-exogenous repair junction was performed.

2.27.2 Reverse Transcriptase PCR

Total mRNA was harvested from polyclonal and monoclonal cultures, to examine the splicing between the Dp427m and Exon 2 of the MD1 exogenous repair template. Reverse transcriptase, is the key enzyme of the RT PCR and used to generate complimentary DNA

(cDNA) from total RNA lysate. Once the cDNA has been synthesised, it is then used as template for exponential amplification of a specific target sequence.

To begin with, 500 µg of the total RNA lysate from polyclonal and monoclonal culture were added to a reaction comprising: 1µl of Oligo-p(dT)₁₅ [50µM] (Roche) and 1µl random primers 10pmol/µl [10µM] made up to 10µl with DEPC treated H₂O. This was then heated to 70°C for 5 minutes and held at 4°C for 5 minutes, to eliminate the any secondary structures that may be present.

Subsequently 10µl of a master mix comprising: 4µl (5x GoScript reaction buffer), 2µl MgCl₂, 2µl dNTPs and 1µl reverse transcriptase and 1µl DEPC treated water. To attain final concentrations of 1x buffer, 2.5mM MgCl₂, 0.5mM dNTPs and 8 units of enzyme respectively.

The cycling conditions: 25°C for 5 mins, followed by 42°C for 1 hr, 70°C for 15 mins and then held at 4°C

In the next step, 1 µl of the cDNA was used as a template for a PCR reaction using primers across the Dp427m-Exon 2 genome-repair template junction.

Primer Name	Primer Sequence 5'-3'	Amplicon Size (Bp)
hINT1 SC F	TAGAGGACTGTTACGAGCGC	250bps
hINT1 SC R	TCCACGTTGTTGTTCTGCAG	

Cycling conditions: 95°C 2minutes, **35 cycles** – (95°C 30s, 57°C 30s, 72°C 80s), **Final Extension** 72°C 10 minutes and held at 4°C.

2.27.3 MD1 Western Blotting

Total protein lysate was also obtained from polyclonal and monoclonal cultures, to assess MD1 expression. Protein was quantified and each sample was prepared at a 50µg concentration. In instances where the protein was too dilute to attain 50µg, such as in repair template only conditions the maximal volume of protein was used in the production of mastermix. This was then subject to western blotting using the Odyssey system as described in section 1.14.

Chapter 3 In Silico Analysis of the 5'-End of the *DMD* Gene and Design of Customisable Endonucleases:

3.1 Introduction

3.1.1 Human *DMD* Gene

The genetic aetiology of Duchenne muscular dystrophy (DMD) has long been understood, with the identification of the *DMD* gene being one of the prominent successes of positional cloning in the 1980s (Koenig et al. 1987). However, the complexity of the *DMD* gene has made correction of disease pathology challenging. Firstly, the *DMD* gene is sizeable, being approximately 2.4Mb in length and constituting 0.01% of the human genome. Spanning this region are 79 exons that encode a 14kb mRNA transcript, which is eventually translated to a 427kDa full-length dystrophin protein (Koenig et al. 1987; Bushby et al. 2010a). Secondly, although it was first anticipated that dystrophin was a muscle-specific protein, it has since been determined that the *DMD* gene encodes numerous full-length and truncated isoforms, expressed in either muscle or brain tissues (Nudel et al. 1989; Nudel 2005). This is mirrored clinically by one third of patients exhibiting a non-progressive cognitive impairment in addition to the dystrophic muscle pathology prominent in the disease (Bushby et al. 2010a; D'Angelo et al. 2011).

The long rod-shaped dystrophin protein encoded by the *DMD* gene comprises four main domains: 1) N-terminal F-actin binding domain, 2) central rod domain of 24 spectrin-like repeats, interrupted into 3 sub-regions by 4 proline-rich hinges, 3) cysteine-rich domain which binds dystroglycan and the DAPC, and 4) C-terminal domain involved in syntrophin and dystrobrevin binding (Koenig et al. 1988; Banks et al. 2010). These domains are all important to the structural link dystrophin provides between intracellular cytoskeletal actin and extracellular matrix, pertinent to muscle integrity (Allen et al. 2016). However, it has since been determined that in addition to its structural role, spectrin-like repeats 16 and 17 of dystrophin protein interact with neuronal nitric oxide synthase (nNOS) and 6 and 7 with Par1b signalling proteins (Lai et al. 2009; Dumont et al. 2015). Studies suggest nNOS has a role in modulating muscle hemodynamics, and Par1b in microtubule dynamics and asymmetric division in satellite cells. These activities and others serve to extend the anticipated functions of the dystrophin protein (Dumont et al. 2015).

Beyond these features, further challenges to the development of genetic treatments are posed by the wide spectrum of mutation types that give rise to DMD, that result in the destruction of the open reading frame. These mutation subsets include: large deletions (68%), duplications involving one or more exons (12%) and small mutations (20%) (Bladen et al. 2015). Notably,

these mutations are not uniform throughout the *DMD* gene. Rather they are clustered around two mutation hot-spots between exons 2 -20, and between exons 45-55, situated near genetic recombination points (Aartsma-Rus et al. 2006; Jarmin et al. 2014). As a consequence of this, genetic therapeutic strategies have largely been mutation-specific, and focused upon restoration of the open reading frame and conversion of a DMD phenotype to the clinically milder BMD-like phenotype, with the outlook of stabilising and slowing disease progression. In contrast to DMD, BMD mutations retain the reading-frame of the transcript, and give rise to truncated dystrophin protein products with internal deletions (England et al. 1990; Aartsma-Rus et al. 2006; Taglia et al. 2015). One notable example of a mutation-specific genetic therapy is exon skipping in which antisense oligonucleotides are delivered to block splice enhancer sites within dystrophin pre-mRNA, promoting the exclusion of specific frame-shifting exons and restoration of the open reading frame (Aartsma-Rus et al. 2009; Cirak et al. 2011; Jarmin et al. 2014). In contrast, universal gene augmentation strategies are normally reliant upon the delivery of full-length or truncated dystrophin cDNAs, driven by either viral or muscle-specific promoters (Romero et al. 2004; Le Guiner et al. 2017). This often results in expression of dystrophin at lower levels than that observed physiologically, with the extent of muscle correction being dependent upon the route of administration. Moreover, these gene based therapies may require repeat administration and may carry adverse immunological risks.

In addition to the complexity of the *DMD* gene architecture and diversity of mutation types, dystrophin protein has a complex spatial expression profile. Shortly after discovering the muscle isoform of dystrophin protein, a full-length dystrophin transcript was also identified in the brain (Nudel et al. 1989; Nudel 2005). It was quickly determined to be a distinct isoform, differing at the initiating exon and converging from exon 2 onwards to the previously identified full length sequence. This was originally denoted as a brain isoform, but the nomenclature was later revised to cortical upon the elucidation of two further isoforms expressed in the cortical and cerebellar Purkinje regions of the brain (Górecki et al. 1992). Thus the full-length isoforms were referred to as Dp427c, Dp427m and Dp427p1 and Dp427p2, denoting cortical, muscle and Purkinje expression respectively. In addition to these full length variants, shorter DMD gene transcripts deriving from internally located promoters and with non-muscle expression patterns were also identified. These shorter transcripts encoded shorter protein isoforms including Dp260, Dp140, Dp116 and Dp71, with the numbers indicating the size of the protein in kDa (Bar et al. 1990; Byers et al. 1993; Lidov et al. 1995; D'Souza et al. 1995; Austin et al. 1995). It should be noted that most therapeutic strategies aim to ameliorate the prominent muscle pathology of DMD by correcting or reintroducing expression of the muscle isoform.

Notably, our understanding of the *DMD* gene continues to evolve. Recently an internal ribosomal entry site (IRES) within exon 5 has been identified, and it is speculated that the *DMD* mRNA can be translated from this point in the presence of some 5' gene mutations (Wein et al. 2014). This serves to explain why some 5' mutation types that are predicted to result in the *DMD* gene being out-of-frame, actually give rise to a BMD phenotype. In addition, a further dystrophin isoform Dp412e has also been recently discovered which initiates from a promoter active in embryonic cells (Massouridès et al. 2015). Continued investigation of the *DMD* gene architecture and scrutiny of reported findings over coming years is still warranted, particularly as historically a fifth full length isoform Dp427l, was identified in lymphocytes and later discredited on the basis of inconsistent splicing evidence and low physiological significance (Nishio et al. 1994; Wheway & Roberts 2003; Nudel 2005). Ongoing advances in our understanding of the *DMD* gene through dystrophin protein structure and function studies thus continue to inform and refine the development of new translational therapeutic strategies.

At the genomic level, the current series of studies aim to assess the relative positions of the full-length *DMD* promoters (Dp427c, Dp427m, Dp427p1 and p2), and identify a putative intron 1 genome editing target sequence common to all. It is anticipated that targeting this common intron 1 sequence with a customised endonuclease and an exogenous repair template would enable the reconstitution of dystrophin cDNA under the control of all endogenous promoters. It is speculated that this would enable correct spatial expression in cortical, muscle and Purkinje cells consistent with physiological levels observed. The exploitation of this could provide clinical benefit in the context of both the degenerative muscle pathology and the non-progressive cognitive impairment.

3.1.2 Murine *Dmd* Gene:

Interestingly, even prior to identification and cloning of the human *DMD* gene and cDNA, a naturally occurring mouse model of DMD was identified with relatively mild dystrophic muscle pathology, this being the *mdx* mouse (Bulfield et al. 1984). However, prior to becoming an established animal model of *DMD*, an examination of the genetic aetiology was required. This initially proved elusive until a point mutation within exon 23 was identified resulting in a premature stop codon (Sicinski et al. 1989). Moreover, research into the murine *Dmd* gene, indicated it had a similar genomic localisation and expression profile to the human orthologue (Heilig et al. 1987; E. K. Brown 1987; Chamberlain et al. 1987; Chamberlain et al. 1988). Thus it is perhaps unsurprising that when a range of novel putative promoters driving expression of full-length dystrophin were identified within the human genome, analysis revealed parallel sequences were found in the mouse genome (Koenig et al. 1987; Nudel et al. 1989; Górecki et

al. 1992). In combination, the dystrophic phenotype and the similar genomic organisation of the *DMD* gene served to establish the mouse as an attractive model of DMD in the context both basic and translational research.

Generally, mice are among the most favoured animal models. Pertinent to this is the comparable nature of the murine and human genomes, with 99% of human genes having a murine orthologue as discovered by the Human Genome Project and Mouse Genome Sequencing Consortium (Waterston et al, 2002). In addition to the high number of orthologues, 90% of human and mouse gene show conserved synteny, this being a comparable gene organisation or structure resulting from inheritance from a common ancestor (Waterston et al, 2002). Aside from these genetic factors, logistically, the small size of mice, quick maturation times and relatively easy husbandry, all serve to make them beneficial models to scientific research (Larcher et al. 2014; Whitmore & Morgan 2014; Manning & O'Malley 2015). The *mdx* mouse model is no exception and has remained the most favoured experimental animal model of DMD. This stated, there has been growing demand for animal models that exhibit more clinically representative phenotypes (Larcher et al. 2014; Whitmore & Morgan 2014). As a consequence therapeutics are often tested in mouse as a small animal model prior to progressing to larger models with more dystrophic muscle pathology such as the Golden Retriever Muscular Dystrophy (GRMD) model (Nakamura & Takeda 2011). Interest has also shifted to the use of rat models (Nakamura et al. 2014; Larcher et al. 2014). Most recently a CRISPR-generated rat model of DMD has been described (Huchet et al. 2017).

Importantly, *DMD* orthologues are present in a multitude of mammalian and non-mammalian species and exhibit a high level of sequence similarity both at the DNA and protein levels as summarised in table 1. This is crucial to the development of representative clinical models and the translation of therapeutic strategies demonstrated in *in-vitro* cell cultures into *in-vivo* models. This investigation will utilise the comparable *DMD* gene architecture across both human and mouse species, aiming to assess and compare the genomic sequence of the *DMD* gene at the 5' end to identify common endonuclease target sites in both species. This would enable the development of dystrophin cDNA integration techniques in human cell culture, with the outlook for expedited clinical translation through the *mdx* mouse model. This is proposed by the means of corrected human myoblast engraftment into immunodeficient *mdx* mice (Vallese et al. 2013), or indeed via direct viral-mediated effector gene delivery.

Comparison to <i>H. Sapiens</i>	% Sequence similarity	
	DNA	Protein
Species		
<i>P. Troglodytes</i>	88.7%	85.2%
<i>C. Lupus</i>	92.9%	94.9%
<i>M. Musculus</i>	90.4%	91.3%
<i>G. Gallus</i>	78.4%	79.0%
<i>D. Melanogaster</i>	48.2%	39.0%
<i>A. Gambiae</i>	48.2%	39.0%

Table 1 The percentage sequence similarity of the *DMD* gene and Dystrophin protein, in a range of species relative to the human DNA and protein sequences. (NCBI Homologene 2015).

3.2 Chapter Aims:

This investigation is concerned with the development of a genome editing strategy with near-universal applicability to DMD patients. This is to be achieved by customised endonucleases and an exogenous repair template compatible with a variety of full length minidystrophin and microdystrophin cDNA configurations, targeted to the 5' end of the *DMD* gene, preferably intron 1. Previous work by our lab demonstrated that a cDNA block containing exons 45-52 flanked by synthetic splice acceptor and donor sequences could be correctly introduced into DMD patient myoblasts harbouring deletions of the same exons, thereby restoring the full dystrophin mRNA open reading frame (Poplewell et al. 2013). Thus it is speculated that exon 1 pre-mRNA sequence transcribed from the main full length endogenous promoters would be able to splice to dystrophin cDNAs delivered in a promoter-less exogenous repair template format with a suitable 5'-splice acceptor site. If site-specific integration was to occur, the dystrophin cDNA within this template would be expressed at physiological levels and mirror the spatio-temporal expression profile observed endogenously. Such expression could ameliorate all clinical phenotypes of *DMD*.

To facilitate attainment of these outcomes, this chapter identified the initiating exons of full-length dystrophin isoforms. These were mapped for both human and mouse genes, relative to exon 2 where all full-length isoforms have been determined to converge. An intron 1 sequence common to all dystrophin isoforms for both species was then subjected to a rigorous pairwise sequence alignment to assess sequence similarity.

The identified region of sequence similarity was then assessed for the presence of *Streptococcus Pyrogenes* CRISPR/Cas9 guides and TALEN target sites, being common to the human and mouse genome, it was speculated that this would permit facilitated translational development. Further analysis was also undertaken for the presence of species specific

endonucleases. Once suitable gene-editing designs were resolved upon, CRISPR guide expression plasmids were generated and this was later used to identify a gene-editing reagent able to introduce a DSB in intron 1 of the DMD gene with high efficiency.

3.3 Results:

3.3.1 Mapping the 5' End of the *DMD* Gene to Locate Full Length Dystrophin Promoters and Exon 1 Elements Relative to the Common Exon 2 Sequence:

Promoter motifs and downstream initiating exon sequences for each of the four human full-length dystrophin isoforms were accessed on NCBI: Dp427c NM_000109.3, Dp427m NM_00406.2, Dp427p1 NM_004009.3 and Dp427p2 NM_004010.3. The exon 2 where all full-length dystrophin isoforms converge and exon 3 sequences were obtained in a similar fashion from the Dp427m isoform. Once all of the sequences were identified and downloaded in a standard FASTA format, they were subject to a Basic Local Alignment Search Tool (BLAST) analysis against the full human X-chromosome sequence [Homo Sapiens Chromosome X, GRCh37.p13 Primary Assembly (Accession NC_000023.10)].

The full-length dystrophin isoform promoters were shown to occur in the following order on the X chromosome, Dp427c, Dp427m, Dp427p1 and Dp427p2, consistent with literature (Nudel 2005). Moreover, this analysis provided co-ordinates for all of the exons in question and by extension the intron 1 sequence common to all isoforms between Dp427p2 and exon 2, shown schematically in (Figure 3-1). Notably, the co-ordinates were consistent with the human *DMD* gene being encoded on the reverse strand, running in a 3' – 5' direction, as anticipated in accordance with the European Bioinformatics Institute (EMBL/EBI). The sequence spanning Dp427c to exon 3, with an additional 1kb flanking sequence, was downloaded in reverse orientation between the co-ordinates c33357826-33037000 of the X-chromosome. This ensured the sequence was obtained in a 5' – 3' orientation which was later confirmed via the Ensembl genome browser (Ensembl 2017).

This sequence retrieval and evaluation process identified two intron sequences at the 5' end of the *DMD* gene common to all full-length isoforms that would be suitable for targeting with customised genome editing reagents. The first sequence referred to as “common intron 1” spanned 107kb located between Dp427p2 and exon 2. The second sequence referred to as “common intron 2” spanned 170kb between exons 2 and 3. In terms of selecting a target site for genome editing the preferred candidate sequence in this instance is common intron 1 due to its close proximity to endogenous promoters which is anticipated to be beneficial for splicing and to potentially increase the breadth of DMD mutation applicability.

Once candidate target sequences were identified for the human genome attention was shifted to the murine *Dmd* gene. The high sequence similarity of human and mouse exons enabled the derived human sequences to be utilised in BLAST analysis against the full murine X-chromosome sequence [Mus Musculus Strain C57BL/6J Chromosome X, GRCm38.p1 C57BL/6J Primary Assembly (Accession NC_000086.7)]. As anticipated, corresponding promoter and exon 1 sequences were identified across the mouse X-chromosome occurred in the same order as on the human X-chromosome. One important distinction between human and mouse, however, is that the murine *Dmd* gene is encoded on the forward strand of the mouse X-chromosome. Thus in contrast to the human genome, the mouse coordinates attained in a 5' - 3' orientation and the same common sequence between mouse Dp427c and exon 3 was downloaded directly between co-ordinates of 82815000-83317000 of the mouse X-chromosome.

In the case of the murine genome, this analysis served to identify a common intron 1 spanning 93kb between Dp427 and p2 and a common intron 2 spanning 209kb between exon 2 and 3 respectively. This disparity in intron size across human and mouse genomes is not uncommon and is often attributable to alternative evolutionary patterns (Waterston et al. 2002). Importantly, the different sized human and mouse common intron sequences had implication for the type of sequence analysis that can be performed. The two main forms of alignment are global in which the 5' end of two sequences are 'anchored' and the alignment of the nucleotides occurs sequentially or local alignment where regions of high sequence similarity are identified and aligned regardless of their position in the sequence (McWilliam et al. 2013). On account of the different sizes it was decided the latter would provide a more suitable form of assessment.

3.3.2 Rigorous Local Pairwise Sequence Alignment of Human and Mouse Intron 1 Sequences: the Identification of a Region of Homology within *DMD* Intron 1.

The common intron 1 sequence was selected as being 3kb downstream of the identified Dp427p2 promoter and upstream of the common exon 2 sequence, prior to a rigorous pairwise local alignment being performed. These 3kb terminal sequences were excluded from analysis, to ensure modulatory signals and native splice sequences, necessary for transcription and RNA splicing were retained. This was deemed crucial as the native transcriptional machinery would be utilised to splice to the integrated dystrophin cDNA introduced by the exogenous repair template. It was predicted that once integrated the cDNA would then be driven by the endogenous promoter and give rise to physiological expression levels of dystrophin. This premise was the novel foundation of the therapeutic approach proposed by this work. These marginally shortened intron sequences were then subject to a pairwise local alignment using the EMBOSS matcher algorithm, provided by the EMBL EBI bioinformatics institute (https://www.ebi.ac.uk/Tools/psa/emboss_matcher/). Notably, this bioinformatics tool provides an optimal alignment with the exclusion of excessive adjacent sequence. As a consequence, the alignment of the adjusted 101kb and 87kb human and mouse common intron 1 sequences respectively identified a 12kb sequence as being optimally aligned. Interestingly, this 12kb alignment began 7kb into the human sequence input spanning from 7-19kb and 4kb into the mouse input sequence spanning 4-15kb (with a sizeable central gap in the middle of the alignment identified).

Importantly, the high level of sequence exclusion from the local alignment was anticipated on the basis that introns, or so-called 'non-coding' sequences, are identified as being less highly conserved across different species than exons. Interestingly however, the 12kb optimal alignment of human and mouse common intron 1 only exhibited 53.9% sequence similarity, with 15.8% gaps and dispersed mismatches throughout the sequences. This prompted a closer visual inspection of the sequence alignment, which in turn yielded a 345 nucleotide stretch with high sequence similarity (Figure 3-1).

This 345 nucleotide region of homology was then identified in human and mouse sequences, and independently aligned, this alignment indicated 84.9% sequence similarity between the two sequences with 5.5% gaps. Moreover, within this sequence, a continuous 100 nucleotide stretch with only 4 mismatches could clearly be identified. These observations in turn established the 345 region and the 100 nucleotide stretch, as sequences of interest to assess and select putative target sites for custom endonucleases. The rationale being designing an endonuclease to this sequence with the avoidance of polymorphisms, would potentially

mediate DSBs in both human and mouse genomes. This common genome editing reagent could then be exploited for the purposes of facilitated translational development.

The identification of these relatively short sequences prompted further assessment of the common intron 1 for more regions of high sequence similarity between human and mouse genomes. To this purpose, the human common intron 1 sequence was divided into 1kb fragments and aligned to the full mouse common intron 1. This was undertaken on the premise that the larger sequence inputs may have produced an alignment that obscured shorter regions of high sequence similarity between human and mouse genomes. This stated, no homology in excess of the 345 nucleotides identified was observed for the first 20kb of human common intron 1 aligned in this manner (results not shown); the failure to identify further regions of sequence similarity resulted in this approach being halted. Finally, a rigorous local pairwise alignment was undertaken for human and mouse common intron 2 sequences. However, no regions of comparable sequence similarity to that identified for intron 1 were observed (results not shown). Overall in this alignment regions of sequence similarity were fewer and shorter than observed for intron 1, with the longest continuous stretch being 24 nucleotides. Thus assessment for endonucleases progressed with the region of homology identified in the common intron 1 sequence alignment.

3.3.3 Identification of TALEN Target Sites and CRISPR Guide Designs to the 5' End of the *DMD* Gene:

3.3.3.1 Main Design Strategy: Targeting the Region of Homology Identified between Human and Mouse Sequences.

The recent development of the endogenous CRISPR/Cas9 system for the purpose of mammalian genome editing meant this system was untested in our lab. In addition, TALEN synthesis is relatively complex and expensive process. Finally, the desire for an expedited translation of this gene editing approach from *in vitro* cell culture to *in vivo* DMD mouse models was also among considerations which indicated that the design of a common endonuclease would be beneficial to this investigation. Thus the primary focus was to analyse the 100 nucleotide region of homology identified between the human and mouse common intron 1 for TALEN target sites and CRISPR guide sequences. Designs for both the established TALEN platform and novel Cas9 system, were undertaken in parallel enabling exploration of this new versatile genome editing system whilst also ensuring the best chance of successful outcomes for the project.

TALEN target sites were identified with the use of the Collectis TALEN hit software, with designs comprising of 17 nucleotide TALE left and right recognition units separated by a

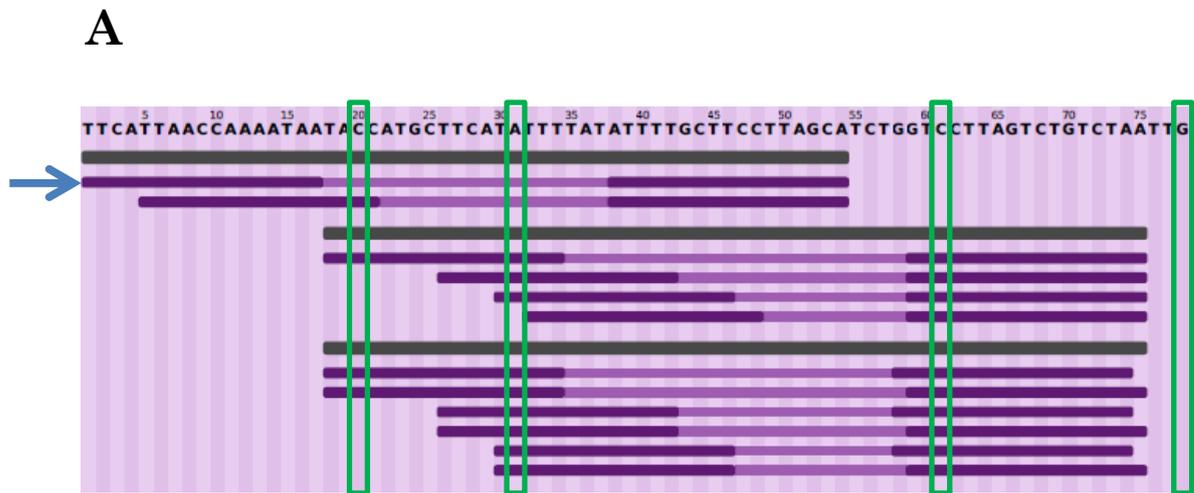
variable 10-25 nucleotide spacer region (Collectis Bioresearch 2011). This analysis yielded 12 TALEN pairs spanning 3 hit zones within the homologous region. Of these designs only 1 TALEN pair comprised of left and right TALE recognition arms targeting completely homologous sequences between both species (Figure 3-2). Notably, the design in question harboured two polymorphisms within the spacer region. However, this intervening sequence, serves to allow FokI nucleases tethered to the TALE arms to undergo obligatory dimerization thereby introducing a DSB. This is in contrast to the TALE arms which confer sequence specificity, enabling the endonuclease to bind to the desired genomic locus. In consideration of this, it was speculated that polymorphisms confined to the spacer region would have little bearing on TALEN specificity and efficacy. Based upon this assertion, this TALEN pair was firstly selected as a potential genome editing reagent common to both human and mouse intron 1 of the *DMD* gene.

Following the successful identification of a TALEN target site, the same common sequence was analysed for CRISPR guide targets with the use of the CRISPR MIT software, (Optimised CRISPR Design 2013). CRISPR guide RNA designs were produced in line with the requirements of the well characterised *Streptococcus Pyrogenes* Cas9, which is a 20 nucleotide sequence upstream of a protospacer adjacent motif (PAM). The PAM being a key recognition sequence either 5'-NGG-3' or 3'-NCC-5', where N denotes any nucleotide. The Cas9 nuclease is directed to cleave target DNA at a position 3 nucleotides upstream from this PAM recognition sequence.

Notably, the CRISPR MIT identifies CRISPR guides, and generates a list of potential off-targets by identifying similar sequences within a defined genomic build. As a consequence mouse and human sequences were analysed for CRISPR guide RNA designs individually against their genomic builds respectively. The results obtained from these two separate analyses were compared and compiled to assess for similar CRISPR guide designs. Overall, 10 CRISPR guides were identified: 4 human and 6 mouse (Figure 3-2). One guide, 5'-TATTTTGCTTCCTTAGCATC-3', was identified to target a sequence completely homologous between both species. This was noted as guide 2 for human and guide 5 for mouse on the basis of the numerical ranking system of the CRISPR MIT tool. Thus the Cas9 nuclease could be directed to cleave this sequence in both genomes.

The CRISPR MIT tool ranks guides on the basis of their specificity, using a specificity score (1-100) and a colour coding system to simplify the guide selection process; guides coloured green, amber or red have respective specificity scores of >50 (high specificity), 20-50 (moderate specificity) or <20 (poor specificity). In accordance with this ranking system, all guides designed with the exception of mouse guide 6 were coded as amber or higher (Figure 3-2).

Interestingly, in contrast to the CRISPR MIT tool, the Collectis TALEN design software provided no predicted specificity or efficacy profiles for the TALENs identified. Nonetheless provisional analysis of guides appeared promising and indicated that gene modification assessments were warranted to assess whether they functioned with high enough efficacy for inclusion in the investigation. Figure 3-3 provides a schematic of the TALEN and CRISPR guide and illustrates the designs locations across the 100 nucleotide region of high similarity between human and mouse sequences.



B

Human

	score	sequence
Guide #1	62	AGACTAAGAACCAGATGCTA AGG
Guide #2	46	TATTTTGCTTCCTTAGCATC TGG
Guide #3	43	GATGAAGCATGATATTATTT TGG
Guide #4	42	TCTGTCTAATTACATTATAA TGG

Mouse

	score	sequence
Guide #1	62	AGACTAAGGACCAGATGCTA AGG
Guide #2	60	TAATGCAATTAGACAGACTA AGG
Guide #3	56	TCTGTCTAATTGCATTATAA TGG
Guide #4	51	TATGAAGCATGGTATTATTT TGG
Guide #5	50	TATTTTGCTTCCTTAGCATC TGG
Guide #6	16	AAAATATAAAATATGAAGCA TGG

Figure 3-2 Design of TALEN and CRISPR/Cas9 Endonucleases to the region of high sequence similarity between human and mouse in common intron 1.

A) The raw output schematic of the collectis TALEN hit software: the input sequence is along the top, TALE recognition arms are illustrated in dark purple and intervening spacer in pale purple. In this instance, 12 TALEN designs were apparent across three sub-regions, the position of sequence mismatches between the human and mouse genome were highlighted by green rectangles. Notably, a single TALEN design with TALE arms targeting completely homologous sequences between human and mouse, was identified as indicated by the blue arrow. **B)** The guide design outputs for human and mouse from CRISPR MIT: guides are ranked numerically in accordance with their specificity 'score' (1 -100). Furthermore, they are colour coded on the left hand side: green optimal, amber moderate off targets and red high off-target effects. Finally, the protospacer 'guide' sequence is denoted in black whilst the PAM recognition motif is shown alongside in green. Of these designs human guide 2 and mouse guide 5, target a common sequence without any mismatches.

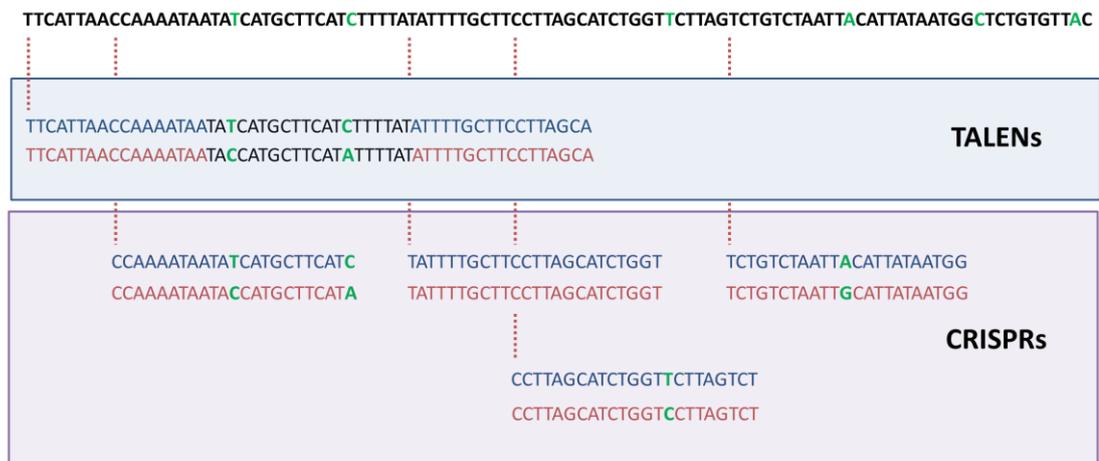


Figure 3-3 A schematic outline of endonuclease designs and the occurrence of single nucleotide polymorphism for the highly homologous sequence input.

The 100 nucleotide human sequence with high sequence similarity to the mouse genome can be seen at the top of the diagram and is denoted in black. Human and mouse endonuclease designs are indicated in blue and red respectively and sequence mismatches between the species are shown in green lettering. TALEN designs illustrate the two mismatches between the human and mouse sequence confined to the spacer region, whilst the TALE recognition arms are shown to be homologous. Of the CRISPR guide designs identified, only the centrally located human guide 2 and mouse guide 5 are completely homologous. All other human guide designs harbour at least a single nucleotide mismatch relative to their murine counterpart.

3.3.3.2 Alternative Design Strategy 1: Targeting DMD Exons 1 -3 in Human and Mouse

The low sequence similarity between human and mouse intron sequences common to the full-length isoforms, served to limit endonuclease designs. This was not unexpected as 'non-coding' intron sequences tend to be less conserved than exons across different species' genomes. Thus in an effort to expand the repertoire of common genome editing reagents, rigorous pairwise local alignment of the first three exons of the Dp427m isoform was also undertaken.

As predicted exon sequences showed a higher level of sequence similarity compared to the intron alignments shown above. The Dp427m exon 1 sequences showed 82.9% similarity with 5.0% gaps. Exon 2 sequences had 98.4% similarity with 0.0% gaps. Exon 3 sequences had 87.1% similarity with 0.0% gaps. Whilst higher homology exists between human and mouse exon sequences, the mismatches present were dispersed throughout the exon which served to limit identification of TALEN targets and CRISPR guides common to both species.

The pairwise sequence alignment of Dp427m exon 1 (Figure 3-4) harboured five notable regions of continuous sequence similarity spanning 33 nucleotides (position 243-276), 22 nucleotides (position 20-42) and 13 nucleotides (#1 at position 134-147, #2 at position 175-188, #3 at position 222-235). The former more extensive runs of homology of 22 and 33 nucleotides were situated at the 5' and 3' regions of the exon respectively. Thus whilst the homologous sequences would be large enough to permit the binding of left or right TALE recognition arms, the intervening sequence is too extensive and would not permit obligatory dimerization of the FokI nuclease. Importantly, in the absence of FokI obligatory dimerization, the reagent will not orchestrate DNA cleavage at the genomic locus. Furthermore, the remaining regions of 13 nucleotide sequence similarity, whilst more closely positioned are of insufficient length to permit TALEN designs common to both species to be undertaken. Thus the profile of sequence similarity for Dp427m exon 1 indicates that a common TALEN design to this region was not feasible.

The Dp427m exon 1 was also assessed for CRISPR guides. While 22 guides could be designed to the exon sequence, at least one sequence polymorphism between human and mouse was present in all cases. Similarly to the TALEN analysis, no common guide was present that would enable DSB induction in both species (Figure 3-4).

Contrastingly, exon 2 pairwise sequence alignment showed only a single centrally located polymorphism, at position 42 out of 63 nucleotides. When analysed with Collectis TALEN hit software 11 TALEN pairs across 3 hit zones were identified. Interestingly, in 8 of the 11 TALEN

designs, the single nucleotide polymorphism was constrained to the spacer region (Figure 3-5). Thus these designs represent potential genome editing reagents to induce DSBs in both the human and mouse genomes.

The CRISPR guide designs to this exon 2 sequences were more limited, with only 2 human and 1 mouse guide designs identified. The two guides from the human analysis were targeted to completely homologous sequences, between the two organisms. Interestingly however, only one of these guide sequences appeared when designs were made in the context of the mouse genome. Nonetheless, their low specificity scores, being 28 and 23 for human and 18 for mouse, contraindicated their development in this study. The limited identification of guide sequences targeting exon 2, was attributed to the A-T rich nature of this sequence, (71.4% being A or T) which reduces the occurrence of PAM sequences.

Finally, exon 3 pairwise sequence alignment identified 3 regions of high sequence similarity. These regions spanned 17 nucleotides (position 22-37) and 15 nucleotides (#1 at position 52-67 and #2 at positions 78-94) and they had insufficient length for TALEN or CRISPR guide designs to be undertaken (Figure 3-4).

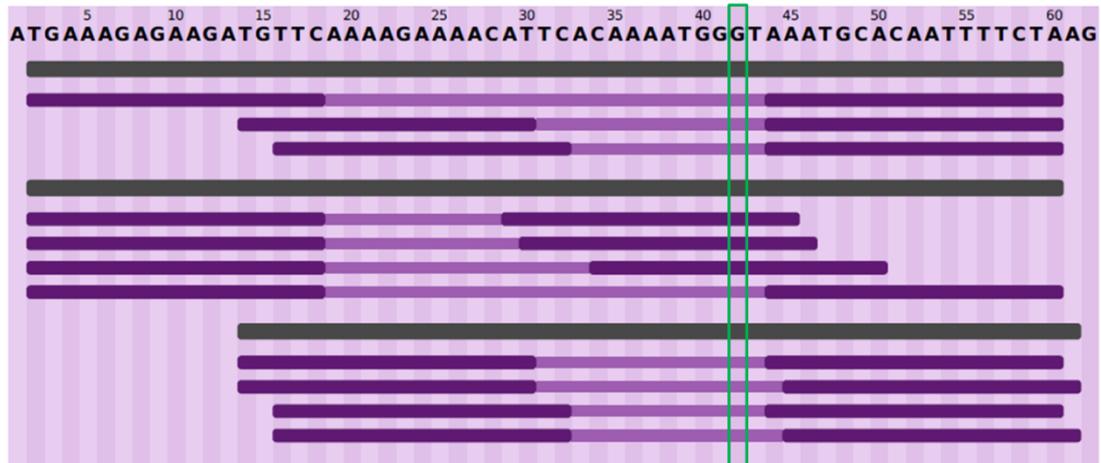


Figure 3-5 TALEN designs to Dp427m exon 2 using the Collectis TALEN hit software

A raw schematic output of the Collectis TALEN hit software: the human exon 2 sequences is along the top, TALE recognition arms are illustrated in dark purple and intervening spacer in pale purple. The sequence mismatch between human and mouse genomes is indicated by a green rectangle. In this instance, 11 TALEN designs were apparent across 3 sub-regions. Notably, there are 8 TALEN designs in which the TALE arms bind to sequences completely homologous between human and mouse. This would represent a series of endonuclease designs to develop, should guides designed to the common intron sequence show low efficacy.

3.3.3.3 Alternative Design Strategy 2: Species-Specific CRISPR Guide Designs.

Thus far, theoretical genome editing reagents have been designed to target homologous sequences between the human and mouse at the 5' end of the *DMD* gene. The ability to target a common site remains important as de novo synthesis of TALENs, remains expensive and thus it would not be possible to produce TALENs targeting multiple genomic sites. Thus it was speculated that a single reagent functioning in both organisms would be more cost effective and enable facilitated translational development. However, the re-purposed CRISPR/Cas9 system does not warrant the same considerations. In contrast to TALENs, this genome-editing platform forms an RNA-DNA duplex rather than protein-DNA interaction. Thus rather than extensive protein engineering, which is required to produce a novel TALEN targeting a distinct genomic site, the Cas9 nuclease can be re-targeted by changing a 20 nucleotide guide RNA sequence. This is achieved simply by self-annealing two primer sequences and sub-cloning them into a plasmid downstream of a U6 promoter. This ease of construction and versatility makes this an attractive genome editing platform but also enables multiple designs to a genomic locus to be undertaken and tested.

In addition to the distinctions of these genome editing platforms, some limitations were identified with regard to designing common genome editing reagents to both human and mouse species. These included: the low availability of homologous sequences, the distinct genomic localisation of identified sequences e.g. 7-19kb human and 4-14kb mouse and finally the recognition that even upon design of common genome editing reagents, distinct repair templates with species-specific arms of homology would be necessitated. These considerations, taken into combination, prompted the design of species-specific CRISPR guides with the intention of expanding potential target sites across the genomic locus of interest.

In pursuit of designing species specific genome editing reagents targeting the intron 1 sequence common to the full length dystrophin isoforms, human and mouse intron sequences were analysed independently for CRISPR guides. A new guide design tool 'CHOP' became available and was used for this purpose. In contrast to the previous CRISPR MIT tool, which only screened 250 nucleotides per input, this tool was permissive to larger sequence inputs. Hence the use of this tool was deemed more suitable to analyse the full intron 1 sequences, which were divided into 1kb fragments that were screened sequentially.

In addition to enabling increased sequence inputs, the 'CHOP' tool also utilised a ranking and colour coding system analogous to CRISPR MIT. In order to identify promising regions for gene targeting, the numbers of highly ranked guides were plotted for each 1kb fragment across the intron 1 sequence (Figure 3-6).

These one kilo base sequences were selected for further examination, if they contained 5 'highly ranked' guide designs positioned within a 300 nucleotide stretch. It was decided that 5 guide designs at each site would be desirable, to ensure at least one guide design functioned with high efficacy (Ran et al. 2013). Furthermore, the close proximity of guide designs would permit the design of a single asymmetric PCR spanning the region identified that could be used in subsequent mismatch detection assays.

This bioinformatics analysis identified 3 candidate sequences with sufficient CRISPR guide RNA designs for each of the human and mouse intron 1 sequences. Human sites were positioned at 67, 83 and 88 kilobases (Figure 3-7) and mouse sites were positioned at 28, 29 and 41 kilobases into the common intron 1 sequence identified (Figure 3-8). This served to expand the repertoire of intron 1 targeting options, ensuring that if high efficacy guides were not present targeting both mouse and human homologous sequences, then investigation would remain viable utilising alternative species-specific target sites.

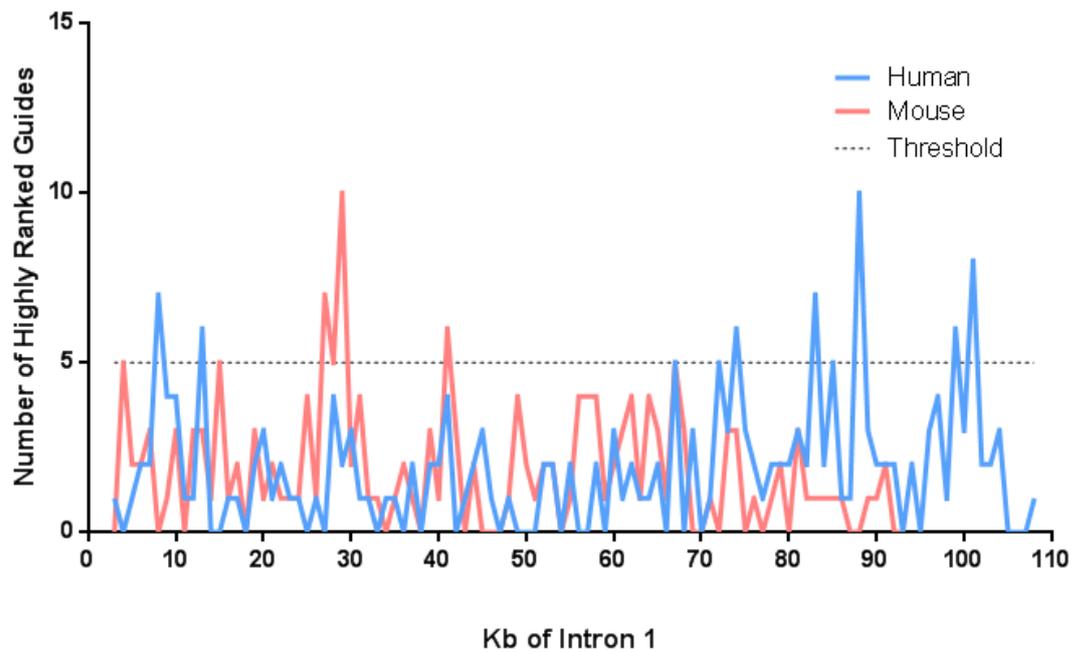
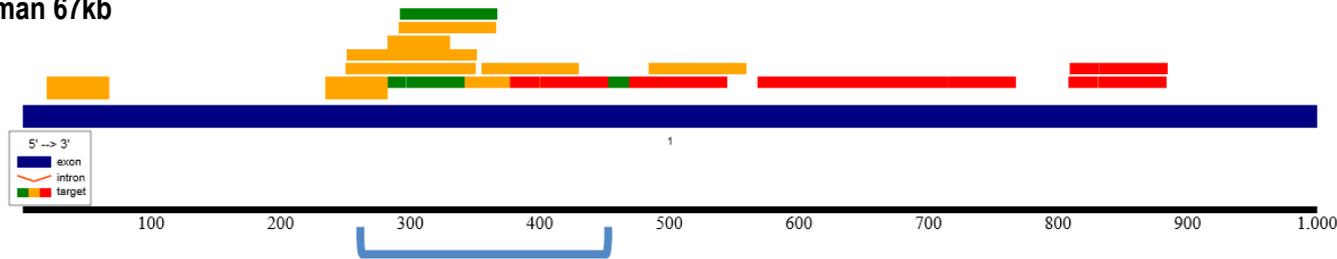
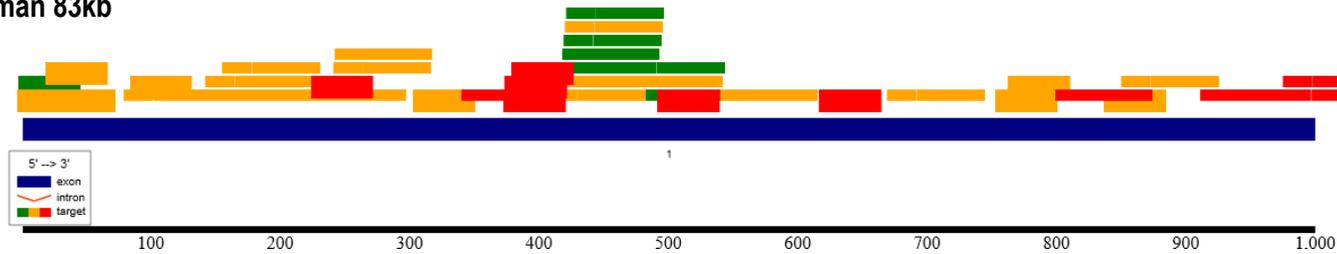


Figure 3-6 A graph showing the number of highly ranked guide designs for human and mouse, common intron 1 sequences.

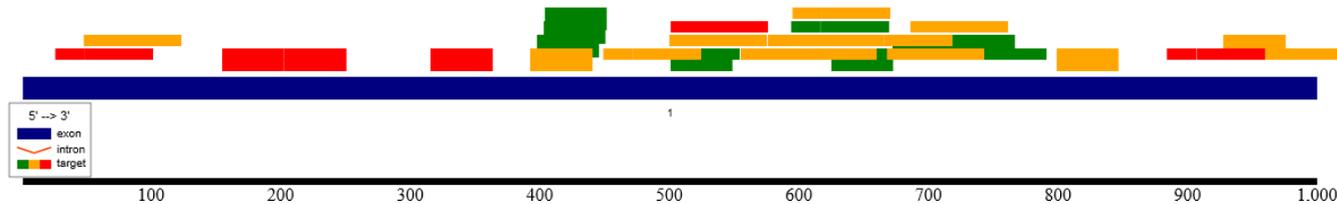
The number of highly ranked guides with a specificity score of 50 or above was plotted for every kilobase of human and mouse intron 1 sequence: blue (human) and red (mouse). Furthermore, a threshold of 5 highly ranked guides was used as a criterion to select genomic sites of interest, on the basis that guide designs cleave the genome with variable efficacy. In this manner, species specific CRISPR/Cas9 guide designs were undertaken across the whole common intron sequence and positions of interest were short-listed.

A**Human 67kb**

275 – 442 bps

B**Human 83kb**

412 – 460 bps

C**Human 88kb**

403 – 506 bps

Figure 3-7 Schematic diagrams showing the guide designs for three of the short-listed target sites within the human common intron 1 sequence performed using the CHOP CHOP guide design tool.

A) Site 1 positioned 67kb into the common intron 1 sequence. B) Site 2, 83kb C) Site 3, 88kb. Notably, the guides are colour-coded to indicate the predicted specificity ranking: green - optimal with few off target effects, amber- moderate off-target effects and red – extensive off-target effects. Importantly, the regions the 5 guides are localised to, is indicated by a blue horizontal bracket.

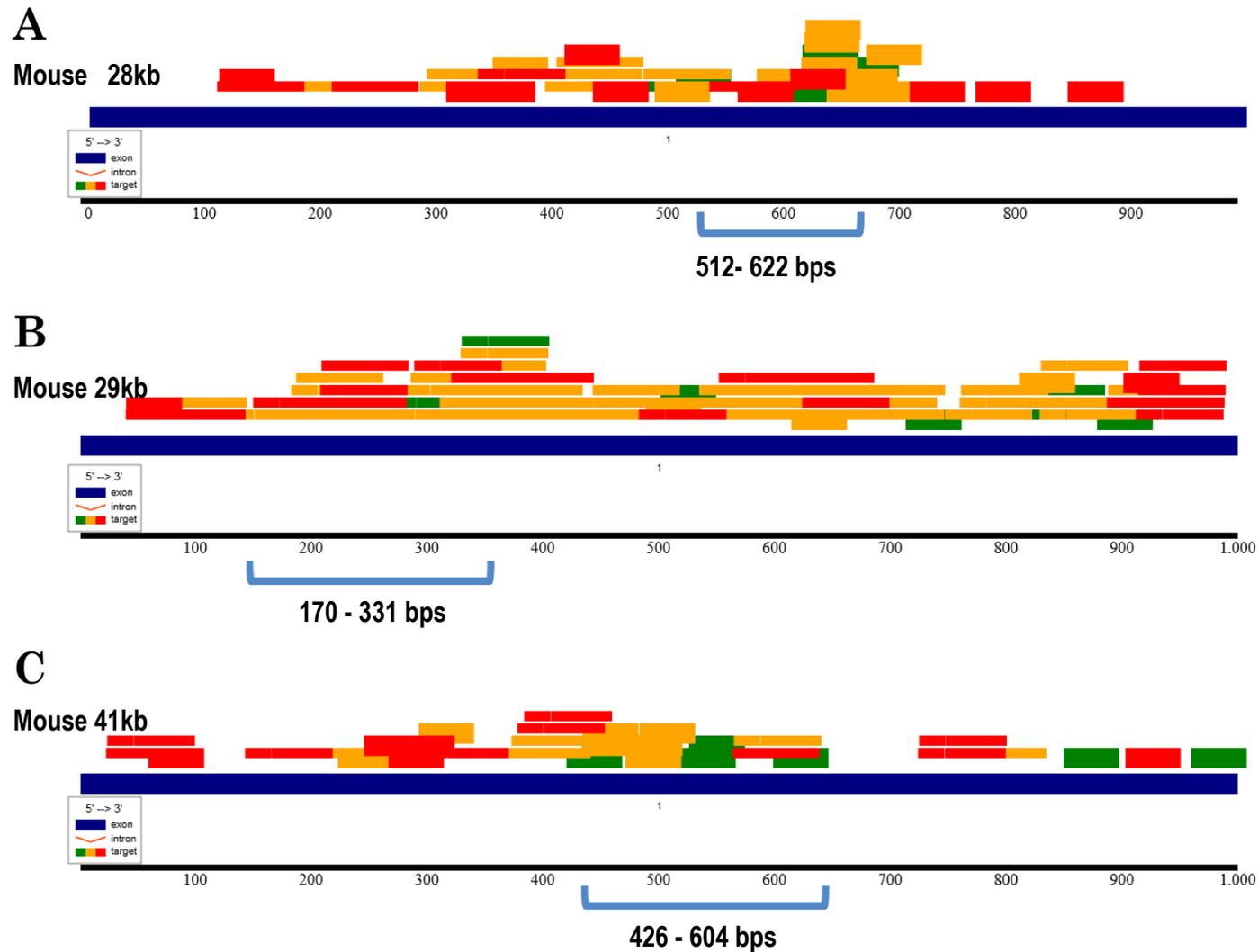


Figure 3-8 Schematic diagrams showing the guide designs for three of the short-listed target sites within the mouse common intron 1 sequence performed using the CHOP CHOP guide design tool.

A) Site 1 positioned 28 kb into the common intron 1 sequence. B) Site 2, 29kb C) Site 3, 41kb. Notably, the guides are colour-coded to indicate the predicted specificity ranking: green - optimal with few off target effects, amber- moderate off-target effects and red – extensive off-target effects. Importantly, the regions the 5 guides are localised to, is indicated by a blue horizontal brackets.

3.3.4 Sub-Cloning of Guide Sequences

3.3.4.1 U6 Plasmids:

Due to the ease of construction and increased versatility of the CRISPR/Cas9 genome editing platform relative to TALENs, it was resolved that the former nuclease would be used throughout the investigation. This decision was made on the basis that if the system proved challenging to establish, the investigation could revert to targeting with TALENs, thereby increasing the potential for successful project outcomes.

To facilitate the generation of an expression plasmid for the CRISPR guides, an empty U6 plasmid was used kindly donated by the lab of Professor Jin-Soo Kim (Figure 3-9). This U6 plasmid comprised of a 700bp spacer, flanked by asymmetric BsaI restriction sites, where self-annealed oligonucleotides constituting the guide sequence should be cloned. In addition, U6-plasmids encoding human guide designs 1 -4 were kindly cloned by Dr. Taeyoung Koo, to facilitate the development of mismatch detection assays. For mouse guides, oligonucleotides containing the 20 nucleotide guide sequence in forward and reverse orientations were synthesised. These were flanked with guanine and cytosine residues on the 5' forward and 3' reverse primer termini, and suitable overhangs for the cloning site respectively. These sequence additions are necessary to ensure the cloning of the guide and transcription by RNA polymerase (Gagnon et al. 2014).

To generate the mouse guide RNA expression plasmids, the U6 plasmid was digested with BsaI to obtain the plasmid backbone by gel purification (Figure 3-10), oligonucleotides self-annealed and finally the two constituents were mixed in a ligation reaction. These mixtures were used in the heat shock transformation of Top10 chemically competent *E. coli*, which once recovered were plated onto agar plates inoculated with ampicillin. Subsequent colonies were analysed for the presence of the correct guide sequences by colony PCR. In instances where the guide sequence had not been successfully sub-cloned an 817bp band was anticipated in which the 700bp spacer was present and amplified. In contrast if successful guide sub-cloning had occurred a 140bp product was expected, in which the guide sequence was inserted in place of the spacer. This process was shown to occur with high efficiency as exemplified by the cloning of one of the mouse guides shown in Figure 3-10. In this instance of the 16 guides screened, 15 were positive for the presence of guide sequence and 1 was negative on the basis of colony PCR. Following this primary indication, 2 colonies positive for the presence of the guide sequence were grown and plasmid DNA was isolated and sequenced. If sequencing provided findings consistent with the colony PCR, the mouse guide plasmid was expanded and purified for use in transfection experiments.

3.3.4.2 U6-Guide-CBh-Cas9-2A-GFP Plasmids

The methodology of cloning guide sequences into expression plasmids remains conserved across a range of differing plasmid types. An analogous method was utilised with regard to the U6-guide-CBh-Cas9-2A-GFP plasmid, the only notable distinction being the U6 guide cloning site in this construct which harbours a single BbsI restriction site with no spacer region (Figure 3-11). As a consequence of this, no preliminary screening was possible, rather a multitude of bacterial colonies were cultured, from which plasmid DNA was extracted and sequenced. This stated the efficiency appears comparable to that shown in the previous section. As high frequency success was obtained, only a few colonies needed to be expanded to identify the guide sequence anticipated.

This plasmid was utilised for the dual expression of Cas9 and GFP, with GFP expression being used as a proxy for transfection efficiency. This made this plasmid extremely useful when extending methods into patient and mouse myoblast lines which relative to HEK293T are more challenging to transfect.

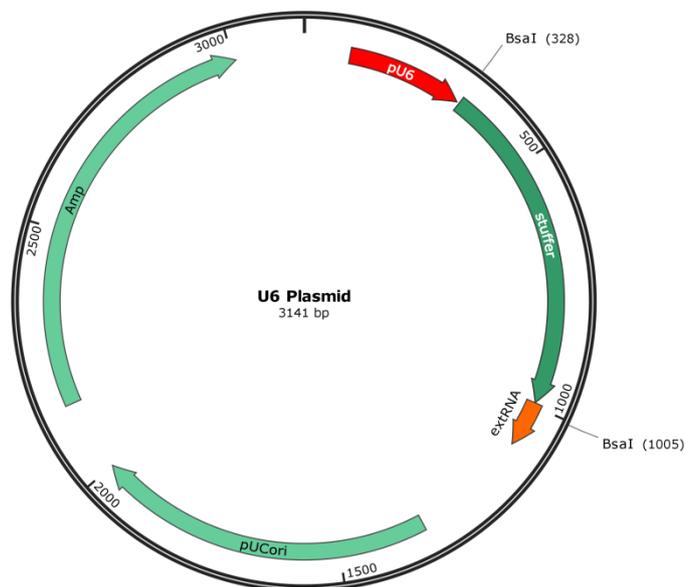


Figure 3-9 A schematic of the U6 plasmid, for the sub-cloning and transcription of CRISPR/Cas9 guides:

This plasmid has a U6 RNA polymerase III promoter to transcribe guides, shown in red. Following this, two asymmetric BsaI restriction sites flank a 'spacer region' shown in dark green, this being a 677 nucleotide fragment that allows for identification of accurate restriction digestion. Downstream of the 3' BsaI site, is the invariant tracrRNA sequence shown in orange, which is necessary for the reconstitution, correct functioning and activation of CRISPR/Cas9 guides. Finally, bacterial sequences of the plasmid backbone are denoted in light green. This plasmid was used for the sub-cloning of all guide sequences, which once cloned formed complete single guide RNA architecture.

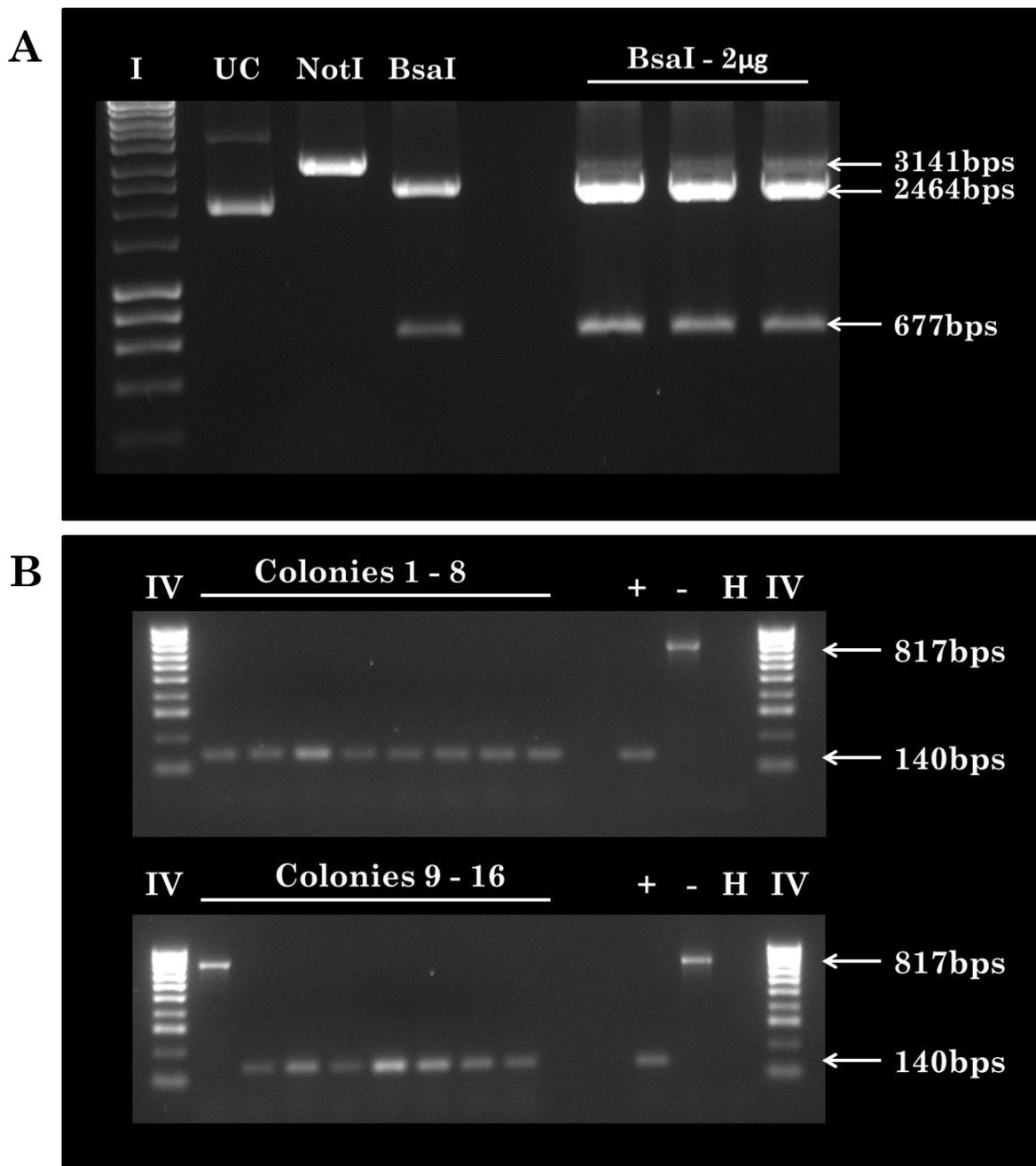


Figure 3-10 Guide-Cloning and subsequent colony PCR screening with the U6 guide plasmid.

A) U6 plasmid restriction digests. On the left there are internal controls: (UC) Uncut, (NotI) Linearise and BsaI digest all performed with 1µg of U6 plasmid. On the right 2µg BsaI restriction digests performed in triplicate to resolve a 2464bp destination vector backbone, this band was subject to Quiagen Gel purification whereas the 677bp spacer region was disregarded. Notably, there was some evidence of partial digest, where the spacer region had not been removed accounting for the faint 3141bp band of unmodified U6 plasmid, this produced some background post-cloning and transformation. All samples were loaded and resolved on a 1% agarose gel alongside hyper ladder I. B) The amplicons from a Colony PCR screen, with primers designed within the U6 plasmid and tracrRNA region. Amplicons positive for the sub-cloned guide sequence were 140bp, in contrast colonies negative for the guide sequence retained the spacer region and were 817bps in size. These were run alongside a positive control (+) a guide construct kindly generated by Dr. Taeyoung Koo and a negative (-) an empty U6 expression cassette and (H) water control. All samples were resolved on a 2% agarose alongside hyperladder IV.

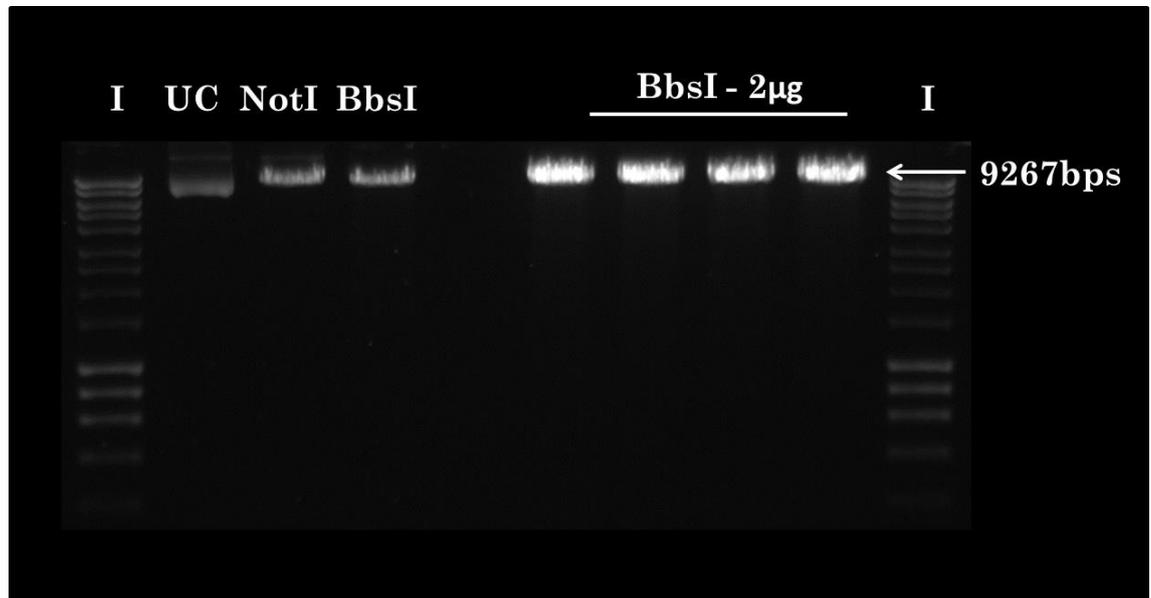
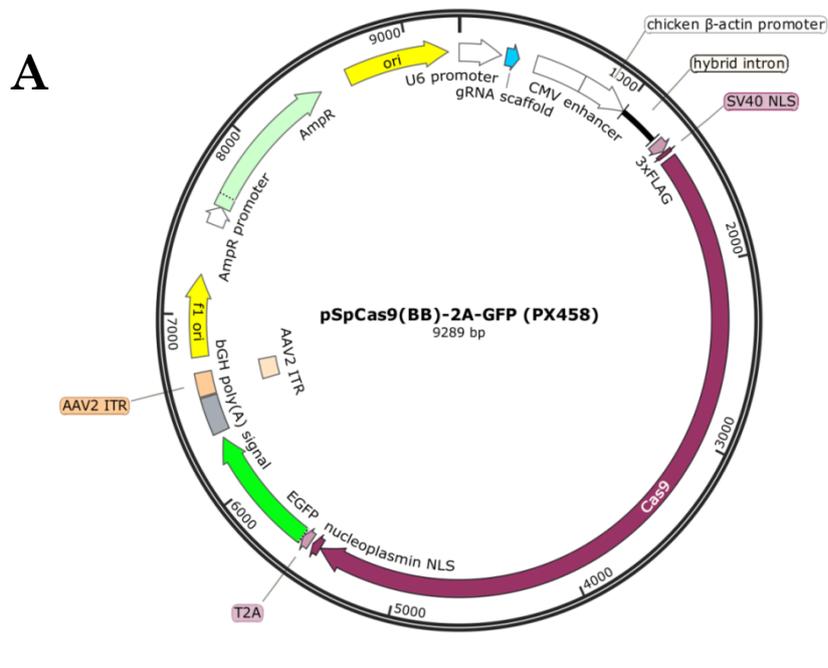


Figure 3-11 Guide Cloning into the U6-Guide-CBh-Cas9-2A-GFP (Px458) plasmid.

A) A schematic plasmid map of U6-guide-CBh-Cas9-2A-GFP (Px458), this all-in-one plasmid has the U6 RNA III polymerase U6 promoter upstream of two asymmetric BbsI restriction sites, separated by 22 nucleotides that were used to clone in guide sequences. Downstream of the CBh promoter, a modified version of the CAG promoter drives high SpCas9 and GFP expression, interrupted by a 2A peptide linker cleavage site. In addition this plasmid contains the bacterial backbone sequence and ampicillin resistance. B) Restriction digests of the Px458 plasmid for sub-cloning guide sequences. On the left internal controls were present including (UC) Uncut, (NotI) Linearise and BbsI all of which were carried out on 1µg of plasmid. On the right four replicates of the BbsI digests were undertaken. In the absence of the spacer, there was little distinction between the UC and BbsI digested plasmid. As a consequence, the subsequent colonies were not screened by PCR but rather by sequencing.

3.3.5 Retrospective Bioinformatics Analysis of the Specificity of Guide RNA Sequences Directed to the Homologous Region in DMD Intron 1:

To be established as a safe therapeutic strategy, genome editing techniques must be both precise and efficacious. Moreover, in approaches requiring the introduction of genetic material, the integration must occur with high fidelity. One concern relating to the use of endonucleases in the correction of disease causing genetic mutations is the potential for off-target effects. This refers to genome editing reagents having the potential to non-specifically cleave sequences with similarity to the desired target locus, in large eukaryotic genomes. Numerous strategies for improving CRISPR/Cas9 specificity have been developed by the field throughout the course of this investigation. These include: improved computational models based upon specificity-governing rule-sets (Benchling 2015; Haeussler et al. 2016), improved In-silico prediction of potential off-targets (Hsu et al. 2013b; Fu et al. 2013; Mali et al. 2013), improved Cas9 nuclease architecture resulting from protein engineering and novel delivery strategies (Slaymaker et al. 2016; Kleinstiver et al. 2016).

At the outset of this study, rule sets determining the tolerance of mismatches between RNA guide sequences and genomic DNA upon hybridisation were under investigation. Notably, a series of investigations were undertaken and specificity predicting computational tools were not yet developed (Fu et al. 2013; Pattanayak et al. 2013; Mali et al. 2013; Hsu et al. 2013b; Carroll 2013). As a consequence, provisional scoring from CRISPR MIT, a first generation CRISPR/Cas9 guide design tool, was used as an indication that designs should be pursued (Figure 3-2). However, when more advanced tools were developed, these were used to analyse human guide designs identified by CRISPR MIT more thoroughly. In this study, the CRISPOR tool was selected for this purpose (Haeussler et al. 2016); particularly as this tool gave more insight into potential off-target site positions within the human genomes, than its predecessor. Note this was performed for human guides as the investigation later only focused upon gene-editing in human *in vitro* cell based models (Table 2). The off-target sites identified by the CRISPOR tool were positioned either within exons, intergenic regions or introns. Furthermore, gene names were provided to facilitate more streamlined analysis of off-target loci. Finally, this tool also provided quantitative prediction of guide efficacy, which was not available at the time initial designs were undertaken.

Notably, guide 1 remained the best ranked with the fewest off-target sites and had the best predicted efficacy of the guides identified, this remained the most favourable for further development. Guide 4 remained the least favourable with the most off-targets and least efficacy predicted. Finally, all guide designs had a specificity score of above 50, which CRISPOR sets as a threshold for guide development.

For each guide a graphical plot indicating the number of off-targets was produced, containing 1 - 4 mismatches between the guide RNA and DNA predicted on- target site. These potential off targets were then further classified to their position in the genome (Figure 3-12 and Figure 3-13). As anticipated, as the number of mismatches increases the proportion of predicted off-target sites also increases, which can be attributed to the increased number of sequence permutations that may arise within the genome. This in combination with the preceding CRISPR MIT analysis indicated that the identified guides remained suitable for further development.

Position/ Strand	Guide Sequence + PAM + Restriction Enzymes + Variants	Specificity Score	Predicted Efficiency Show all scores		Out-of-Frame score	Off-targets for 0-1-2-3-4 mismatches + next to PAM
			Doench	Mor. Mateos		
Guide 1	AGACTAAGAACCAGATGCTA AGG Enzymes: <i>BstDEI, MwoI</i>	69	54	41	57	0 - 0 - 2 - 24 - 163 0 - 0 - 0 - 4 - 5 189 off-targets
Guide 2	TATTTTGCTTCCTTAGCATC TGG Enzymes: <i>LweI</i>	57	39	34	54	0 - 2 - 3 - 42 - 278 0 - 0 - 2 - 5 - 6 325 off-targets
Guide 3	GATGAAGCATGATATTATT TGG <u>T</u> Enzymes: <i>Tru1I</i>	61	19	17	61	0 - 0 - 4 - 33 - 204 0 - 0 - 2 - 3 - 13 241 off-targets
Guide 4	TCTGTCTAATTACATTATAA TGG <u>C</u> Enzymes: <i>SmiI, AaI</i>	57	36	13	56	0 - 1 - 5 - 26 - 212 0 - 0 - 3 - 7 - 5 244 off-targets

Table 2: An output table from the CRISPOR tool, showing the specificity and efficacy analysis of the human guides previously identified:

Analysis of the four guides identified with the CRISPR MIT tool. . Importantly, all of the guides have a specificity score of above 50, which is considered the threshold for inclusion in a scientific research project, using the algorithm developed by Hsu, et al 2013. Furthermore, their efficiency scores for two algorithms Doench, et al 2016 – largely used for U6 guide targeting strategies and Moreono-Mateos, et al 2015 – largely for transcribed RNA guides are included. Finally, the numbers of off-targets categorised by the number of mismatches they contain are present on the right hand side. Not all of these parameters were available at the time of the initial guide RNA designs, but nonetheless would not have prevented the guides being further developed.

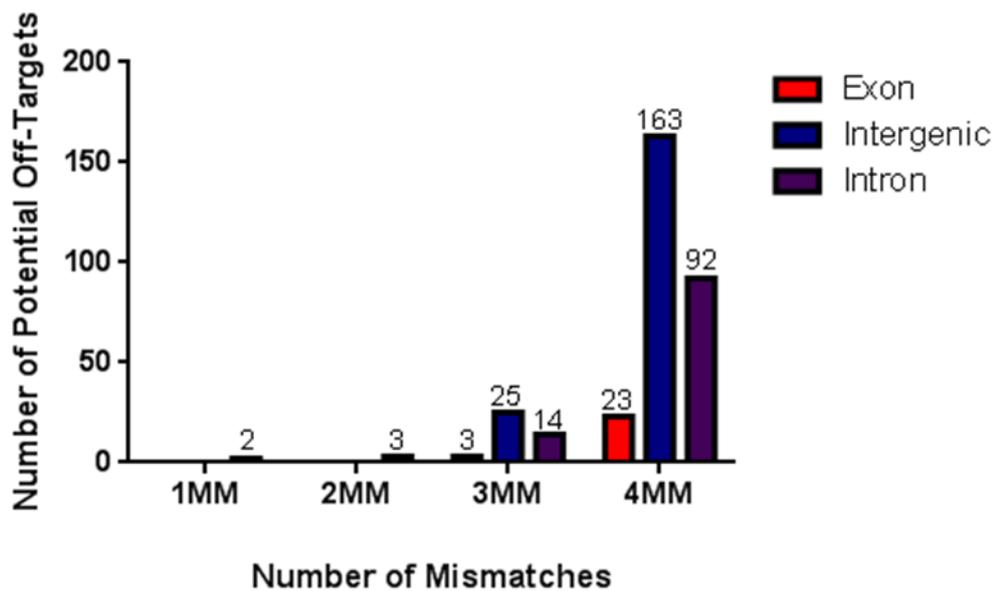
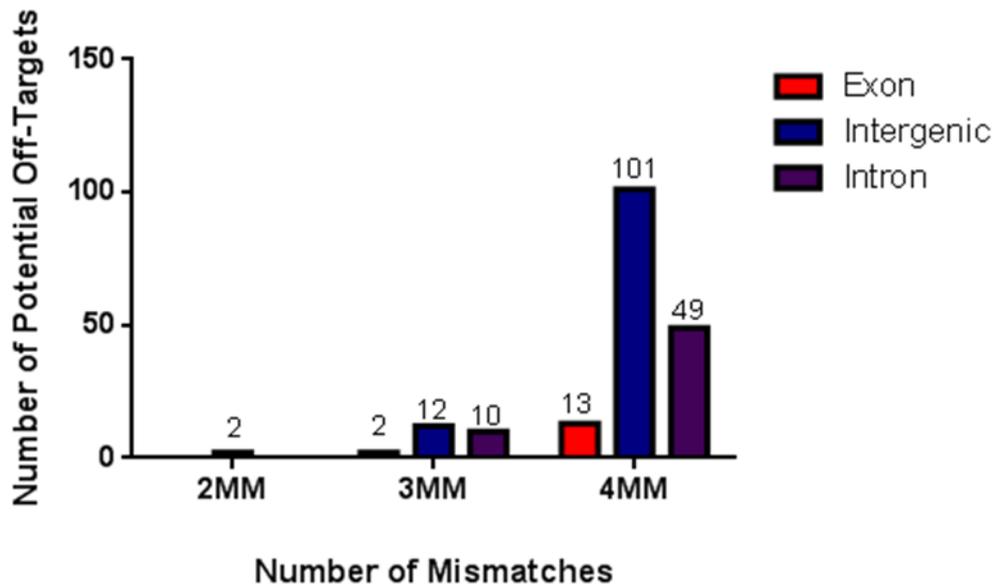


Figure 3-12 Graphical off-target profiles of Guide designs 1 and 2 divided into the location of their off-targets, being within: exons, intergenic or intron sequences.

(Top) Off-target profile of guide 1, the numeracy of mismatches (MM) is indicated along the bottom and their position in the genome being within the Exon (Red), Intergenic (Blue) or within an intron (Purple) is provided. No off-target sites with a single MM are identified and of the 26 off target sites with 3 MM or less, 22 are within intergenic or introns sequences and thus are not likely to affect the reading frame of genes. **(Bottom)** Off-target profile of guide 2. There are two off-targets with a single MM and 47 with 3 MM or less, of which 44 are within intergenic or intron sequences.

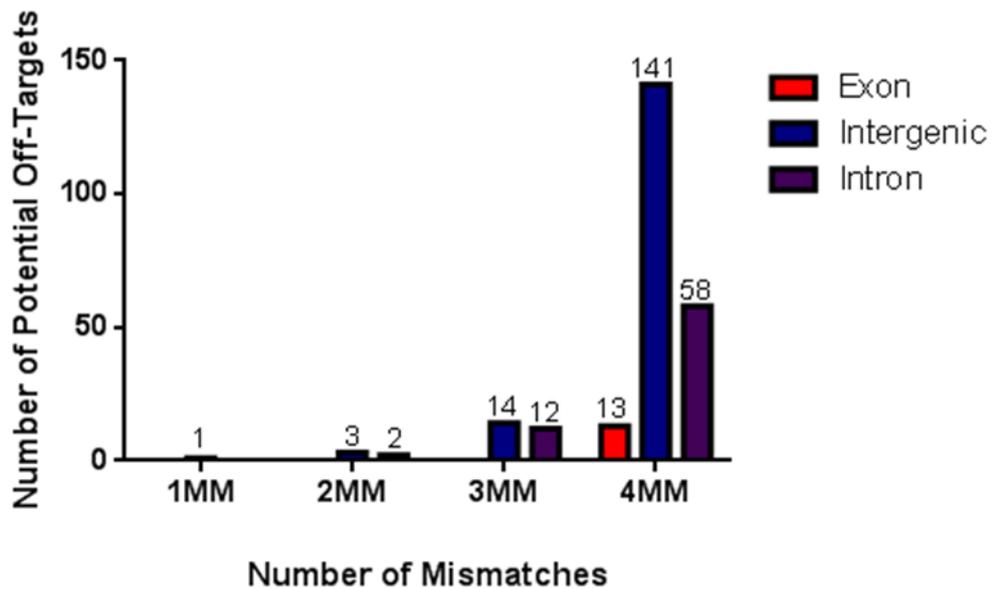
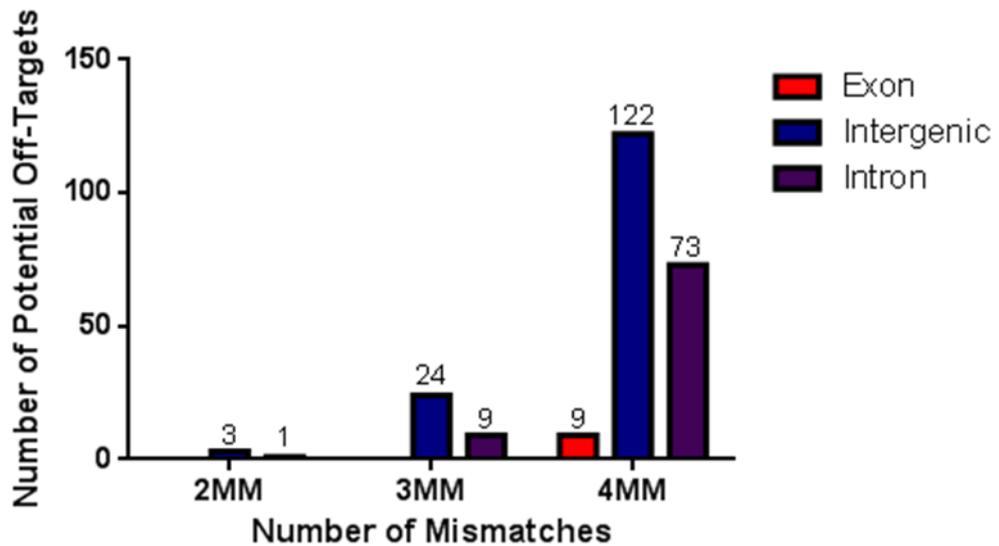


Figure 3-13 Graphical off-target profiles of Guide designs 3 and 4 divided into the location of their off-targets, being within: exons, intergenic or intron sequences.

(Top) Off-target profile of guide 3, the numeracy of mismatches (MM) are indicated along the bottom and their position in the genome being within the Exon (Red), Intergenic (Blue) or within an intron (Purple) is provided. No off-target sites with a single MM are identified and of the 37 off target sites with 3 MM or less, 34 are within intergenic or intron sequences and thus are not likely to affect the reading frame of genes. **(Bottom)** Off-target profile of guide 4. There is one off-target with a single MM and of the 34 off-targets with 3 MM or less, of which 33 are within intergenic or intron sequences.

3.4 Discussion

Intron sequences common to the full-length dystrophin isoforms for human and mouse were successfully identified and subjected to a pairwise sequence alignment. Due to the size distinction between the two intron sequences, a rigorous local alignment using the EMBOSS matcher tool from the EMBL/EBI suite was undertaken (McWilliam et al. 2013). In consideration of the extensive size of the *DMD* gene intron sequences, this alignment in isolation was far from exhaustive. The alignment produced may have obscured shorter stretches of high sequence similarity, which may have been of interest to this investigation. One way to address this could have been to use the sequence similarity tract within the UCSC genome browser, to examine the human *DMD* gene (Kent et al. 2002).

In contrast to the pairwise alignment of the human and murine *DMD* gene sequences presented; the UCSC genome browser would produce a sequence alignment for multiple species. This could be used to identify similar sequences between the human *DMD* gene and multiple larger animal models with more representative clinical pathology; including the Golden Retriever and the newly developed *DMD* rat models (Larcher et al. 2014; Brinkmeyer-Langford & Kornegay 2013). If candidate sequences with similarity between the human *DMD* gene and one or more animal models were identified; these could be subject to a rigorous multiple sequence alignment with the use of CLUSTAL; in a similar fashion to the analysis described (McWilliam et al. 2013). If undertaken this analysis would serve to identify further sequences suitable to the design of common endonucleases, which would be particularly useful in the context of TALENs. Further to expanding the repertoire of genome editing reagents if such sites were identified for larger animal models this would extend the translational scope of the investigation. A demonstration of the proposed therapy outlined in this study in a larger animal model with more comparable dystrophic pathology to *DMD* patients, would provide a beneficial insight into potential clinical benefit to patient cohorts.

The current alignment process however, was sufficient to identify a 345 nucleotide stretch of high sequence similarity between human and mouse, within the common intron 1 sequence. Importantly it was determined early in the investigation that endonucleases targeting the common intron 1 sequence, would be most beneficial to the therapeutic strategy proposed for three reasons. Firstly, intron sequences are deemed 'non-coding'; hence modification within this region will not disrupt an open reading frame of a gene. Thus it was anticipated an intron target site would circumvent the risk of producing further deleterious *DMD* gene mutations. and reduce the risks associated with insertional mutagenesis (Howe et al. 2008). Secondly, it was anticipated that targeting the most 5' non-coding intron of the *DMD* gene, would increase the applicability of this genome-editing approach to patients. The targeting of intron 1

described here, is estimated to have 95% applicability to patients, making this therapeutic strategy near-universal (Bladen et al. 2015). Notably any mutations occurring upstream of this region, particularly those within the dystrophin promoters, which have a documented association with X-linked cardiomyopathy, will not be amenable to correction via this approach (Muntoni et al. 1993). Thirdly, it was speculated that targeting intron 1 would improve the potential of the endogenous promoter and exon 1 elements, to splice to the dystrophin cDNA insertions, introduced by the exogenous repair template.

A 100 nucleotide region of this candidate sequence was then analysed for potential TALEN target sites and CRISPR guide designs. Endonuclease designs were found with the propensity to cleave genomic sequences common to both human and mouse species for both platforms, denoted as common gene editing reagents. Notably, the Collectis TALEN hit software provided no information pertaining to the specificity or efficiency of the single common TALEN design identified (Collectis Bioresearch 2011). However, if this construct was to be used a BLAST analysis could be undertaken to provide an insight into its potential off-target effects. In contrast, the CRISPR MIT software provided some information on guide specificity, with the list of predicted off-target sites derived from a BLAST analysis of the common guide sequence against the species genome (Optimised CRISPR Design 2013). Whilst an improvement, this tool provided little information as to the location of predicted off-targets within the genome and were not based upon specificity governing rules, which were later derived experimentally from examination of large numbers of guide designs (Hsu et al. 2013b; Fu et al. 2013; Pattanayak et al. 2013; Mali et al. 2013).

The limited information with regard to endonuclease specificity and efficacy, coupled with the versatility of targeting the CRISPR/Cas9 gene editing platform, prompted the examination of alternative endonuclease sites at the 5' end of the *DMD* gene. Interestingly, the alignment of the first three exons of the Dp427m isoform highlighted high sequence similarity at exon 2 for both human and mouse sequences. Importantly, 8 TALEN designs identified had the propensity to cleave both sequences; this could represent an interesting and alternative genomic target site for development. Whilst the original concerns for targeting exon sequences were the production of further deleterious mutations within the genome, the risk of this could be outweighed by the potential benefit, particularly if corrected colonies were screened on an individual basis, clonally expanded and engrafted. Thereby following an *ex vivo* modification and subsequent engraftment strategy. If the exon 2 sequence is conserved between multiple species, this could also provide the benefit of enabling the translation of this therapeutic approach to larger animal models. Furthermore, if the synthesis of TALENs was optimised all 8

TALEN pairs could be screened to identify the common gene-editing reagent with the highest efficacy.

In addition, to the strategy of addressing sequences of high similarity between human and mouse genomes, the full common intron 1 was screened in a species specific manner. This was performed to further expand the repertoire of guide designs, particularly as variable efficacy of guides to direct Cas9 cleavage was anticipated and later confirmed by a variety of studies (Doench et al. 2014; Doench et al. 2016; Liu et al. 2016a; Moreno-Mateos et al. 2015). It was speculated that multiple high-ranked guide designs in close proximity would be beneficial to the targeting of the intron 1 sequence and thus this was the criteria adhered to. Moreover, increased designs to the intron 1 sequence would maintain the proposed targeting strategy but also allow for higher efficacy guides to be identified. However, it has since been indicated by others that high numbers of PAM motifs in close proximity can be detrimental to guide efficacy (Malina et al. 2015). As a consequence whilst specific designs continue to hold merit, if this approach was to be developed further, additional bioinformatics analysis with newer in-silico prediction tools would be required.

Although further guide designs and target sequences were identified to ensure the most successful outcomes for the project, the four guides designed to the region of homology had provisional specificity profiles from the CRISPR MIT tool that made them favourable for further development and assessment. Hence these became the foundational designs utilised throughout this study. Importantly however, during the course of this investigation, other research groups extensively screened guide libraries to identify guides with high specificity and efficacy (Hsu et al. 2013b; Doench et al. 2014; Doench et al. 2016; Liu et al. 2016a). The subsequent findings identified guide RNA specificity and efficacy-governing rule sets that were used in the development of computer algorithms, to facilitate optimal guide selection. Once available the original four guide designs were re-assessed with the CRISPOR.org tool, which utilises these algorithms (Haeussler et al. 2016). Interestingly, their specificity profile still indicates that the development of these guides would be feasible in the context of this project. Furthermore, this tool provided more insight into the location and future screening of off-target sites.

If time had permitted it, off-target analysis would have been undertaken in a biased manner for off-targets with up to three mismatches. In this instance, primers would have been designed to a predicted off-target site and the subsequent amplicons are analysed by a mismatch detection assay (Qiu et al. 2004; Vouillot et al. 2015; Tycko et al. 2016). A biased approach would be opted for due to its increased scalability. However, if the proposed

therapeutic strategy of this investigation was to be further developed a more robust and/or stringent genome-wide assessment using a variety of newly developed methods would have to be undertaken, including: Targeted deep gene sequencing (Ran et al. 2013), Guide-seq (Tsai et al. 2014) or Digenome seq (Kim et al. 2015). This is particularly pertinent as it has been indicated that as much as up to 6 mismatches can be tolerated by the CRISPR/Cas9 system (Tsai et al. 2014).

Nonetheless, the guides designed to this region were selected for further development, to improve the near-universal applicability of this approach. If successful it is speculated that the integration of the dystrophin transgene at the 5' end of the *DMD* gene will ameliorate the pathology associated with a wide range of mutations. Notably however, internal dystrophin isoforms such as the Dp260, Dp140, Dp116 and Dp71, will not be reconstituted by this process. These isoforms are non-muscle expressing isoforms, implicated in the non-progressive cognitive impairment; in light of this some of the cognitive symptoms displayed by DMD patients may not be addressed by this therapeutic strategy. These constraints noted, the muscle pathology is the most debilitating aspect of the disease, which ultimately leads to premature death (Bushby et al. 2010a). Furthermore, by targeting the common intron 1 region, it is speculated that in addition to the muscle isoform, the cortical, Purkinje 1 and 2 full length isoforms will also be reconstituted and expressed at physiological levels.

Thus this near universal strategy could provide permanent correction of a range of dystrophin mutations and be therapeutically beneficial for a significant proportion of patients addressing a clear unmet medical need. The few mutation types not encompassed by this near-universal strategy, could be addressed later by mutation-specific means in the future or by gene therapy approaches such as utrophin up-regulation or MD1 gene addition (Tinsley et al. 2011; Guiraud, Squire, et al. 2015; Le Guiner et al. 2017).

3.5 Conclusion

This chapter presents a series of pragmatic design strategies to identify suitable TALEN and CRISPR guides for the introduction of a DSB, at the 5' end of the *DMD* gene. These designs focused upon areas of high sequence similarity between human and mouse genomes at both exon and intron sequences, but also upon species-specific designs across the common intron 1 sequence. Notably, a series of designs were undertaken to increase the likelihood of the project having successful outcomes. This stated, it was resolved that targeting the intron 1 sequence common to all full length isoforms, using CRISPR guides would facilitate the development of the therapeutic strategy proposed. However, in light of this CRISPR/Cas9 bacterial immune system being newly re-purposed for mammalian genome editing, TALEN

designs were also undertaken, to permit the use of the established TALEN endonuclease platform if required.

Chapter 4 Establishing a Delivery System for the Cas9 Nuclease and Validation of Guide RNA Efficacy at the Human *DMD* Intron 1 Site

4.1 Introduction

Successful gene editing is reliant upon the delivery of customisable nucleases into mammalian target cells and the subsequent ability of the endonuclease to introduce double stranded breaks (DSBs) at a specific genomic locus. Notably, the occurrence of DSBs at the genomic locus dictate the extent to which NHEJ or HDR DNA repair pathways can be exploited to modify the mammalian genome and by extension address or compensate for disease-causing mutations (Ciccia & Elledge 2010; Liang et al. 1998; Jasin & Rothstein 2013). Thus in the context of this investigation which aims to utilise the low efficiency HDR DNA repair pathway (Mao et al. 2008; Heyer et al. 2010) to mediate an MD1 transgene integration, highly efficient delivery of gene-editing reagents and identification of guide RNAs that introduce DSBs with the high efficacy will be crucial to attaining successful outcomes. It is important to establish that throughout this investigation a single guide RNA (sgRNA) scaffold was used comprising of two part a 20 nucleotide crRNA and an invariant 83 nucleotide tracrRNA sequence. Notably, the crRNA associates with Cas9 nuclease and directs it to DNA via Watson-Crick complementary base pairing. This 20 nucleotide sequence within the sgRNA is referred to as “guide” RNA sequence throughout the field.

Interestingly, publications near the start of this investigation highlighted the inherent variability in the ability of CRISPR guide RNAs to direct the Cas9 nuclease in producing DSBs and the importance of screening multiple guide RNA designs targeting the genomic locus (Mali, Yang, et al. 2013; Ran et al. 2013). Established methods for assessing the efficacy of endonucleases use the low fidelity of the NHEJ repair pathway and the resultant small insertions or deletions (InDels) as a proxy for gene-modification at a locus (Ciccia & Elledge 2010; Qiu et al. 2004). Numerous methods have been developed to assay the diverse resulting InDel spectrum including mismatch detection assays such as High Resolution Melt Curve Analysis (HRMCA) (Vossen et al. 2009), Surveyor (Qiu et al. 2004) and T7 Endonuclease I (T7E1) (de Massy et al. 1987; White et al. 1997) assays, sequencing assays including deep gene sequencing, tracking of InDels by decomposition (TIDE) (Brinkman et al. 2014) and restriction digest assays in the form of Restriction Fragment Length Polymorphism (RFLP) (Jeffreys 1979; Kim et al. 2014). These methods have differing inherent capacity to identify InDel types and quantify the frequency of InDel occurrence. Moreover, in many instances the DNA termini following the introduction of DSBs are often ligated perfectly, as a consequence these

methods provide an underestimate of the amount of DSB present in situ (Ciccia & Elledge 2010).

Importantly, the in-silico guide RNAs designed to DMD intron 1 will introduce DSBs and InDel modifications to varying degrees. This process has since been elucidated to be largely influenced by guide sequence determinants (Xu et al. 2015; Liu et al. 2016b; Doench et al. 2014), local chromatin structure (Malina et al. 2015), the ability of the guide to form secondary structures once transcribed (Moreno-Mateos et al. 2015) and the format in which the guide RNA is delivered, whether it be encoded within plasmid vectors, or directly in an RNA (Doench et al. 2014; Moreno-Mateos et al. 2015; S. Kim et al. 2014). Thus this chapter will be concerned with trialling alternative gene-editing reagent delivery strategies and the identification of the most efficacious in-silico guide RNA design.

4.2 Chapter Aims:

The primary aims of this chapter were to assert successful delivery and expression of gene-editing reagents by transient transfection into HEK293T cells and to establish a functional gene-modification assay, to enable the detection of InDels. These methods were then employed to assess the efficacy of guide RNAs designed to target intron 1 of the *DMD* locus. Once the delivery, expression and efficacy of the gene-editing reagents were demonstrated in HEK293T, these approaches were then transitioned to $\Delta 45-55$ immortalised patient myoblasts. As a secondary aim, the Cas9 nuclease was also produced and delivered in mRNA and protein formats, as it was speculated this would improve the efficiency of the system and by extension the approach being trialled within the immortalised patient myoblasts.

4.3 Results:

4.3.1 An Examination of Cas9 Nuclease Delivery and Expression using HEK293T Cells as a Model System:

In an attempt to achieve successful delivery and subsequent expression of the Cas9 nuclease, two plasmids CBh-Cas9 and U6-Guide1-CBh-Cas9-2A-GFP (For plasmid maps see Figure 4-1), hitherto referred to as Px165 and Px458 respectively, were delivered to HEK293T by transient transfection. Notably Cas9 expression in both cases is driven by the chicken beta-actin hybrid promoter (CBh), reported to give robust expression in a range of mammalian cell lines (Gray et al. 2011). Furthermore the Px458 utilises a T2A self-cleaving peptide between Cas9 and a GFP transgene [for a review of 2A linkers, see (Liu et al. 2017)]. The outlook here was to optimise a western blotting protocol to assess the expression of Cas9 nuclease by harvesting total protein lysate 48 hours post transfection. Upon establishing a suitable and robust methodology, dose responses of the gene-editing reagents were undertaken to identify a dose that provides suitable expression and to assess the cell culture for potential toxicity. In the case of Px458, GFP could be used as a proxy for Cas9 expression with the proportion of GFP positive cells as assessed by FACS analysis indicating the number of cells where Cas9 gene delivery and expression has been successful.

4.3.2 Establishing a Western Blotting Protocol for the Detection of Cas9 Protein.

In an effort to establish a suitable western blotting procedure for Cas9 protein, 2 μ g of the Px165 plasmid was transiently transfected into HEK293T, using a lab established 5:1 Viafect transfection reagent (μ l): DNA (μ g) ratio. Protein was harvested 48 hours post transfection and 50 μ g total protein lysate was subjected to Western blotting with the Diagenode Cas9 antibody. Initially, antibody dilutions ranging from 1:1000-1:5000 were trialled in the Western blotting procedure (Figure 4-2). This was performed with the aim of identifying Cas9 protein expression and obtaining a Western blot with low background. Importantly a protein of 150kDa corresponding to the anticipated size of Cas9 protein was present in each of the dilution series. This stated, the protein did not resolve as a sharp distinct band, but rather a smear, with marginal differences observed in the resolution and background in the dilution series trialled (Figure 4-2-A).

The 1:5000 dilution of the Cas9 antibody was selected for further optimisation investigations. The next parameter to be altered related to the amount of total protein lysate loaded to the 3-8% Tris-acetate gel during the Western blot procedure. To this end, 25 μ g and 12.5 μ g total protein lysate respectively were loaded and assessed for Cas9 expression (Figure 4-2-B). The 12.5 μ g total protein lysate loading condition yielded a distinct band, albeit with a large surface area, after 3 seconds exposure using the HRP method. This indicated that the Px165 plasmid

was expressing Cas9 protein. Furthermore, the results suggested that for future detection of Cas9 protein a 1:5000 dilution of the Diagenode Cas9 antibody and 12.5µg of total protein lysate should be utilised.

4.3.3 Examination of Cas9 Protein Expression Following Transient Transfection in HEK293T Cells with Dose Escalation.

A dose escalation study was performed for both Px165 and Px458 plasmids, to evaluate 4µg, 2µg, 1µg and 0.5µg doses. The transient transfection was undertaken in parallel for both plasmids, total protein harvested 48hours post-transfection, and 12.5µg of total protein lysate was subject to western blotting. In addition, microscopy imaging and FACS analysis for eGFP expression was undertaken in cells treated with Px458.

As anticipated the western blots show both the Px165 and Px458 show a dose dependent expression profile (Figure 4-3). Importantly bands corresponding to the 150Kda Cas9 protein were in both transfection series. Notably doses between 1µg and 4µg produced strong bands, whereas at the 0.5µg dose bands are absent in the Px165 and only marginally detected in the case of Px458. Interestingly, additional bands were present in the dose ranges of both plasmids. In the case of P165, below the expected 150kDa band, two additional bands were present estimated to be 140 and 130kDa in size, which was more pronounced in 4µg and 2µg doses. In contrast in the case of Px458, two bands were present both above and below the band corresponding to Cas9 protein. The upper two were estimated to be 177kDa and 167kDa and the lower two of 140kDa and 130kDa respectively. The lower bands, consistent across both samples, were speculated to be degradation products. In contrast the larger bands observed for Px458 were anticipated to reflect Cas9 and GFP fusion proteins that have not undergone 2A peptide cleavage.

The dose dependent expression of Cas9 nuclease was mirrored in fluorescence microscopy and FACs analysis of the Px458 samples, which were undertaken alongside a pCI-GFP positive control. This plasmid was selected as a positive control as it utilises a strong viral Cytomegalovirus (CMV) promoter to drive green fluorescent protein (GFP) expression (Figure 4-4). The mean percentage of GFP⁺ cells for each condition were as follows: 0.5µg 6.50% ± 0.04, 1.0µg 40.23% ± 0.44 , 2.0µg 71.70% ± 2.09 and 4.0µg, 92.67% ± 3.00 (N=3: mean ± SEM in all cases). A significant difference was detected between doses of 0.5 and 1µg (P>0.001), 1 and 2µg (p=0.023) and 2 and 4µg (p=0.0143). Interestingly, there was no significant difference between the GFP⁺ population at 4µg dose of Px458 and 2µg dose of pCI-GFP both giving expression levels of almost 100%.

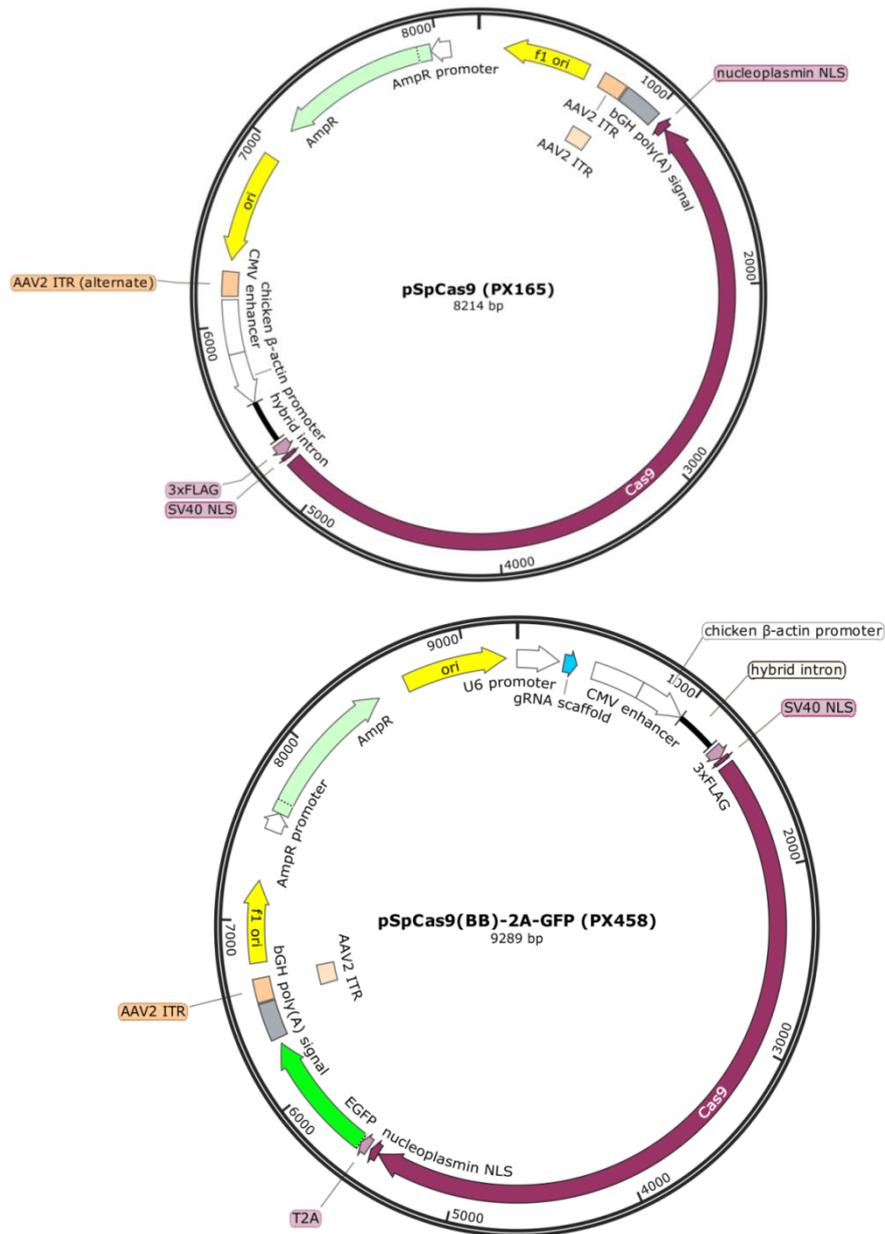


Figure 4-1 Schematic of *Streptococcus Pyogenes* Cas9 expression plasmids used in transfection efficiency and gene-modification investigations; both of which were obtained from Addgene.

The Px165 construct (Top) comprises: the Chicken Beta Actin (CBh) promoter, a modified CAG promoter with validated robust expression in a range of mammalian cells, immediately flanked by human sequence-optimised SpCas9, terminated by a Bovine Poly A signal. This construct requires dual transfection alongside a U6-guide plasmid to provide the guide expression necessary for mediate gene-modification. The Px458 construct (Bottom) contains the same CBh promoter and Cas9 components as the Px165. However, it also contains a U6 promoter upstream with a sgRNA scaffold in place allowing for a guide sequence to be directly cloned with the use of BbsI, a type II restriction enzyme, circumventing the requirement for dual transfection. In addition, a GFP marker terminating on a Bovine Poly A signal is present downstream of the SpCas9 using a T2A promoter to aid assessment of transfection efficiency. Both plasmids harbour ampicillin resistance cassettes, to facilitate bacterial transformation and selection.

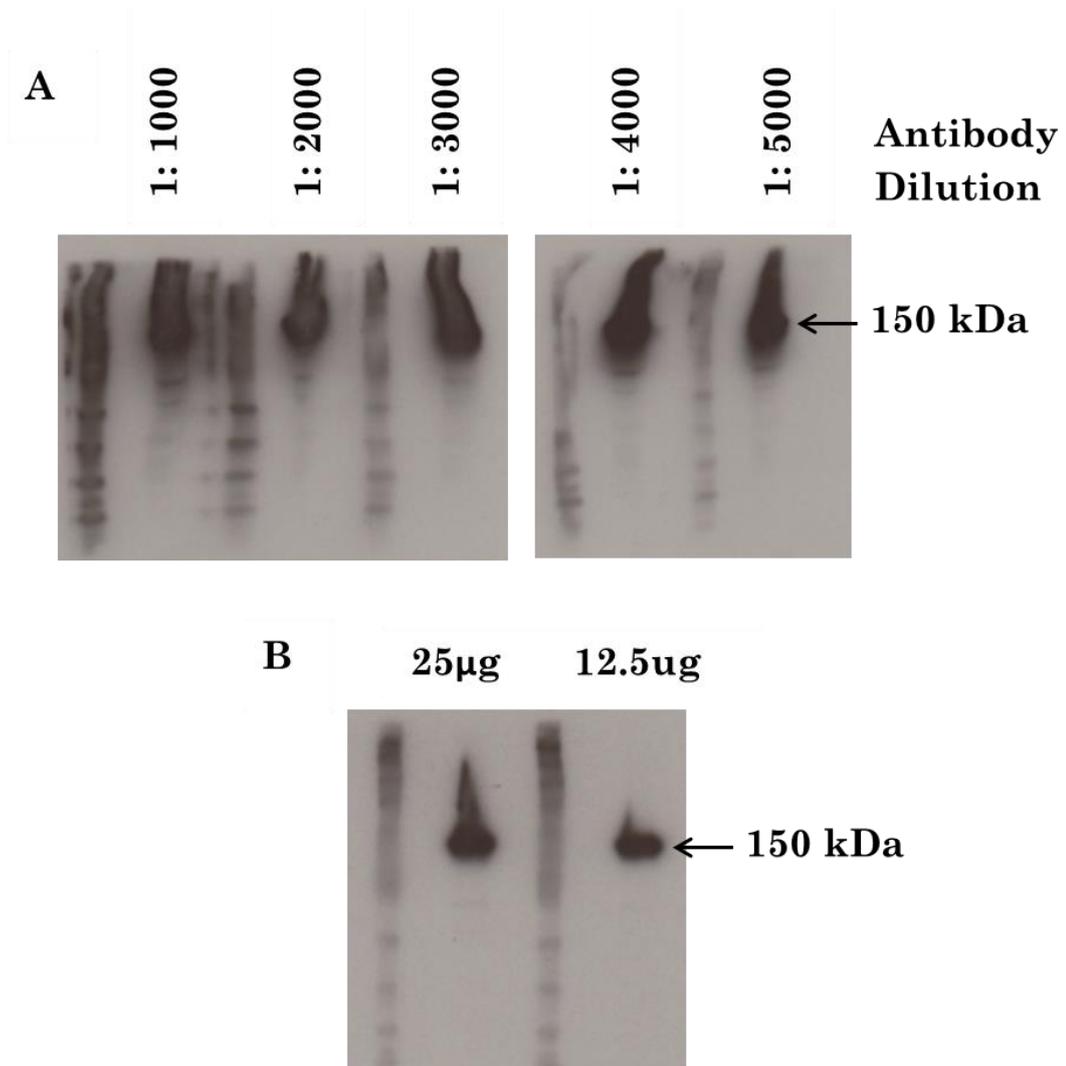


Figure 4-2 HRP Western Blotting undertaken to optimise the detection of Cas9 protein, using the Diagenode Cas9 antibody.

A) Western blotting was performed on 50µg total protein lysate harvested from HEK293T cells 48hours post transfection of 2µg of Cbh-Cas9 (Px165) using 5:1 Viafect reagent (µl): DNA (µg) ratio, using serial dilution ranges of 1 in 1000- 1 in 5000 of the Diagenode Cas9 antibody raised to the N-terminal of Cas9 in mouse. **B)** Western blotting of samples from A diluted to 25µg and 12.5µg total protein lysate, using a 1 in 5000 dilution of the Diagenode antibody. Samples were resolved on a 3-8% Gradient Tris-Acetate gel, alongside NuPage HiMark ladder, prior to transfer to a 0.45µM nitrocellulose membrane. The samples were then incubated for an hour in the presence of a horse radish peroxidase (HRP) conjugated goat anti mouse secondary antibody, prior to the blot being developed by chemiluminescence for 1 minute. Film (GE healthcare) was then positioned over the blot and imaged after a 3 second exposure time.

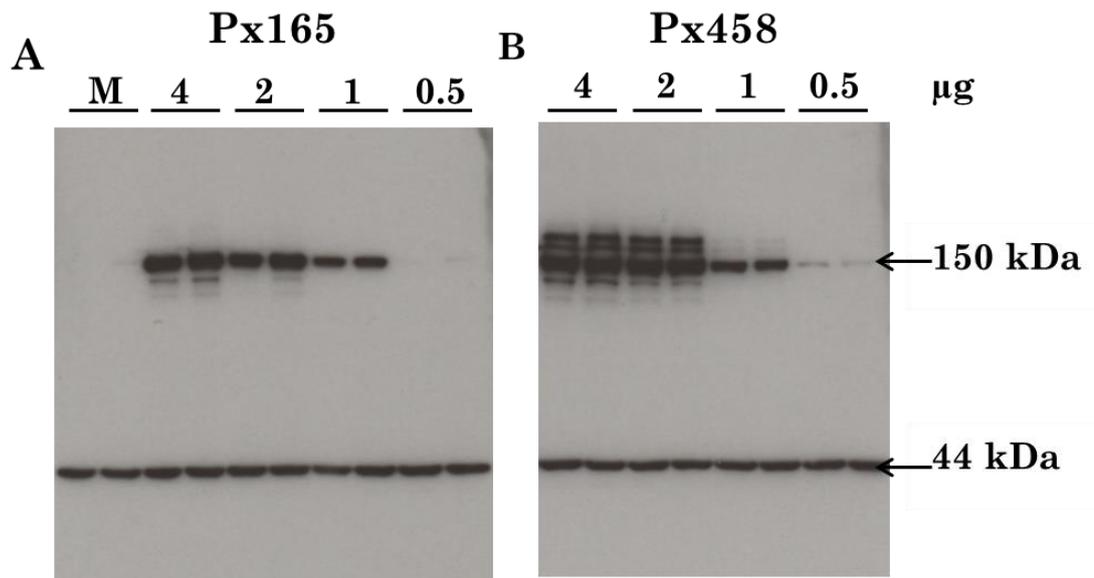


Figure 4-3 HRP Western blotting to detect Cas9 protein in HEK293T culture, subjected to a dose escalation of Px165 and Px458 Cas9 expression plasmids.

Western blotting was performed on 12.5μg total protein lysate harvested from HEK293T cells 48hours post transfection of Cas9 expressing plasmids **(A)** Px165 and **(B)** Px458 at doses of 4μg, 2μg, 1μg or 0.5μg using 5:1 Viafect reagent (μl): DNA (μg) ratio. Protein samples were assessed for Cas9 expression using a 1 in 5000 dilution of the Diagenode Cas9 antibody. Loading control was assessed using of an alpha tubulin antibody (Abcam 40774) at a 1:2500 dilution. Cas9 protein at the 150kDa and Alpha tubulin at 44kDa are indicated. Samples were resolved on a 3-8% Gradient Tris-Acetate gel, alongside NuPage HiMark ladder, prior to transfer to a 0.45μM nitrocellulose membrane. The samples were then incubated for an hour in the presence of a horse radish peroxidase (HRP) conjugated secondary antibody, goat anti mouse (Cas9) or a goat anti-rabbit (alpha-tubulin), prior to the blot being developed by chemiluminescence for 1 minute. Film by GE healthcare was then positioned over the blot and imaged after a 3 second exposure time in both instances.

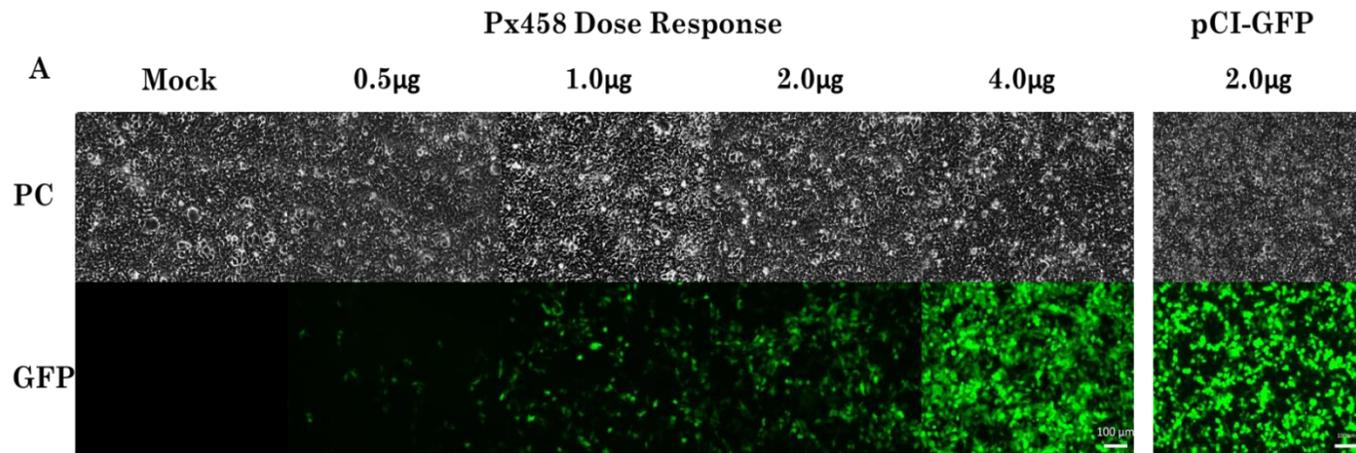
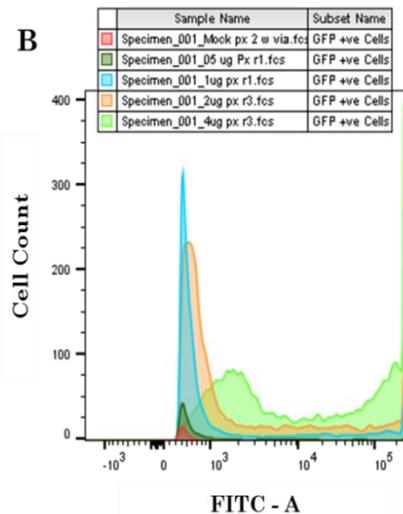
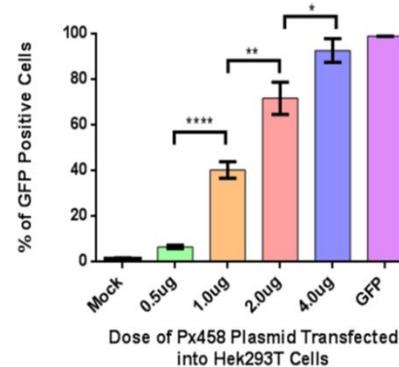


Figure 4-4 Microscopy Images of HEK293T Cells 48 hours post transient transfection Px458 plasmid at increasing doses, alongside a 2µg pCI-GFP control.

(A) Representative Microscopy images of HEK293T transfected with the Px458 plasmid at doses of 4, 2, 1 and 0.5µg using 5:1 Viafect reagent (µl): DNA (µg) ratio, 48 hours post-transfection. Transient transfections of Px458 were undertaken alongside mock samples, in which cells were treated with 20µl of Viafect transfection reagent in isolation and 2µg of pCI-GFP as a positive control. **(B)** A histogram overlay of FITC-positive cells determined by FACS analysis and produced by FlowJo software, for the transfected populations from **A)** harvested 48 hours post transfection into 2% PFA FACS buffer. Samples: Red: Mock (1% gating), Dark green: 0.5µg, Light blue: 1µg, Orange: 2µg and Light Green: 4µg respectively. **(C)** A bar chart of mean percentage GFP positive cells, determined by FACS analysis for each of the transfected Px458 doses, with error bars representing standard error of the mean. Differences between doses transiently transfected were deemed significant by means of an unpaired t-test: comparisons 0.05µg-1µg $P < 0.0001$, 1µg -2µg $p = 0.023$ and 2µg-4µg $p = 0.0143$.



C A Bar chart showing the percentage of GFP positive Hek293T cells, 48 hours post transient transfection with various Px458 Doses



4.4 Establishing the Surveyor and T7E1 Endonuclease (T7E1) Cleavage Assays for the Detection of Targeted Genome Modification.

4.4.1 Efficient PCR Amplification of the Human *DMD* and Mouse *Dmd* Intron 1 Locus

In order to evaluate guide RNA efficacy in producing DSBs, this being how efficiently in silico guide designs identified can direct the Cas9 nuclease to introduce DSBs at the *DMD* intron 1 locus, PCRs were designed and optimised spanning the region of homology. These PCRs used primers 1960F and 2741R or 1960F and 2541R for the human *DMD* intron 1 locus and 1950F and 2775R for mouse. In all instances a single amplicon band was obtained when 250ng of untreated HEK293T DNA was amplified using the high fidelity polymerase Hotstart phusion flex.

Importantly, for Surveyor or T7E1 mismatch detection to be successful, PCR amplicons need to be asymmetric with regard to the predicted cleavage point of the guide RNA under examination. This is due to the inherent mechanism of mismatch detection assays which rely upon the hybridisation of wild-type and modified amplicon sequences, the latter harbouring InDels as a result of subsequent DNA cleavage and repair. Importantly, these sequences hybridise imperfectly as a result of the sequence modification and heteroduplex or 'loop' structures form that can be cleaved by enzymes such as Surveyor and T7E1 (Figure 4-5). For cleavage products to be successfully resolved via gel electrophoresis, the heteroduplex should be positioned nearer to one terminus of the amplicon. The two PCRs designed for the human intron 1 site are a case in point, with 1960F and 2741R having suitable asymmetry for guide designs 1, 2 and 4, and 1960F and 2541R for guide 3.

Notably, the outlook was to trial all human and mouse guides and assess for efficacy. However, due to the considerable time taken to optimise the *DMD* intron 1 PCRs, and Surveyor and T7E1 mismatch detection assays, the investigation was focused to human *DMD* intron 1 from this point onwards.

4.4.2 Efficacy of Human Guides and Optimisation of the Surveyor and T7E1 Enzyme Mismatch Detection Assays.

4.4.2.1 Co-Transfection of U6 Guide and Cas9 Expressing Px165 Plasmids into HEK293T

Once PCRs asymmetric to predicted guide RNA cleavage points were optimised (Figure 4-6), U6 expression plasmids encoding human guide RNAs 1, 2, 3 or 4, kindly cloned by Dr Taeyoung Koo, were co- transfected alongside the Cas9 expressing Px165 plasmid into HEK293T cells. These plasmids were delivered at either a 1µg:1µg or a 2µg:2µg dose, using a 5:1 Viafect volume: DNA mass transfection ratio in a 6 well plate. DNA was harvested 48 hours post

transfection and 250ng of DNA amplified with primers 1960F and 2741R for guides 1, 2 and 4, or primers 1960F and 2541R for guide 3, alongside mock and Cas9 alone controls.

These samples were then subject to Surveyor and T7E1 mismatch detection assays and resolved on 2% agarose alongside hyperladder IV. Importantly, varying volumes of amplicon 6 μ l, 9 μ l and 12 μ l were utilised in the Surveyor assay, whereas the latter T7E1 was trialled with 200ng of PCR purified amplicon (Figure 4-7). Importantly, faint products consistent with the sizes of the anticipated cleavage fragments were observed (see Table 3) on the agarose gels for the 1 μ g:1 μ g and 2 μ g:2 μ g conditions in both Surveyor and T7E1 mismatch detection assays. This intensity of cleavage products was however a lot more prominent in the higher 2 μ g: 2 μ g dose for both assays, suggesting a dose dependent response of gene modification. Finally, the cleavage products did not appear to differ across the varying amplicon volumes utilised in the Surveyor assay.

Guides	Anticipated Band Fragments
1	512 (uncut), 288 and 224 (Cut)
2	512 (Uncut), 289 and 223 (Cut)
3	582 (Uncut), 332 and 250 (Cut)
4	512 (Uncut), 319 and 193 (Cut)

Table 3 Anticipated bands of the uncut amplicon and cleavage products should InDels be present at the location of the cleavage site of the guide RNA designs.

**Notice that guides 1, 2 and 4 sizes corresponding to the amplicon resulting from the PCR primer pairs 1960F and 2571R and guide 3 sizes corresponding to 1960F and 2541R*

4.4.2.2 Quantification of Gene Modification Efficiency and Polyacrylamide Gel Electrophoresis

The intensity of both uncut amplicons and cleavage products from agarose gels in Figure 4-7 was estimated using Image J analysis. The resultant values were used to attain estimates of the frequency of gene modification in accordance with published equations and methods (Guschin et al. 2010). This method was used throughout this chapter to provide gene modification frequency estimates. In accordance with these calculations, guide RNAs 1, 2 and 4 were estimated to give respectively, 11%, 5% and 12% gene modification frequency in 2 μ g:2 μ g conditions. Unfortunately band resolution using an agarose gel was insufficient to attain gene modification percentage in the 1 μ g:1 μ g condition. Finally, these co-transfections were performed as single treatments only to establish the functionality of the surveyor and T7E1

mismatch detection assays, thus a mean and standard error of the mean could not be calculated in this instance.

In an effort to enhance the signal from the T7E1 mismatch detection assays, 400ng of amplicon was subject to the T7E1 cleavage and resolved on a 10% TBE polyacrylamide electrophoresis gel and visualised by 1 hour incubation in a 1/5,000 dilution of SYBRsafe in 1x TBE buffer (Figure 4-8). This type of gel electrophoresis was selected to improve the resolution and quantification of cleavage fragments. In this system the percentage of gene modification for guide RNAs 1, 2 and 4 was determined to be 7%, 3% and undetectable at the 1 μ g:1 μ g and 27%, 7% and 25% at the 2 μ g:2 μ g treatments respectively.

In consideration of the findings thus far, the assessment of guide RNA 3 was undertaken with primer set 1960F-2541R using 400ng of amplicon and resolved on a 10% TBE gel (Figure 4-8). The gene modification under the 2 transfection conditions was estimated to be 5.5% and 9.2%. However, the accuracy of this estimation was likely influenced by the close proximity of non-specific bands present on the gel.

The efficacy of guides RNAs estimated under a range of transfection conditions in the T7E1 assay on both 2% agarose and 10% TBE polyacrylamide gels are tabulated in Table 4, and presented graphically in Figure 4-9. Notably, gene modification for 1 μ g: 1 μ g samples were only quantifiable on the 10% TBE polyacrylamide gels. Furthermore, the estimates of gene modification of the 2 μ g:2 μ g samples were higher when resolved on the 10% TBE polyacrylamide gel. The guide RNAs 1, 2 and 4 gave estimates on agarose gels of 11%, 5% and 12%, in contrast to 27%, 7% and 24% by 10% polyacrylamide gels.

These results indicate that the use of high resolution polyacrylamide gels to identify T7E1 cleavage products is probably the best system but are more complicated and costly, and it was decided that agarose gels would continue to be used routinely for initial indications of the occurrence of InDels by T7E1 assay. Furthermore, these attempts to establish a mismatch detection assay highlighted that the T7E1 assay was easier to optimise, perform and reproduce than the Surveyor assay. Thus the T7E1 system was selected and utilised from this point onwards in the investigation. Finally, the trend in the data suggests that guide RNA 1 introduces DSBs with the highest efficacy out of the various in-silico designs.

	Agarose 1:1	10% TBE 1: 1	Agarose 2: 2	10% TBE 2: 2
Guide 1	NQ	7.20	11.1	27.1
Guide 2	NQ	3.00	5.30	6.6
Guide 3	N/A	5.50	N/A	9.2
Guide 4	NQ	NQ	12.30	24.9

Table 4 A table to show the gene-modification estimates attained for conditions of 1µg:1µg and 2µg:2µg of U6-Guide and Px165 co-transfected, subject to the T7EI and subsequently resolved on agarose and 10% TBE acrylamide gels. NQ – Not quantifiable and N/A not applicable.

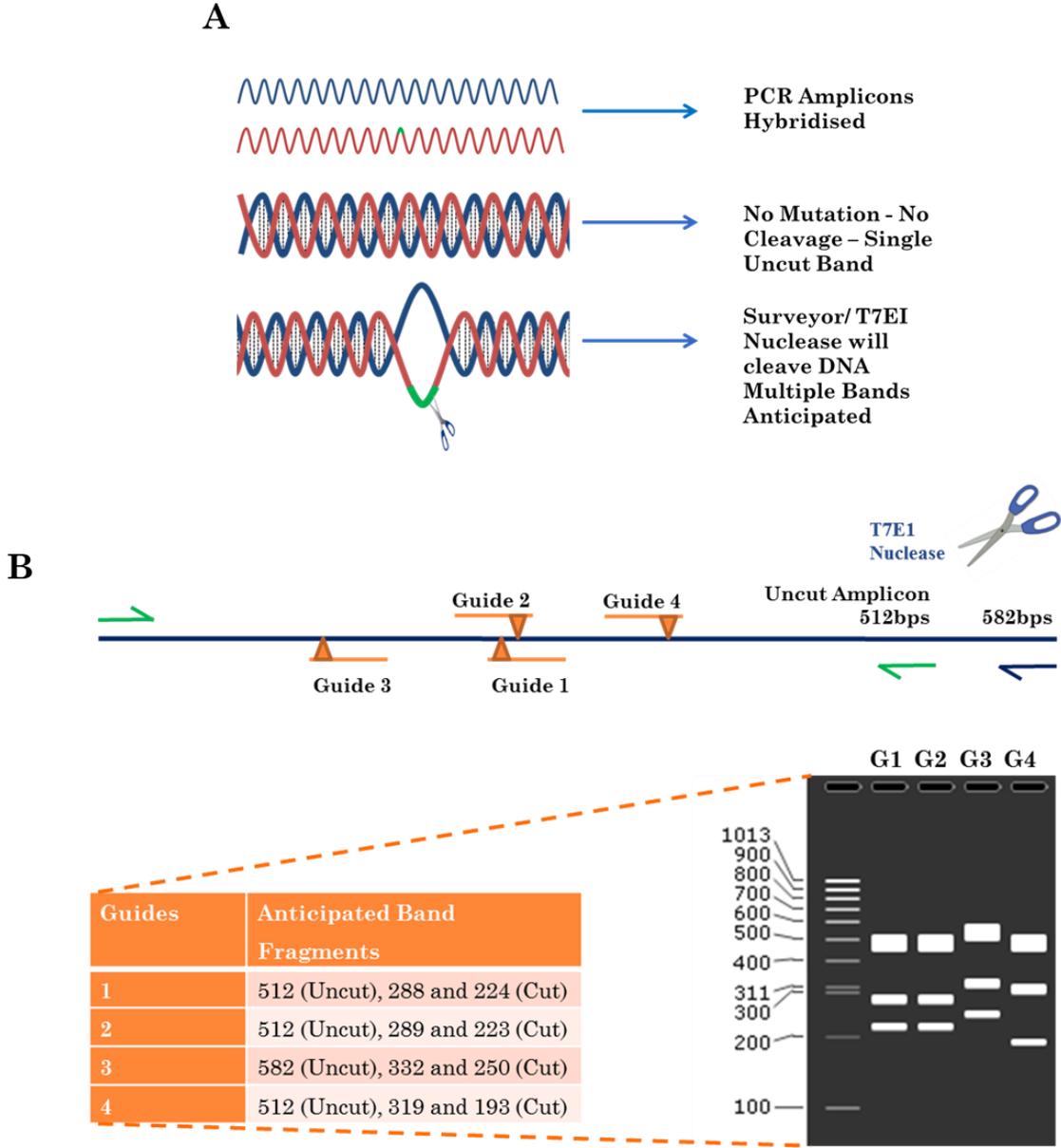


Figure 4-5 A schematic illustrating the underlying mechanism of the T7E1 mismatch detection assay.

A) A diagram showing the underlying mechanism of the T7E1 mismatch detection assay, an asymmetric PCR is performed across the genomic region the guide RNAs have affinity to. Post treatment with the CRISPR/Cas9 system, the DSB introduced by the guides will be repaired by an NHEJ mechanism, causing the local introduction of INDELS. The resultant amplicons are heterogeneous, with a mixed population of amplicons with INDEL mutations (MT) and amplicons without INDEL mutations (WT). The amplicons in this population are denatured to resultant single stranded DNA, INDEL mutations are denoted by the green line on the DNA. The resultant ssDNA are hybridised, WT sequences anneal without secondary structures in the DNA. Whereas when a WT and MT single strands anneal there is the formation of a heteroduplex structure in the DNA (loop in the diagram). The T7E1 nuclease (blue scissors) can cleave these secondary structures in the DNA, and thus the mixed amplicon indicating the presence of INDELS. B) A schematic diagram of the Intron 1 PCRs used for the T7E1 assay, the amplicon (blue line) is produced by either by 1960F and 2471R (green half arrows 512bps amplicon: Guides 1,2 and 4) or 1960F and 2541R (green forward and blue reverse arrows 582bp amplicon: Guide 3). The position of guides is indicated by the orange lines and the position of DNA cleavage and local INDEL introduction is indicated by orange triangles touching the amplicon. Below a table and “predicted” agarose gel image show the anticipated uncut and cut amplicon fragments following digestion with the T7E1 nuclease.

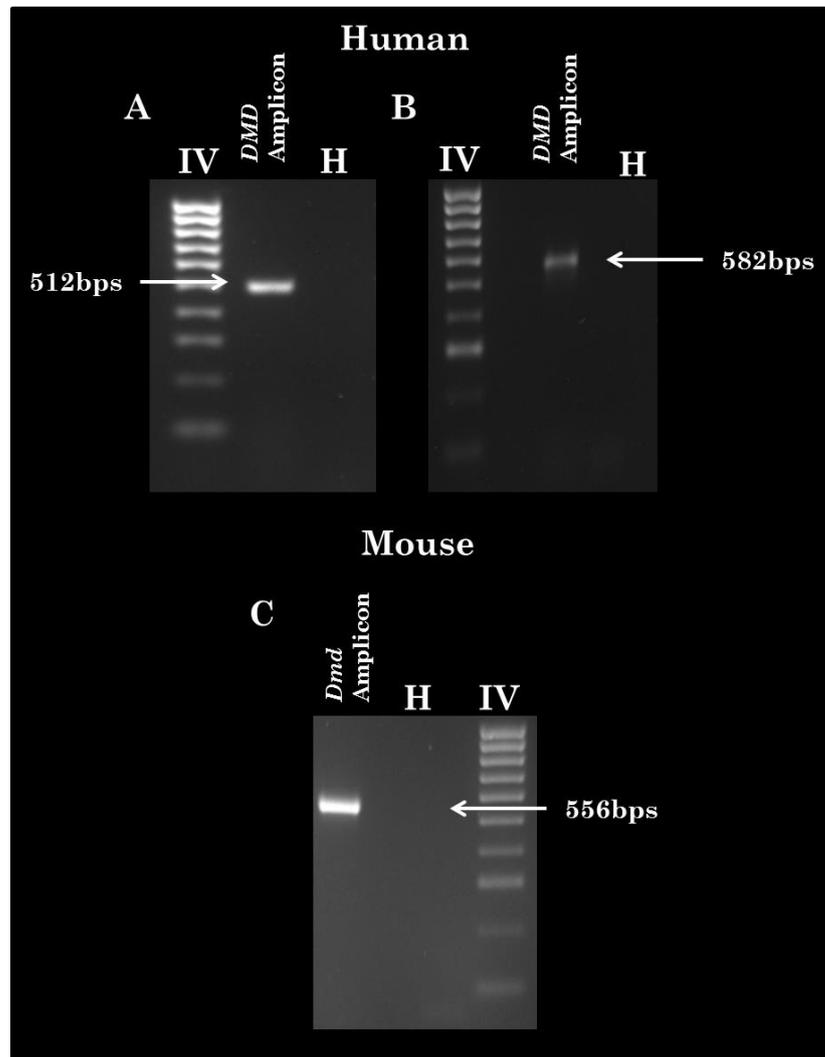


Figure 4-6 Gel electrophoresis of amplicon products from human *DMD* and mouse *Dmd* intron 1 PCRs.

A) Gel electrophoresis of a PCR amplicon, amplified from 250ng wild-type genomic HEK293T DNA. The product was amplified using 1960F and 2471R primers in a 35 cycle reaction using HotStart Phusion Flex (NEB). Once amplified a 10 μ l aliquot of the sample denoted DMD amplicon was resolved on a 2% agarose gel. The size is consistent with the 512bp amplicon anticipated and is asymmetric with regard to guide RNA 1, 2 and 4 putative cleavage sites. **B)** Gel electrophoresis of a PCR amplicon, amplified from 250ng of wild-type genomic HEK293T DNA. The product was amplified using 1960F and 2541R primers in a 35 cycle reaction using HotStart Phusion Flex (NEB). Once amplified a 10 μ l aliquot of the sample denoted DMD amplicon was resolved on a 2% agarose gel. The size is consistent with the 582bp amplicon anticipated and is asymmetric with regard to the guide RNA 3 putative cleavage site. **C)** Gel electrophoresis of a PCR amplicon, amplified from 250ng H2kb *mdx* genomic DNA. The product was amplified using 1950F and 2775R primers in a 35 cycle reaction using HotStart Phusion Flex (NEB). Once amplified a 10 μ l aliquot of the sample denoted *Dmd* amplicon in the figure was resolved on a 2% agarose gel. The size is consistent with the 556bp amplicon anticipated and is asymmetric with regard to mouse guide RNA 1, 2, 3, 4 and 5 putative cleavage sites. All samples were resolved alongside a suitable non-template water control (H) and hyperladder IV.

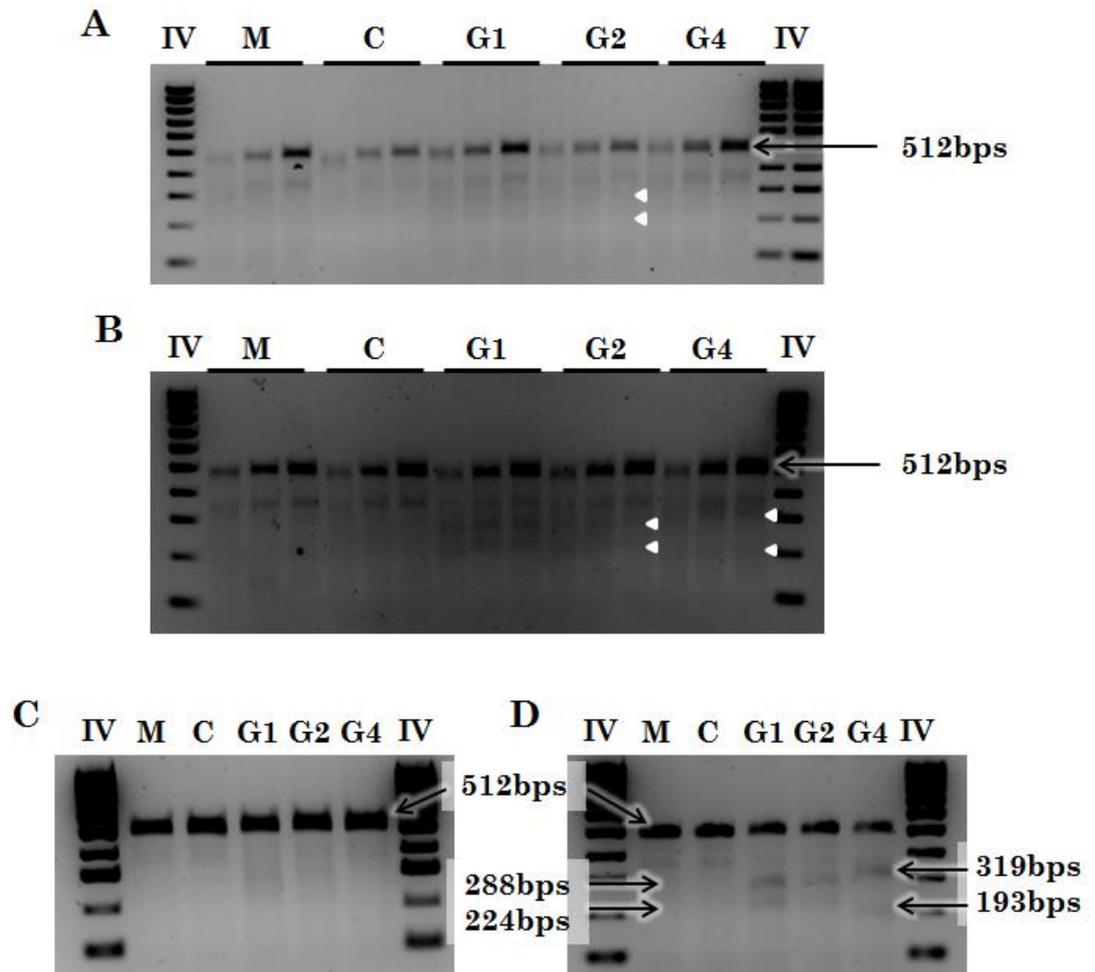


Figure 4-7 Surveyor and T7E1 mismatch cleavage digests, spanning the human *DMD* intron 1 Cas9 target site resolved on 2% Agarose gel.

A) Gel electrophoresis of Surveyor Nuclease digests for the purposes of mismatch detection, undertaken on amplicons from HEK293T genomic DNA following a range of treatments. Including: Mock (M) - addition of 20µl of Viafact transfection reagent to cells, Cas9 Alone (C) - 1µg Cas9 expressing Px165 plasmid alone and co- transfection of 1µg U6- Guide 1, 2 or 4 alongside 1µg Px165 at a 1:1 ratio, labelled G1, G2 and G4 respectively. Once amplified using 1960F and 2741R primers, varying volumes of 6µl, 12µl and 18µl of each PCR condition (loaded left to right in order of ascension), were annealed and subject to a 2µl of the Surveyor nuclease. **B)** The same process as described in A except for 2µg G1, G2 and G4 and 2µg Px165 Cas9 expressing plasmid in dual transfections. **C)** T7E1 nuclease digest undertaken on 200ng PCR purified hybridised amplicon for all treatments of the 1µg U6-guide :1µg Px165 condition described in A. **D)** T7E1 nuclease digest undertaken on 200ng PCR purified hybridised amplicon for all treatments of the 2µg Guide :2µg Px165 Cas9 condition. Notably cleavage products are present for Guide RNA 1: 288bps and 224bps, Guide RNA 2: 289bps and 223bps and Guide RNA 4: 319bps and 193bp, for both conditions when subject to the Surveyor and the T7E1 assay, this stated the presence of bands is much clearer with the 2µg:2µg conditions. Band present are indicated by white triangles for the surveyor gel images and arrows for the T7E1 gel images. All samples were resolved on 2% agarose alongside hyperladder IV using 1X TAE buffer and inversion of the image was performed to enhance the appearance of cleavage products.

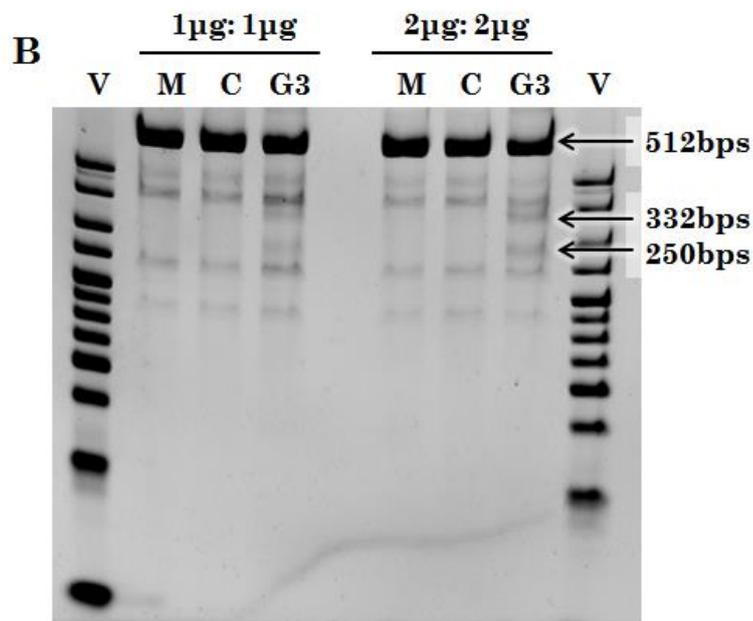
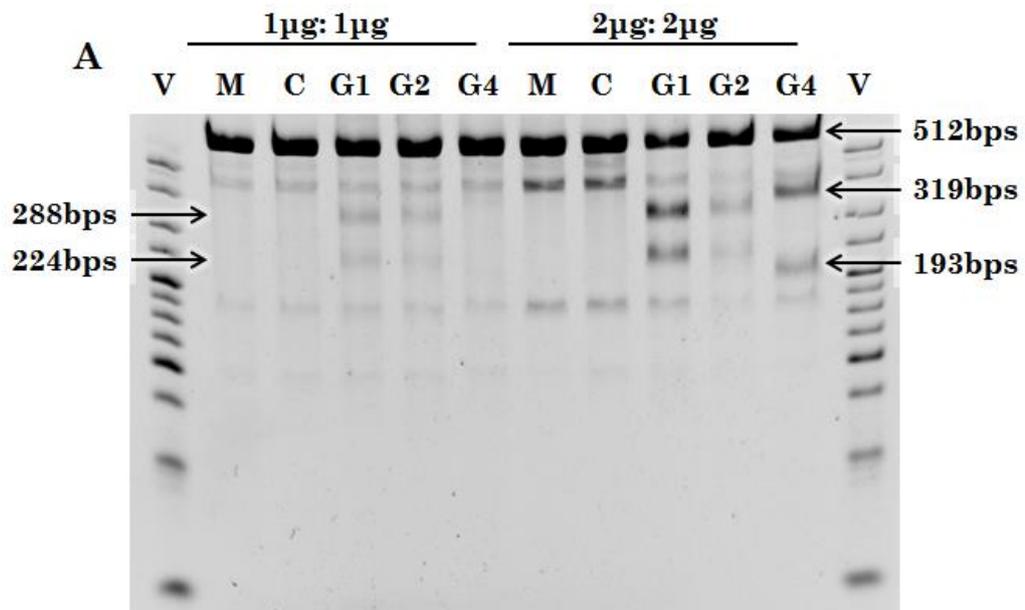


Figure 4-8 T7E1 Digest products of HEK293T hybridised treated amplicons resolved on a 10% TBE polyacrylamide gel.

A) T7E1 digests for mismatch detection were undertaken on 400ng of purified PCR amplicons, performed on 250ng of HEK293T genomic DNA following a range of treatments. PCR amplifications were performed using 1960F and 2741R primers and resolved on 10% TBE polyacrylamide gel. Treatments included: Mock (M) addition of 20µl of Viafect transfection reagent to cells, Cas9 Alone (C) 1µg Cas9 expressing Px165 plasmid alone and co-transfection of U6- Guides 1, 2 or 4 alongside Px165 at a 1:1 ratio, at 1µg: 1µg or 2µg:2µg doses labelled as G1, 2 and 4 in both conditions respectively. **B)** T7E1 digests for mismatch detection on the same treatment conditions, using the U6-guide3 expression plasmid and associated 1960F- 2541R primer pair as previously described. All gels were run at 200V for 45 minutes in line with the Thermofischer (Invitrogen) protocols in 1X TBE buffer. This was subsequently stained with SYBRSafe diluted 1 in 5000 in 1xTBE, which was incubated at room temperature for 1 hour to visualise bands. Sizes of cleavage bands are indicated, note guide RNA 2 gives 289bp and 223bp cleavage products due to the marginal 1 nucleotide difference relative to guide RNA 1, this is not directly labelled.

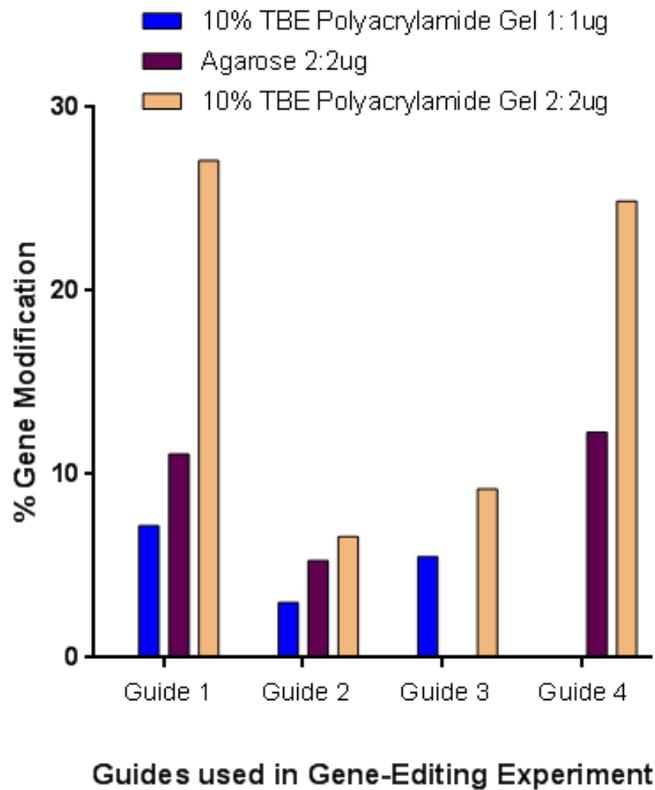


Figure 4-9 Graphical summary of percentage gene modification produced by 1:1 μ g and 2:2 μ g co-transfection of HEK293T cells with U6 and Px165 plasmids as estimated by the T7E1 assay, when resolved on either a 2% agarose or 10% TBE polyacrylamide gel.

Graphical summary of the general trends observed on optimising the surveyor and T7E1 mismatch detection assays. Note that 1 μ g:1 μ g samples could not be accurately quantified using 1% agarose due to low resolution of cleavage products and thus are not included. Furthermore findings suggest that 10% TBE gels give superior resolution of cleavage products and more accurate gene-modification estimates relative to 2% agarose. As a consequence of this, guide RNA 3 quantification was only taken on a 10% TBE gel. Notably, findings indicate that guide RNA 1 appears the most efficacious of the in silico guide designs however; more repetitions are required for assertion of this finding.

4.4.2.3 Continued Assessment of Guide RNA Efficacy Employing the HEK293T.Cas9 Cell Line.

In parallel with the optimisation of the gene-modification assays described above, U6-guide plasmid transfection dose escalation studies were undertaken in HEK293T.Cas9 cells. This cell line, from Gencopeia, harbours the stable integration of the bi-cistronic CBh-Cas9-Ef1a-eGFP cassette at the AAVSI safe-harbour locus (Figure 4-10). Thus only the transient transfection of U6-Guide RNA encoding plasmids is required to provide the necessary constituents for Cas9 mediated DNA cleavage. The constitutive expression of integrated Cas9 within this cell line makes it a useful guide RNA screening tool removing variations in Cas9 expression that could occur as result during the co-transfection of U6 Guide RNA and Cas9 expressing Px165 plasmids.

In this instance, HEK293T.Cas9 cells were transiently transfected with 1 μ g, 2 μ g and 4 μ g doses of U6-gRNA plasmids using Viafect. Consistent with previously described methodologies samples were harvested at 48 hours and amplified in triplicate and pooled prior to 400ng of purified amplicon being hybridised and subjected to T7E1 cleavage.

The resultant cleavage products for each of these treatments were resolved on a 2% agarose gel (Figure 4-10). For guide RNA 1 the mean percentage gene modification were estimated to be 10.45% \pm 0.32, 12.90% \pm 0.72 and 11.63% \pm 1.61 at doses of 4, 2 and 1 μ g (mean \pm SEM; N=3) respectively. In addition, guide RNA 4 percentage gene modification was estimated at 4.1 \pm 0.94 SEM (N=3). For all other guide RNAs the determination of mean percentage gene modification was estimated at 2% or less. The increase in guide RNA 1 activity was deemed statistically significant relative to all other in silico guide designs at the 95% confidence interval using a one way ANOVA. A summary of percentage gene modification estimates obtained for all guide RNAs is presented graphically (Figure 4-11). The in-silico analyses undertaken previously indicated that guide 1 had the fewest off-targets, and taken in combination with the gene modification estimates presented this suggested that guide 1 was the most suitable candidate to progress to co-transfection studies exogenous microdystrophin repair template, in the pursuit of achieving integration at the *DMD* intron 1 site.

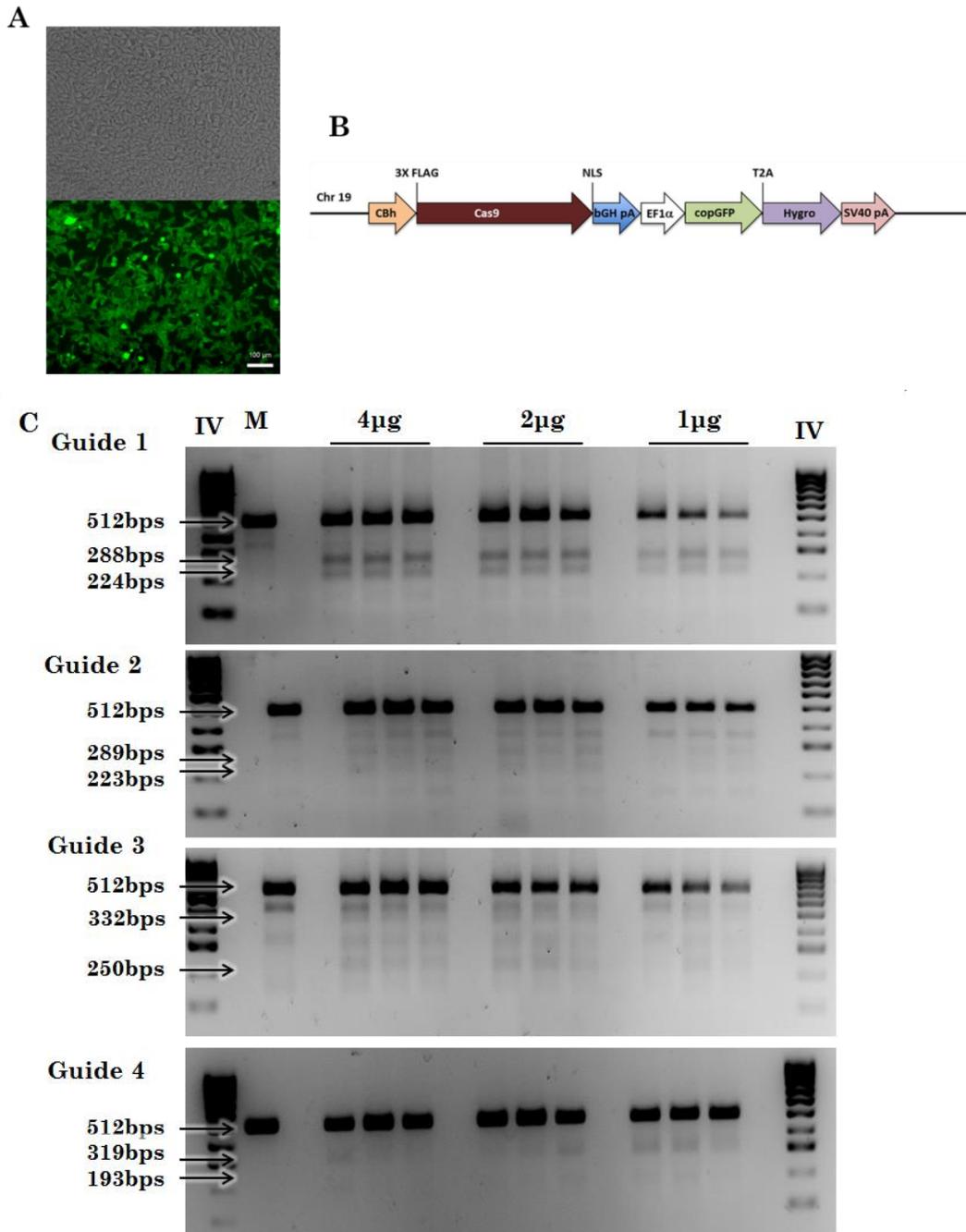


Figure 4-10 Screening guide RNA efficacy with the use of HEK293T.Cas9 cells

A) Constitutive GFP expression of HEK293T.Cas9 cells imaged at 100X magnification, using 1400ms exposure in the GFP channel 48 hours post seeding to a six well plate. **B)** A schematic of the Cas9 expression cassette integrated into the AAVSI locus of the HEK293T cells. The human optimised *Streptococcus pyogenes* Cas9 protein expression in these cells is driven by the CBh promoter, whilst the GFP and hygromycin transfection selection markers are under the control of the Ef1a promoter. This indicates that GFP expression is not necessarily reflective of the level of Cas9 protein in the culture. **C)** T7E1 digests undertaken on purified PCR amplicons, amplified from HEK293T.Cas9 genomic DNA following a range of treatments. Treatments included: Mock (M) HEK293T.Cas9 cells were treated with transfection reagent alone, or doses of 4 μ g, 2 μ g or 1 μ g of U6 expression cassettes encoding guide RNAs 1 -4. Cleavage products consistent with gene-modification were present for each of the guide RNAs, with guide 1 providing the most robust gene editing. All samples were resolved on a 2% agarose gel alongside hyperladder IV.

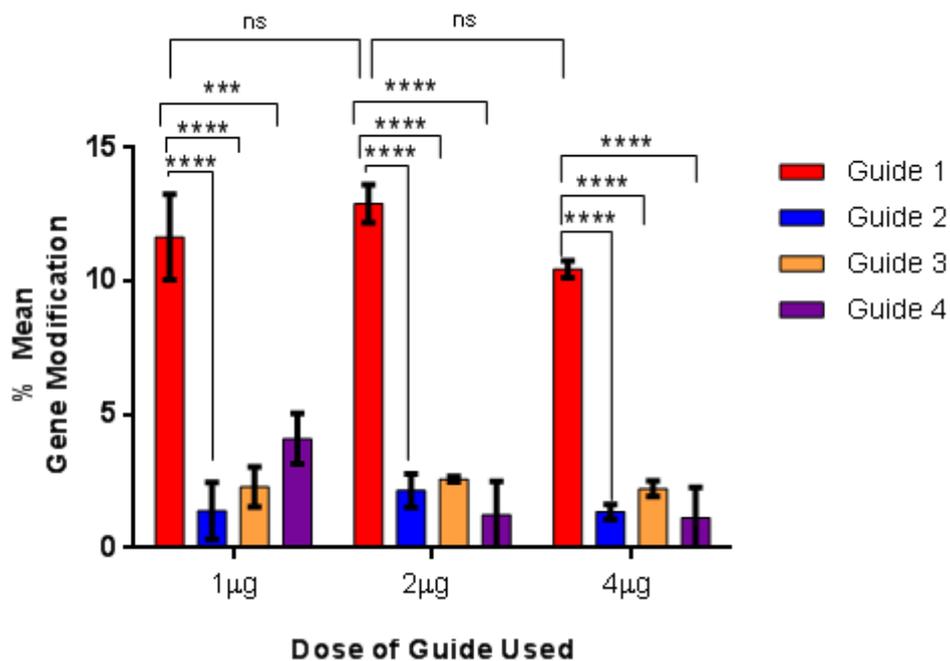


Figure 4-11 Graphical summary of the efficacy of In Silico guide designs, following transient transfection of 1µg, 2µg and 4µg U6 guide plasmids encoding guide RNAs 1, 2, 3 or 4 into HEK293T.Cas9.

Graphical summary of the mean gene modification estimated from the T7E1 treatment of 400ng of 1960F and 2741R amplicon. Each of the samples were amplified from 250ng HEK293T.Cas9 DNA, following transient transfection of U6-guide plasmids containing guide RNAs 1, 2, 3 or 4 at 1µg, 2µg and 4µg respectively. Notably, the mean gene modification was higher in samples treated with guide RNA 1 relative to all other in silico guide designs. This effect was deemed statistically significant to the 95% confidence interval by a 2-way ANOVA. Finally, the effect of U6-guide 1 dose did not significantly change the amount of mean gene modification identified.

4.5 Assessing Non-Viral Delivery Strategies for Cas9 Nuclease into Immortalised Δ 45-52 Patient Myoblasts.

4.5.1 Transient Transfection of Px458 to Immortalised Patient Myoblasts Harboured Deletions of Exon 45 -52

Functional mismatch detection assays were established and guide RNA 1 was identified as the most efficacious of the in silico guide designs. To transition gene-editing approaches to a more therapeutically representative in-vitro dystrophic model, studies were undertaken in the Δ 45-52 immortalised DMD patient myoblasts. To facilitate this transition the empty Px458 plasmid construct was used to assess and optimise transfection efficiency (Ran et al. 2013).

4.5.1.1 Lipofectamine3000 Transient Transfections of Immortalised Patient Myoblasts

Transient transfection of Px458 was trialled initially with Lipofectamine3000, a reagent with a propensity to deliver DNA with high efficiency to difficult-to-transfect cells. The transfection was performed in accordance with the manufacturer's protocol using a transfection reagent (μ l) to DNA (μ g) ratios of 1.5:1 or 3:1 with 2.5 μ g of Px458 or pCI-GFP positive control. In both ratios trialled, the delivery of pCI-GFP was successful, as determined by microscopy imaging and cytometric analysis of the proportion of GFP positive cells (Figure 4-12). In the 1.5:1 and 3:1 ratios with pCI-GFP, mean percentage fluorescence was established to be 10.3% \pm 1.29 and 18.5% \pm 0.647 respectively (Mean \pm SEM, N=3) which was deemed statistically significantly different at the 95% confidence interval by 2-way ANOVA. Whereas, in the case of Px458, transfection efficiency was diminished and mean percentage GFP was established to be much lower at 0.297% \pm 0.083 and 0.327% \pm 0.061 in both conditions yielding a non-significant difference. Finally, a statistical difference at the 95% confidence level was also observed between the number of GFP positive cells yielded from pCI-GFP and Px458 transient transfections.

It was noted that the disparity in plasmid sizes used in the investigation (pCI-GFP is 4709bp, and Px458 is 9289bp) would result in more plasmid copies of pCI-GFP being present relative to Px458 in the 2.5 μ g DNA mass added to the transfection mixture. It was speculated that this could in turn influence the transfection efficiency observed. To address this a dose escalation was required in which a 3:1 ratio of transfection reagent, asserted in the case of pCI-GFP to facilitate higher delivery, was used to deliver 2.5 μ g, 5 μ g and 7.5 μ g of Px458 (Figure 4-13).

At these doses, mean percentage of GFP positive cells was established to be 2.39% \pm 0.118, 4.15% \pm 0.265 and 1.65% \pm 0.186 SEM respectively, with N =3 in all conditions. The difference observed was determined to be significant between the 2.5 μ g and 5 μ g doses, at the 95% confidence interval by 2-way ANOVA, but the difference between 2.5 μ g and 7 μ g was identified

as non-significant. Notably, the baseline transfection of 2.5µg showed improved transfection efficiency relative to preceding experiment. However, the highest mean percentage of GFP positive cells observed in this experiment was ~5% achieved with the 5µg dose. This would provide a very small pool of cells to undergo gene modification and subsequent HDR, thus this method of DNA delivery was deemed insufficient for the requirements of the investigation.

4.5.1.2 Viafect Transient Transfections of Immortalised Δ45-52 Patient Myoblasts

Due to the unexpectedly low efficiency of Lipofectamine3000, a second transfection reagent, Viafect, was trialled in the delivery of Px458 plasmid to immortalised Δ45-52 patient myoblasts. Once more, transient transfections was performed in accordance with the manufacturer's protocol. In this instance three ratios of Viafect, 3:1, 4:1 and 5:1, were used to deliver 2µg of Px458 or pCI-GFP (Figure 4-14). At 48 hours post transfection there was a striking proportion of GFP positive cells, in both pCI-GFP and Px458 conditions. The mean percentage of GFP positive cells at ratios of 3:1, 4:1 and 5:1 for pCI-GFP was established as to be 75.33% ± 0.81, 75.93% ± 0.38 and 74.10% ± 1.40 SEM respectively and for Px458 to be 31.25% ± 1.86, 20.56% ± 1.33 and 29.33% ± 2.53 SEM. A statistically significant 2-3 fold decrease in transfection efficacy was observed when delivering Px458 compared to pCI-GFP (95% confidence interval by 2-way ANOVA).

The different ratios trialled showed comparable transfection efficiency, but toxicity and cell death occurred at higher transfection reagent to DNA ratios, as evidenced by phase contrast images of the 4:1 ratio of transfected patient myoblasts relative to the 3:1 ratio and mock conditions. In this instance the number of GFP positive cells at the 3:1 relative to the 4:1 ratio, was deemed significantly higher ($p < 0.05$). This stated the difference observed between 3:1 and 5:1 ratios was deemed non-significant. In light of the comparable transfection efficiency and visual indications of toxicity it was resolved that the lowest transfection reagent to DNA ratio of 3:1 would be adhered to from this point onwards.

Using a 3:1 transfection reagent ratio, dose escalation delivering 2µg, 4µg and 6µg of Px458 was then undertaken (Figure 4-15) yielding GFP positive cells of 38.92% ± 1.55, 42.89% ± 0.67 and 38.5% ± 0.817 respectively (Mean +/- SEM; N=3). No significant differences between the doses were observed.

Whilst, the Lipofectamine3000 and Viafect transient transfections of Px458 were undertaken using slightly differing conditions as dictated by their associated protocols, comparison of the two transfection reagents highlights that Viafect achieves greater transfection efficiency than Lipofectamine3000. As noted by the maximal transfection efficiency achieved of >40% GFP positive cells following Viafect transfection of 4µg of Px458 relative to <5% following

transfection of 5 μ g of Px458 by with Lipofectamine3000. Thus the Viafect transfection reagent was selected for ongoing assessments of gene editing.

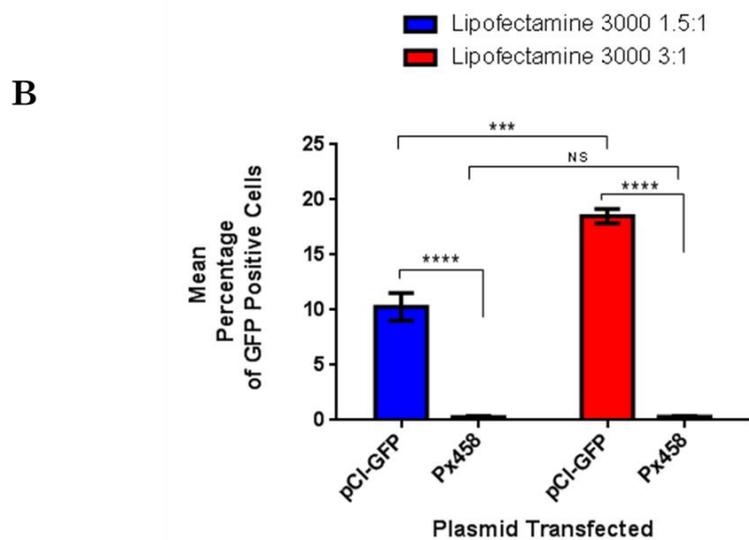
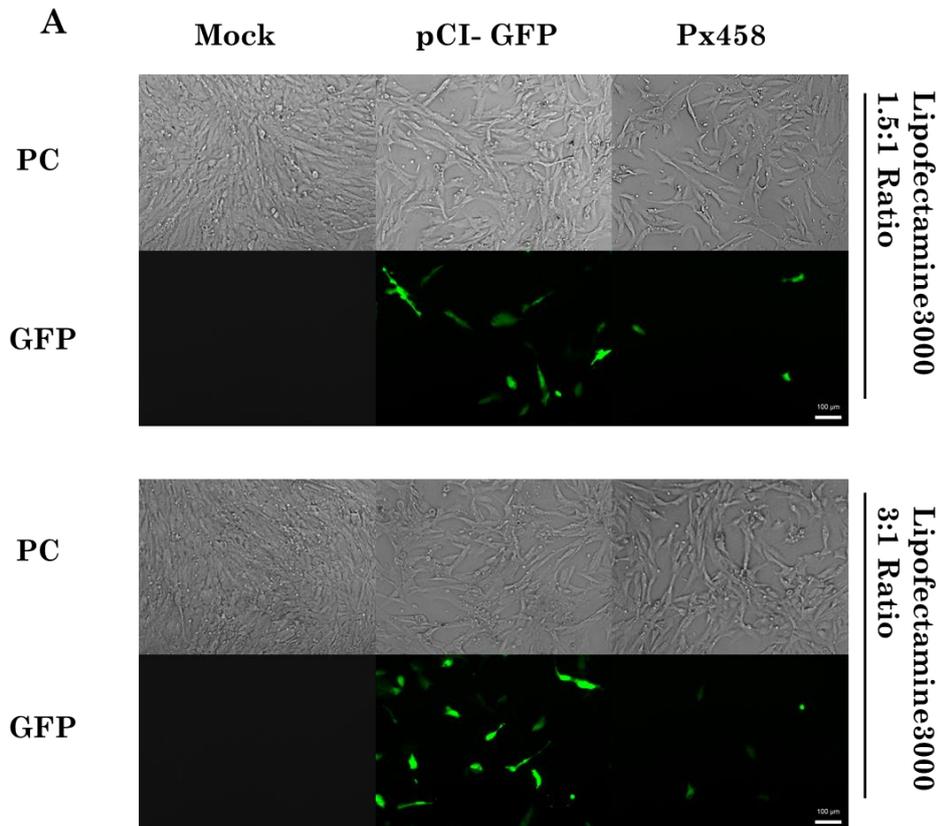


Figure 4-12 Transient Transfection of pCI-GFP and Px458 into Δ 45-52 Immortalised DMD patient myoblasts using lipofectamine3000.

A) Phase Contrast (PC) and fluorescent microscopy images were taken of Immortalised Δ 45-52 patient myoblasts in the FITC (GFP) channel. The cells were transiently transfected with 2.5 μ g pCI-GFP or Px458 plasmids at 1.5:1 and 3:1 Lipofectamine3000 reagent μ l: DNA μ g ratios. All images were taken with the Zeiss microscope 48 hours post transfection, using 100x magnification. Phase contrast images were taken at 30ms exposure and GFP Images at 800ms exposure. The image displayed is representative of 5 images taken per well. B) A bar chart showing the mean percentage of GFP positive cells, determined using flow cytometric analysis of cells fixed in 2% PFA-FACs buffer 48 hours post transient transfection. The transfection efficiencies of GFP and Px458 at the two ratios of lipofectamine3000 reagent, were compared by two-way ANOVA, it is indicated in the graph where differences are statistically significant.

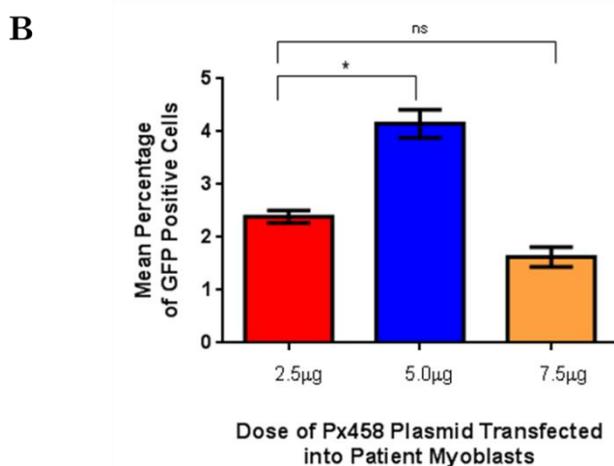
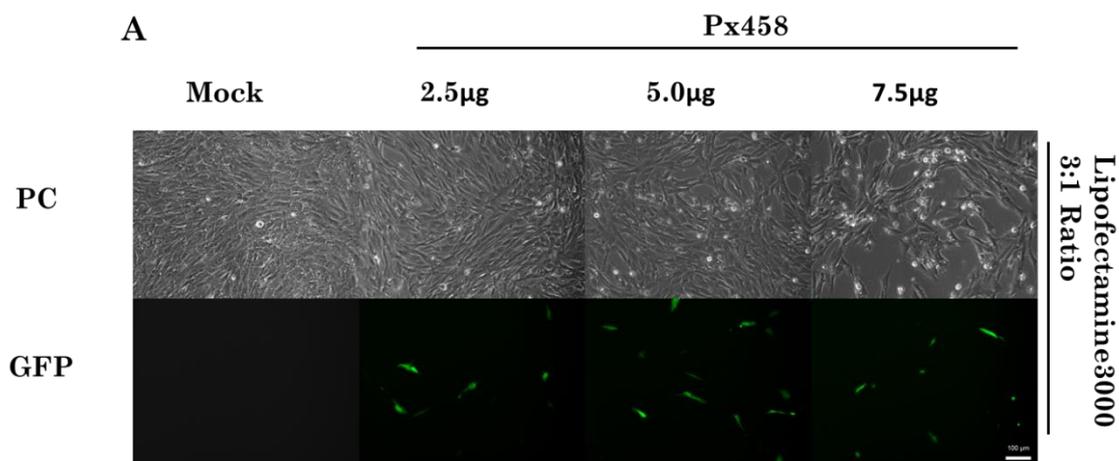


Figure 4-13 A dose escalation of Px458 transfected into immortalised Δ 45-52 DMD patient myoblasts.

A) Phase Contrast (PC) and fluorescent microscopy images were taken of Immortalised Δ 45-52 patient myoblasts in the FITC (GFP) channel. The cells imaged were transiently transfected with doses of 2.5µg, 5.0µg and 7.5µg Px458 at a 3:1 Lipofectamine3000 reagent µl: DNA µg ratio. All images were taken with the Zeiss microscope 48 hours post-transfection at 100x magnification. Phase contrast images were taken at 30ms exposure and GFP Images at 800ms exposure. The image displayed is representative of 5 images taken per well **B**) A bar chart showing the mean percentage of GFP-positive cells, determined using flow cytometric analysis of cells fixed in 2% PFA-FACs buffer 48 hours post transient transfection. A twofold difference between the 2.5µg and 5.0µg conditions was observed and deemed statistically significant by two way ANOVA. However the total transfection efficiency observed is a little under 5%, which is unlikely to enable sufficient DNA cleavage for dystrophin cDNA integration at the DMD intron 1 site.

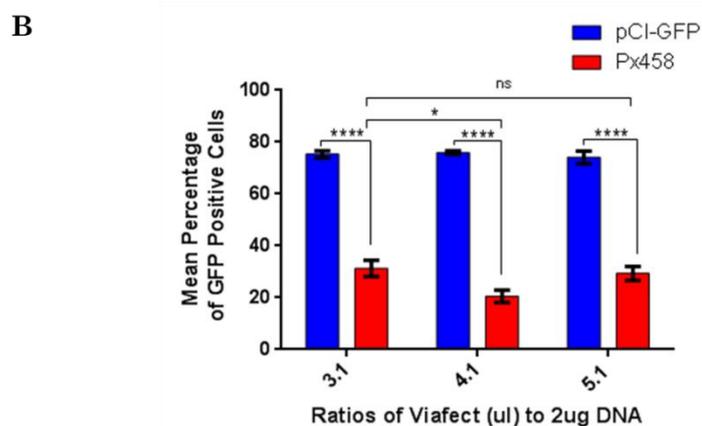
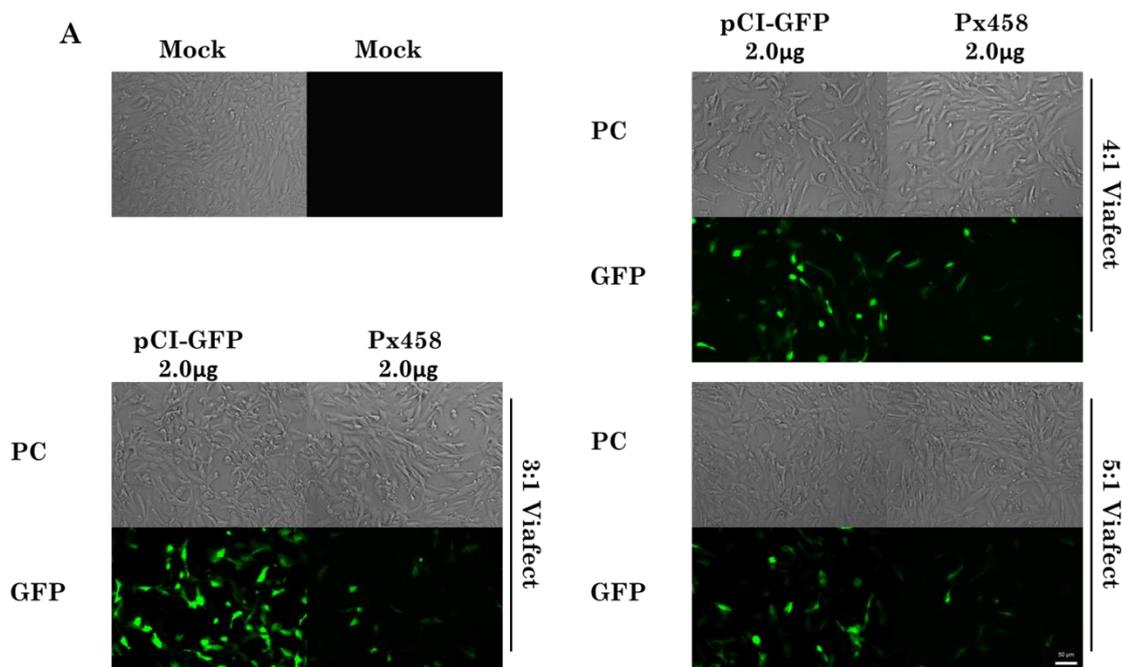


Figure 4-14 Transient transfection of pCI-GFP and Px458 into immortalised Δ 45-52 DMD patient myoblasts using Viafect.

Phase Contrast (PC) and fluorescent microscopy images were taken of Immortalised Δ 45-52 patient myoblasts in the FITC (GFP) channel. The cells were transiently transfected with 2.0 μ g of pCI-GFP and Px458 plasmids at 3:1, 4:1 and 5:1 Viafect reagent μ l: DNA μ g ratios. Transient transfections of Px458-were undertaken alongside mock samples, in which cells were treated with 20 μ l of Viafect transfection reagent in isolation. All images were taken 48 hours post-transfection at 200x magnification on the Zeiss microscope. Phase contrast images were taken at 30ms exposure and GFP images at 800ms exposure. The image displayed is representative of 5 images taken per well. **B**) A bar chart showing the mean percentage of GFP-positive cells determined using flow cytometric analysis of cells fixed in 2% PFA-FACS buffer 48 hours post transient transfection, with differing ratios of Viafect reagent to pCI-GFP and Px458 respectively. The transfection efficiencies of GFP and Px458 at the two ratios of Viafect reagent, were compared by two-way ANOVA, it is indicated in the graph where differences are statistically significant.

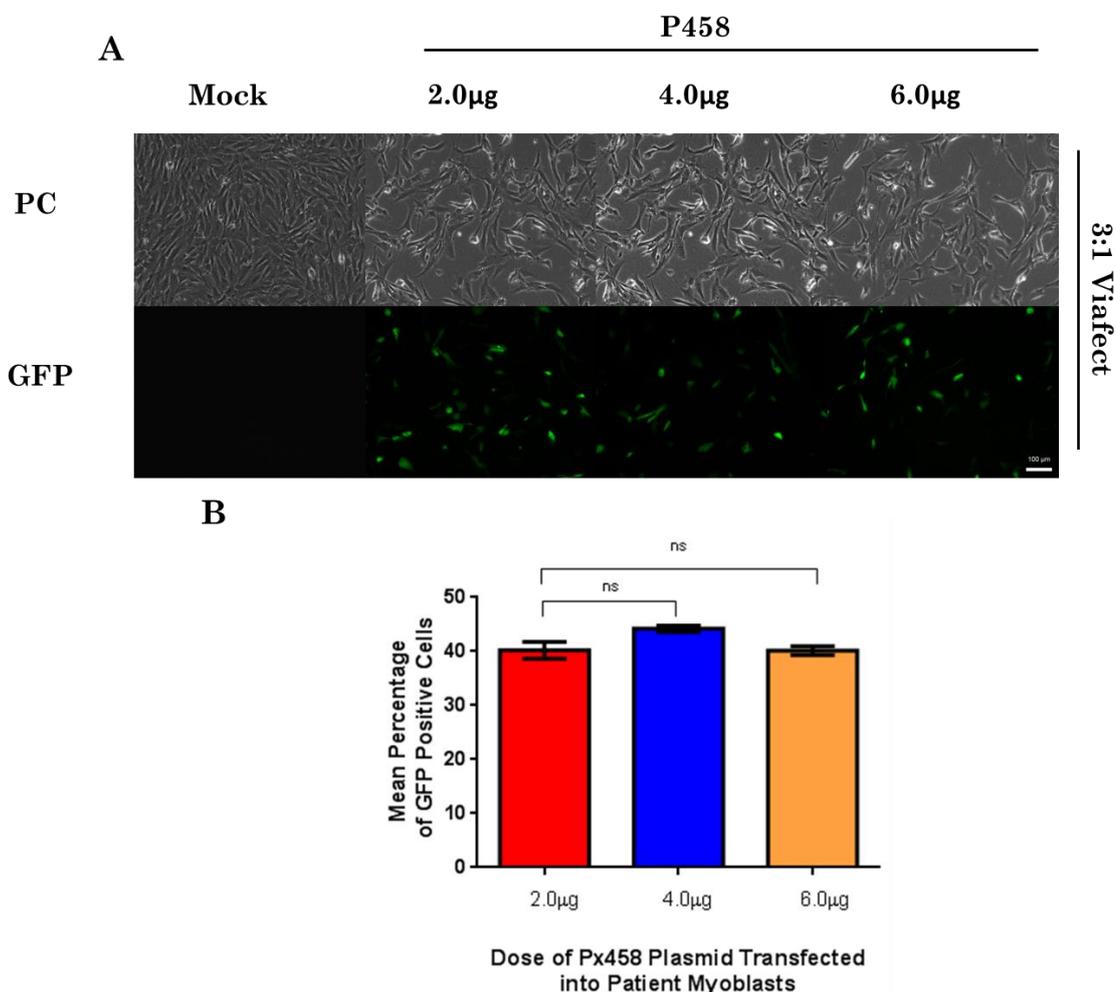


Figure 4-15 Transfected dose escalation of Px458 in immortalised Δ 45-52 DMD patient myoblasts using Viafect transfection reagent.

A) Phase Contrast (PC) and fluorescent microscopy images were taken of Immortalised Δ 45-52 patient myoblasts in the FITC (GFP) channel. Cells were transiently transfected with the Px458 plasmid at doses of 2.0µg, 4.0µg and 6.0µg using 3:1 Viafect transfection reagent μ l: DNA μ g ratio. Transient transfection of the Px458 plasmid was undertaken alongside mock samples, in which cells were treated with 18µl of Viafect transfection reagent in isolation. The GFP expression was imaged with the Zeiss microscope 48 hours post-transfection at 100x magnification, with phase contrast images taken at an exposure of 30ms and GFP images at an exposure of 800ms. **B)** A bar chart showing the mean percentage of GFP-positive cells, determined using flow cytometric analysis of cells fixed in 2% PFA-FACs buffer 48 hours post transient transfection, with differing doses of Px458 cas9 expressing plasmid.

4.5.1.3 Examination of Transfection Using the All-In-One Px458- Guide 1 Plasmid in Gene-Editing Experiments:

After being used to establish transfection protocols, it was sought to use the all-in-one Px458 plasmid for the purpose of gene editing. This was achieved by sub-cloning annealed oligos encoding guide 1, previously determined to be the most efficacious of in silico guide designs, into this plasmid backbone. In this manner GFP expression was used as a proxy to indicate the successful delivery of all constituents necessary for the CRISPR/Cas9 system to introduce DSBs at the DMD intron 1 locus.

Transient transfection of Px458-Guide 1 was undertaken at both 2 μ g and 4 μ g doses in both HEK293T cells and immortalised Δ 45-55 patient myoblasts. Microscopy images were taken 48 hours post transfection, and GFP expression was used to visually assess the delivery of the Px458-Guide 1. The GFP expression was deemed comparable on gross estimation to previous transfections of the respective cell lines with the empty Px458 plasmid (Figure 4-16 and Figure 4-19).

Genomic DNA was harvested from both cell lines 48 hours post transient transfection of Px458-Guide 1, amplified using HotStart phusion polymerase and numerous methods were undertaken to estimate gene modification occurring at the DMD intron 1 locus. The resultant amplicons were hybridised subjected to the standard T7E1 assay and subsequent cleavage products were then resolved and quantified using either a 10% TBE polyacrylamide gel or a high sensitivity DNA chip on the Agilent Bioanalyzer. The Agilent Bioanalyser uses electrophoretic microchips to determine nucleic acid size and concentration with high sensitivity, and it was speculated that this method could accurately identify the low levels of DNA cleavage product anticipated upon treatment of patient myoblast cultures. In addition to this, samples were also subject to tracking of InDels by decomposition (TIDE) analysis (Brinkman et al. 2014). This relatively novel TIDE method analyses sequence traces of wild type and treated amplicon products using a decomposition algorithm to assess for the presence of sequence modifications. TIDE was trialled due to the method in question being less expensive and labour intensive than mismatch detection assays. Furthermore, TIDE provides a greater level of insight into the resulting InDel spectrum at a DNA breakpoint. Collectively, these methods were used to provide robust quantification and insight into the gene editing occurring at the DMD intron 1 locus.

Estimates of gene modification estimated varied in accordance with the method used. In the first instance a 20 μ l T7E1 digest was undertaken on 400ng of PCR purified amplicon, and a 17 μ l aliquot of the digest was resolved on a 10% TBE polyacrylamide gel (Figure 4-16-B). Gene

modification was quantified using Image J. For HEK293T cells treated with 2 μ g and 4 μ g of Px458-Guide 1 plasmid, this was estimated to be 4.00% \pm 0.29 and 7.40% \pm 1.14 respectively (Mean \pm SEM; N=3). The remaining 3 μ l digest sample was stored and 1 μ l was analysed using the Agilent Bioanalyser which produces a pseudo gel image and electrophoretogram for all samples (Figure 4-17). In this Bioanalyser method, cleavage products are presented as a percentage of the total sample and this fraction was used to provide mean percentage gene modification of 2.41% \pm 0.37 and 4.86% \pm 0.83 (Mean \pm SEM; N=3). Notably, a general non-significant trend for the Bioanalyser to yield lower estimates of gene modification relative to samples resolved by polyacrylamide was observed.

Finally sequencing was undertaken on 75ng of the same PCR purified amplicons. The resultant sequencing trace was subject to analysis using the TIDE algorithm web tool (<https://tide-calculator.nki.nl/>) and a percentage estimate of InDels, proportional to gene modification was ascertained (Figure 4-18). The mean percentage of InDels was determined to be 10.93% \pm 1.33 SEM (N=2) and 10.31% \pm 0.40 SEM (N=3), for 2 μ g and 4 μ g conditions respectively. Notably the difference between mean gene modification estimates obtained from 10% TBE polyacrylamide gel or Agilent bioanalyser relative to TIDE was generally denoted as statistically significant in most cases.

The above was all achieved in HEK293T cells in the first instance, but the same analysis was undertaken in parallel for the immortalised patient myoblast samples which in the context of a 10% TBE gel gave mean percentage gene modification values of 1.14% \pm 0.39 and 5.24% \pm 0.40 SEM (N=3), respectively, for the 2 μ g and 4 μ g transfection conditions (Figure 4-19). In comparison the Agilent bioanalyser determined a mean percentage of gene modification of 1.08% \pm 0.33 SEM (N=2) and 4.45% \pm 0.71 SEM (N=3) (Figure 4-20). The difference in these two estimates of gene modification was once again deemed non-significant. In the context of TIDE the percentage of InDels present was deemed to be 6.85% \pm 0.43 SEM (N=3) at 2 μ g and 12.33% \pm 1.88 SEM (N=3) (Figure 4-21). Once more the difference between mean gene modification estimates obtained from 10% TBE polyacrylamide gel or Agilent bioanalyser relative to TIDE was denoted as statistically significant in all cases at the 95% level. Finally, a graphical comparison of the methods undertaken to analyse the two transfection doses, for both cell lines can be seen in Figure 4-22.

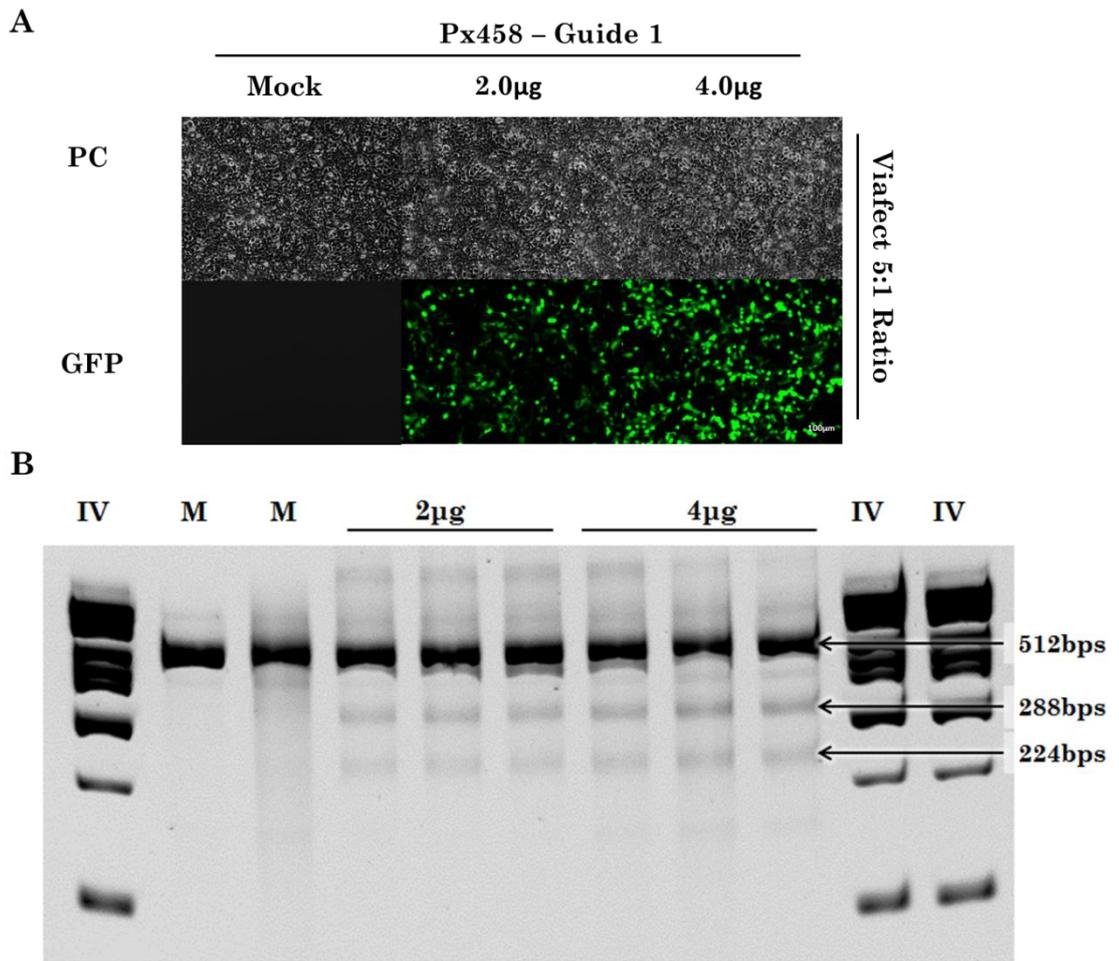


Figure 4-16 Examination of Px458 Guide-1 Transfection efficiency and determination of gene-modification in HEK293T cells.

A) Phase Contrast (PC) and fluorescent microscopy images were taken of HEK293T cells in the FITC (GFP) channel. Cells were transiently transfected with 2.0 μ g and 4.0 μ g doses of Px458-Guide 1 at a 5:1 Viafect reagent (μ l): DNA (μ g) ratio. Transient transfections of Px458-Guide 1 were undertaken alongside mock (M) samples, in which cells were treated with 20 μ l of Viafect transfection reagent in isolation. All images were taken with the Zeiss microscope 48 hours post transfection at 100x magnification, phase contrast images were taken at 30ms exposure and GFP images taken at 800ms exposure. (B) A T7E1 Assay performed on 400ng of DNA amplicon; amplified from genomic DNA harvested 48 hours post transfection using 1960F and 2471R, with the proofreading Hotstart Phusion polymerase (NEB). Samples were resolved on a 10% Tris-Borate-EDTA polyacrylamide gel run in 1X Tris-Borate-EDTA buffer alongside hyperladder IV (Bioline). Notably, across all Px458-Guide 1 treatments gene-modification is present, as indicated by the presence of cleavage products.

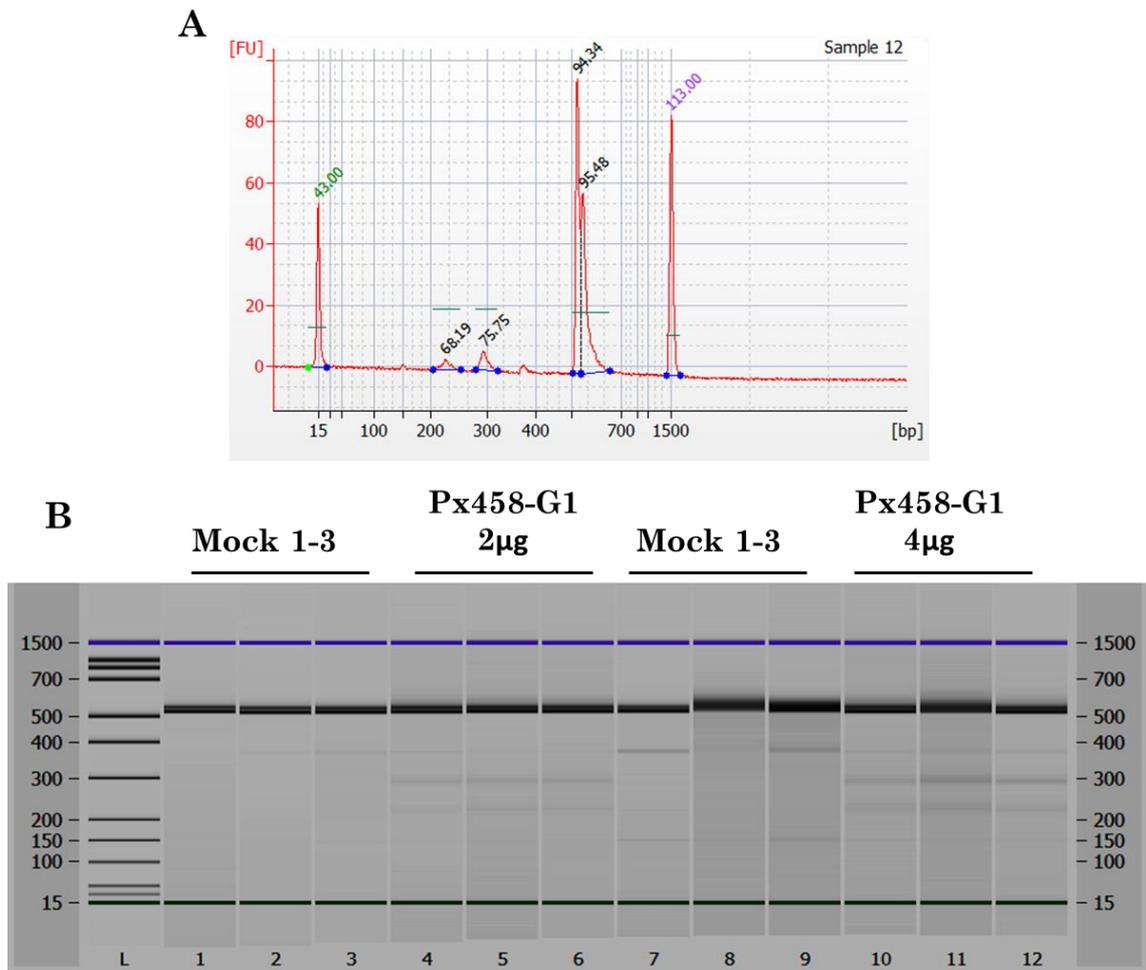


Figure 4-17 Agilent analysis of gene-modification in HEK293T treated with Px458-Guide 1 dual guide and Cas9 expressing plasmid.

A) An example electrophoretogram, showing the analysis of a 1 μ l aliquot of a T7E1 digest, undertaken on hybridised amplicons and resolved on a high sensitivity DNA chip, alongside 1 μ l of high sensitivity DNA ladder. Samples were amplified from 250ng genomic DNA harvested 48 hours post transfection from HEK293T cells using 1960F and 2741R primers. Transiently transfected samples included: Mock in which cells were treated with 20 μ l of Viafect alone and either 2 μ g or 4 μ g of Px458-Guide 1. Peaks at 15bpb and 1500bpb indicate higher and lower markers for the agilent, whilst peaks apparent at 512bpb, 288bpb and 224bpb are indicative of uncut and two cleavage products respectively. The area underneath the peaks was used to calculate levels of gene-modification. Notice in this instance the upper band appears to have a shoulder indicating the occurrence of a doublet. B) A composite gel, generated by the Agilent programme to simulate the amplicon resolution that would be apparent on gel electrophoresis.

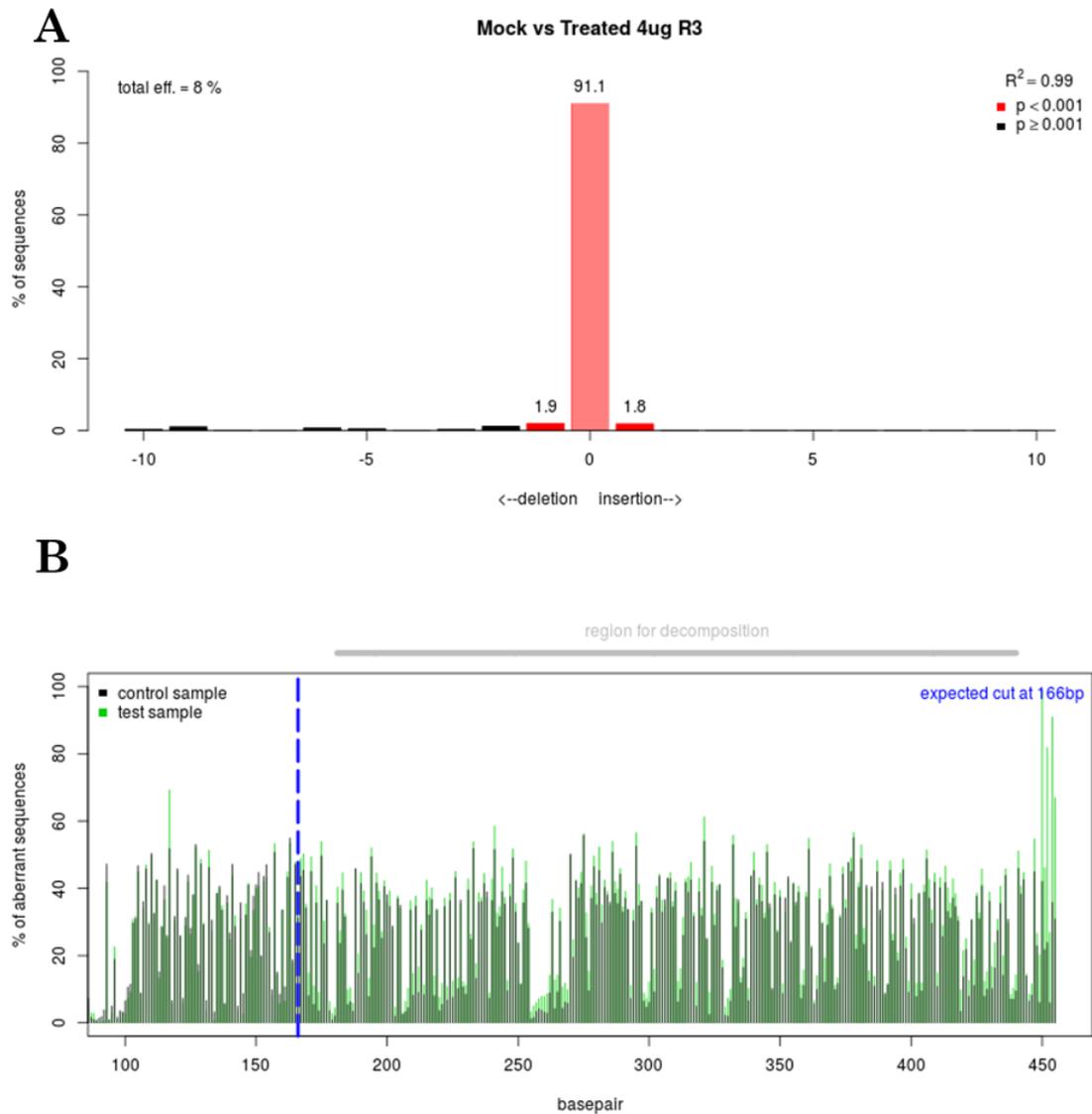


Figure 4-18 Tide analysis of gene-modification in HEK293T cells mediated by Px458-Guide 1 dual guide and Cas9 expressing plasmid.

A) A bar chart generated by TIDE analysis software (<https://tide-calculator.nki.nl/>), to indicate the percentage occurrence of InDels present in amplicons sequenced from a 4 μ g Px458-Guide 1 treatment. To achieve this, 250ng of HEK293T genomic DNA harvested from Mock and Px458 treatments 48 hours post transfection were amplified with 1960F and 2741R primers. The resultant amplicons were purified and an aliquot corresponding to 75ng supplemented with 2 μ l 1960F at 10 μ M, was sent for sequencing by Eurofins MWG. Sequence traces obtained for mock and Px458 conditions were then compared by the TIDE software for the presence of decomposition at the guide 1 target site and the occurrence and type of InDels was estimated computationally. Note mock in this case was HEK293T cells incubated in the presence of 20 μ l of Viafect transfection reagent alone. B) A decomposition trace for the sample described in part A. Dark green indicates comparable reads, light green indicates trace decomposition. The gene modification of this sample is approximately 8% however this sample appears to have a relatively high background.

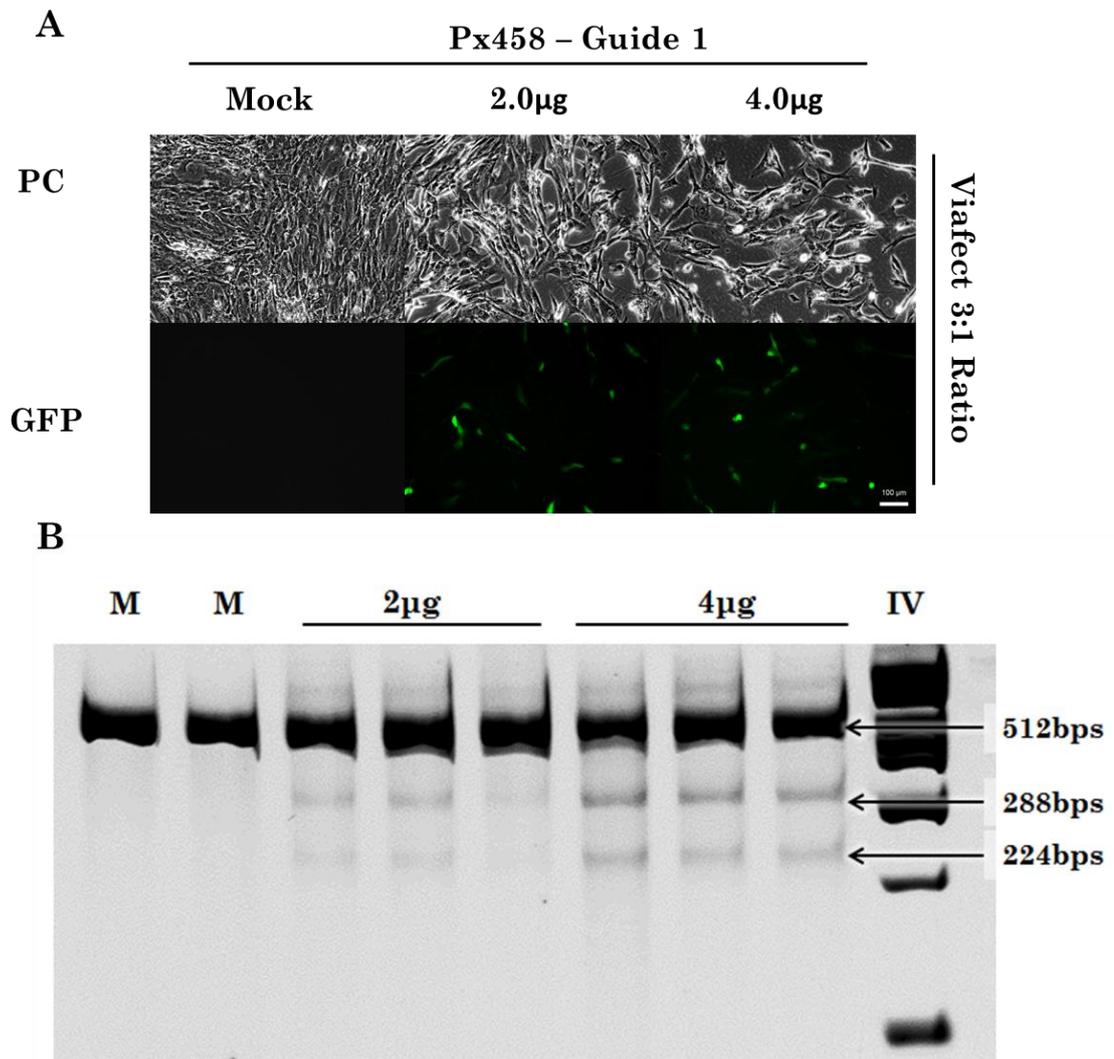


Figure 4-19 Examination of Px458 Guide-1 Transfection efficiency and determination of gene-modification in immortalised Δ 45-52 DMD patient myoblasts.

(A) Phase Contrast (PC) and fluorescent microscopy images were taken of Immortalised Δ 45-52 patient myoblasts in the FITC (GFP) channel. Cells were transiently transfected. Cells were transiently transfected with 2.0 μ g and 4.0 μ g doses of Px458-Guide 1 at a 3:1 Viafect reagent (μ l): DNA (μ g) ratio. Transient transfections of Px458-Guide 1 were undertaken alongside mock (M) samples, in which cells were treated with 20 μ l of Viafect transfection reagent in isolation. All images were taken with the Zeiss microscope 48 hours post transfection at 100x magnification, phase contrast images were taken at 30ms exposure and GFP images taken at 800ms exposure. **(B)** A T7E1 Assay performed on 400ng of DNA amplicon; amplified from 250ng of Immortalised Δ 45-52 Patient myoblast genomic DNA, harvested 48 hours post transfection using 1960F and 2471R, with the proofreading Hotstart Phusion polymerase. Samples were resolved on a 10% Tris-Borate-EDTA polyacrylamide gel run in 1X Tris-Borate-EDTA buffer alongside hyperladder IV (Biolone). Notably, across all Px458-Guide 1 treatments gene-modification is present, as indicated by the presence of cleavage products.

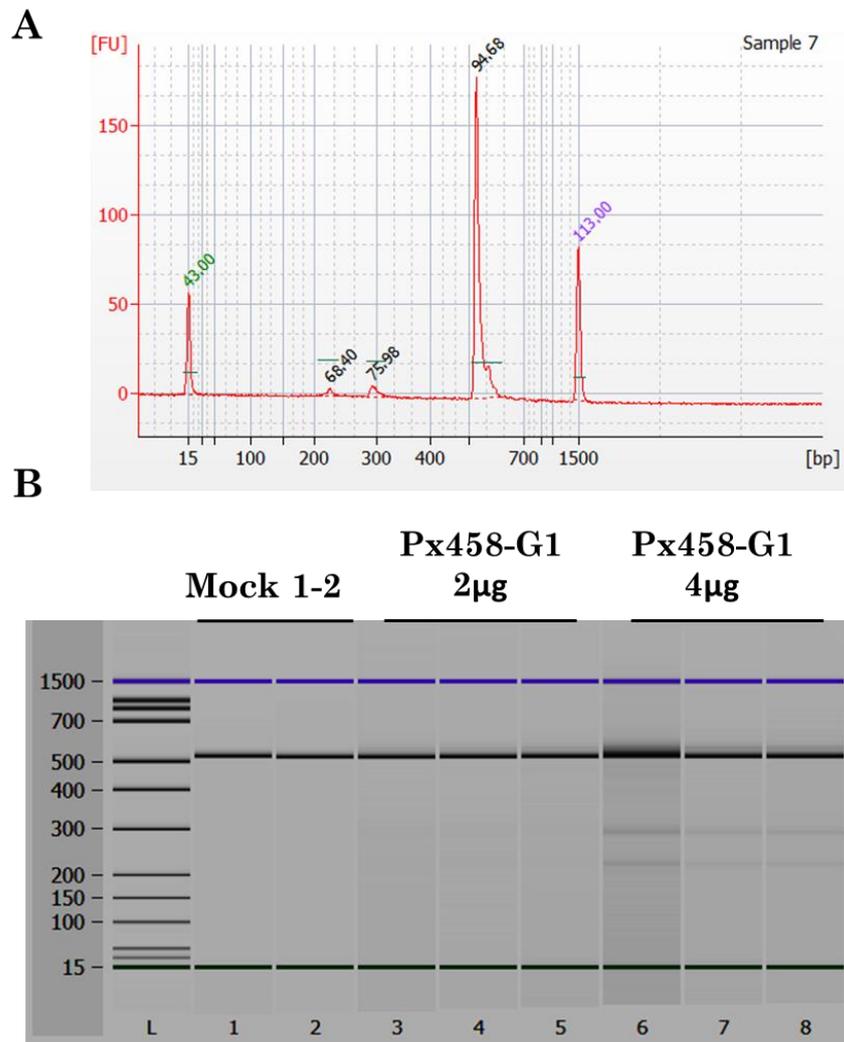


Figure 4-20 Agilent analysis of gene-modification in Δ 45-52 Immortalised Patient Myoblasts mediated by Px458-Guide 1 dual guide and Cas9 expressing plasmid.

A) An example electrophoretogram, showing the analysis of a 1 μ l aliquot of a T7E1 digest, undertaken on hybridised amplicons and resolved on a high sensitivity DNA chip, alongside 1 μ l of high sensitivity DNA ladder. Samples were amplified from 250ng genomic DNA harvested 48 hours post transfection from Immortalised Δ 45-52 patient myoblasts using 1960F and 2741R primers. Peaks at 15bps and 1500bps indicate higher and lower markers for the agilent, whilst peaks apparent at 512bps, 288bps and 224bps are indicative of uncut and cleavage products respectively; the area underneath which was used for a calculation of gene-modification. Notice in this instance the upper band appears to be a single defined band. **B)** A composite gel, generated by the Agilent programme to simulate the amplicon resolution that would be apparent on gel electrophoresis.

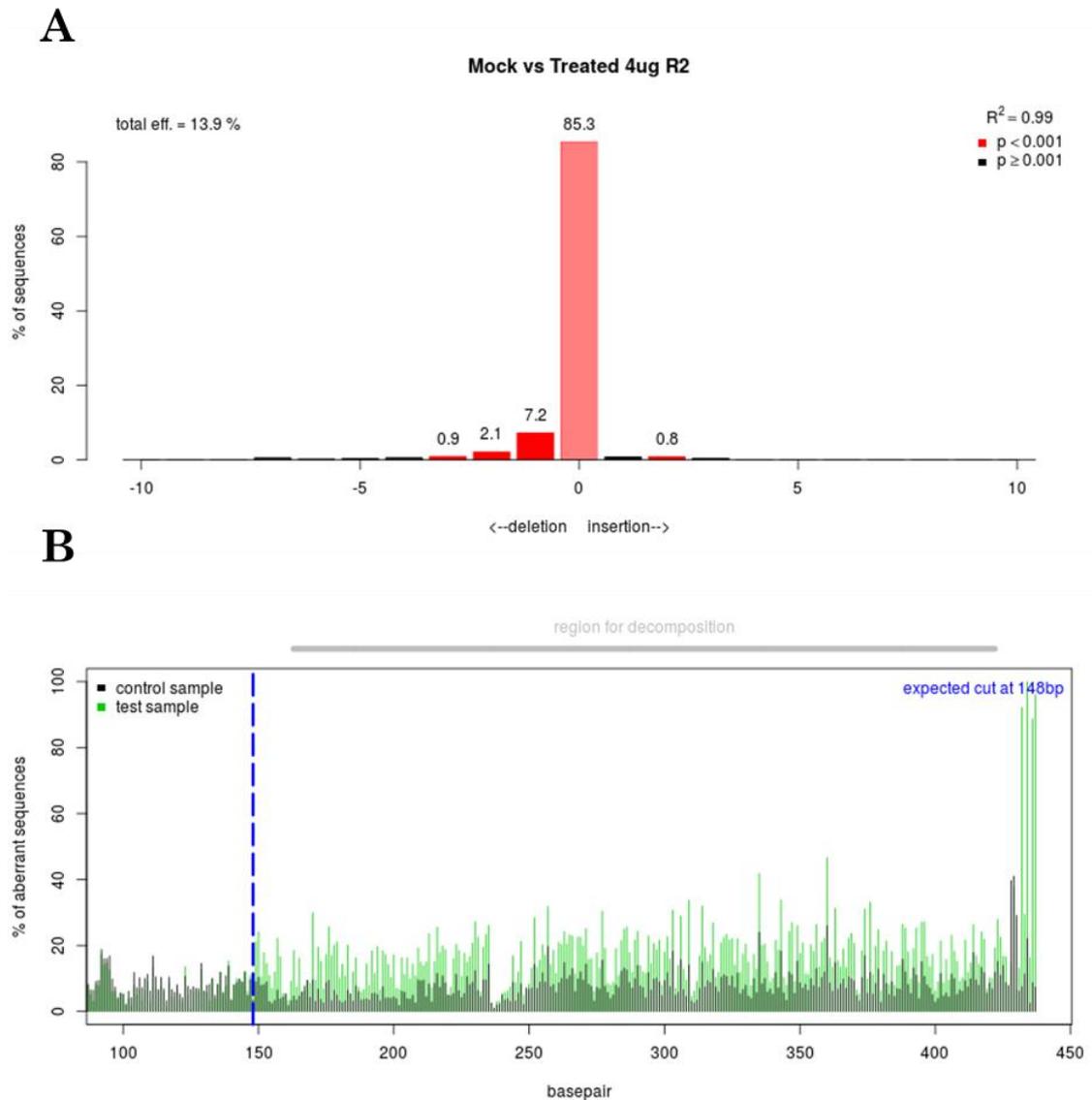


Figure 4-21 Tide analysis of gene-modification in Immortalised $\Delta 45-52$ patient myoblasts mediated by Px458-Guide 1 dual guide and Cas9 expressing plasmid.

A) A bar chart generated by TIDE analysis software (<https://tide-calculator.nki.nl/>), to indicate the percentage occurrence of InDels present in amplicons sequenced from a 4 μ g Px458-Guide 1 treatment. To achieve this, 250ng of Immortalised $\Delta 45-52$ patient myoblast genomic DNA harvested from Mock and Px458 treatments 48 hours post transfection were amplified with 1960F and 2741R primers. The resultant amplicons were purified and an aliquot corresponding to 75ng supplemented with 2 μ l 1960F at 10 μ M, was sent for sequencing by Eurofins MWG. Sequence traces obtained for mock and Px458 conditions were then compared by the TIDE software for the presence of decomposition at the guide 1 target site and the occurrence and type of InDels was estimated computationally. Note mock in this case was immortalised $\Delta 45-52$ patient myoblasts incubated in the presence of 12 μ l of Viafect transfection reagent alone. **B)** A decomposition trace for the sample described in part A. Dark green indicates comparable reads, light green indicates trace decomposition. The gene modification of this sample is approximately 14% however this sample appears to have a relatively high background.

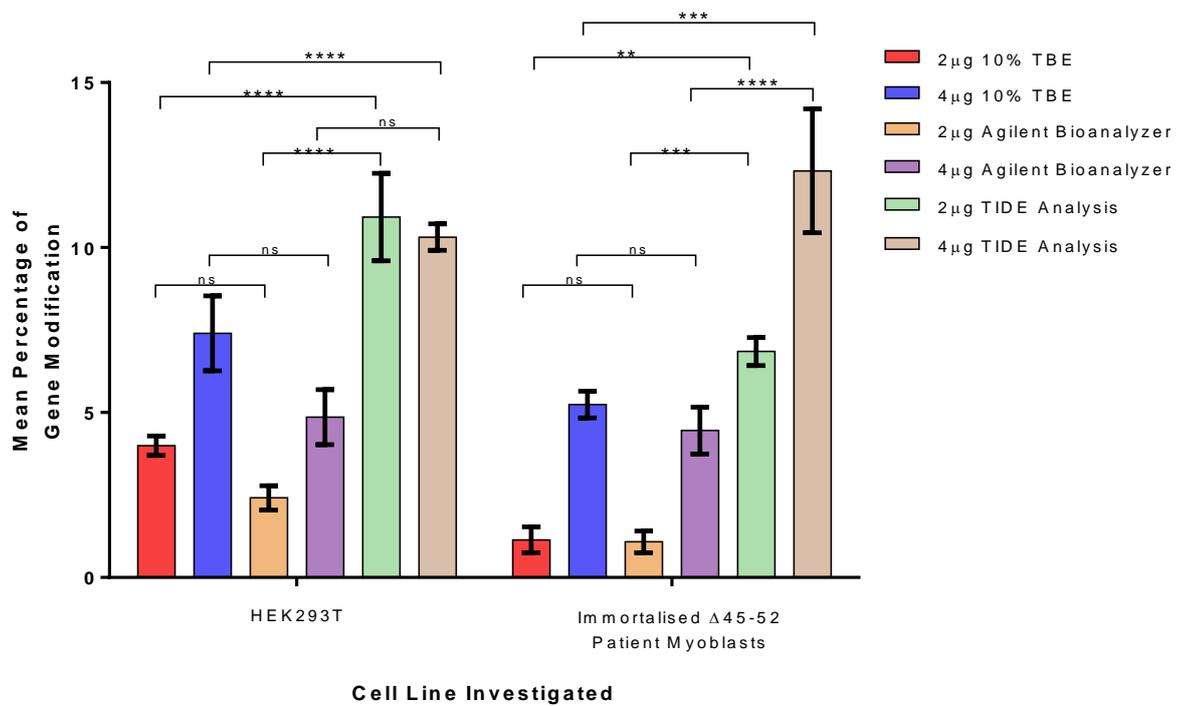


Figure 4-22 A graphical comparison of mean percentage gene modification estimated by the T7E1 assay quantified in two independent ways and by TIDE analysis.

A graphical summary of gene modification estimates attained from transient transfection of Px458-Guide 1 at 2 μg or 4 μg into two cell lines. Estimates were obtained by resolution of T7E1 cleavage products with either the use of 10% TBE polyacrylamide gels or the Agilent bioanalyzer. Moreover, a final method, TIDE analysis was also used to assess the occurrence of InDels. Notably, the mean percentage gene modification estimated from the methods trialled followed the same trends across both cell lines, with TIDE giving the highest estimates and agilent the lowest. All samples N=3 and the error bars represent the standard error of the mean. Samples in exception to this are except N=2 in HEK293T 2 μg TIDE analysis and Immortalised Δ45-52 patient myoblast agilent analysis.

4.6 Establishing In-Vitro Transcription of guide RNAs and Delivery Approaches in the Context of Gene-Editing.

Due to the moderate Px458 plasmid transfection efficiency observed in the immortalised $\Delta 45-52$ patient myoblasts, the ability to produce and deliver CRISPR gene-editing reagents in an RNA format was assessed. It was speculated that this alternative delivery format could improve upon the low to moderate efficiency of plasmid-based approaches and increase the chance of the investigation attaining successful outcomes.

4.6.1.1 Transient Transfection of GFP mRNA to HEK293T and Patient Myoblasts Harboured Deletions of Exon 45 -52:

Initially, assessment of RNA delivery to HEK293T and immortalised $\Delta 45-52$ patient myoblasts was undertaken using Lipofectamine Messenger Max and commercially produced GFP mRNA (Trilink). This was trialled using transfection reagent (μl): RNA (μg) ratios of 1.5:1 or 3:1, to deliver 2.5 μg GFP mRNA in accordance with the manufacturer's protocol. HEK293T and immortalised $\Delta 45-52$ patient myoblasts were seeded at 5×10^5 and 8.3×10^4 per well respectively and transfected when culture had attained 70% confluency in a 6-well plate.

Transfection efficiency was then evaluated using fluorescence microscopy imaging and FACS analysis. In HEK293T this resulted in a mean percentage of GFP positive cells of $58.48\% \pm 2.24$ and $88.83\% \pm 1.45$ SEM (N=3) at the 1.5:1 and 3:1 ratios respectively (Figure 4-23). This was deemed a statistically significant difference by an unpaired t-test ($p=0.0003$). In immortalised $\Delta 45-52$ patient myoblasts this resulted in a mean percentage of GFP cells, of $59.12\% \pm 1.17$ and 63.89 ± 1.32 SEM at 1.5:1 and 3:1 ratios respectively (Figure 4-24). Notably, the transfection efficiency of GFP mRNA into patient myoblasts at both 1.5:1 and 3:1 transfection reagent: mRNA ratios were comparable to transfection efficiencies observed at the lower 1.5:1 ratio in HEK293T cells.

4.6.2 Production and Genome Editing Efficiency of In-Vitro Transcribed Guide RNAs:

Following demonstration of successful GFP mRNA delivery to the cell lines of interest, Guide RNA 1 was in vitro transcribed. Conventional methods of in-vitro RNA production depend on linearization and purification of plasmid DNA, prior to proceeding with an *in vitro* transcription process. However the short length of chimeric single guide RNAs utilised by the Cas9 system renders them amenable to a 'one pot' PCR method described by Liang (X. Liang et al. 2015).

Briefly, this process anneals a customised forward primer and an invariant reverse primer to provide a short stretch of double stranded DNA encoding a single guide RNA that can be in-vitro transcribed. The 67bp forward primer comprises a T7 RNA polymerase promoter directly

upstream of a GX20 guide sequence. The GX20 guide sequence functions as the 'crRNA' in this system conferring specificity to the genomic locus of interest. This sequence is contiguous with 23bps of tracrRNA. Importantly, the invariant reverse primer initiates on 23 nucleotides complementary to the 23 nucleotide tracrRNA sequence encoded by the forward primer. Beyond this sequence the reverse primer comprised 60 nucleotides that serve to complete the tracrRNA and encode a polyA sequence.

A customised extended forward oligo was designed and synthesised to enable the production of guide 1 RNA. This was combined with an invariant extended reverse primer in a HotStart phusion PCR reaction to produce a double stranded DNA amplicon encoding single guide RNA 1, downstream of a T7 promoter. This amplicon was then purified using the Qiagen PCR Purification kit and 1.5µg of this amplicon was then subject to an overnight in-vitro transcription reaction. The resultant transcribed RNA was then DNase treated and purified again by a Qiagen PCR column system, with or without prior isopropanol precipitation and resolved on a 2% agarose gel to assess purity (Figure 4-25). The isopropanol precipitation of transcribed RNA was investigated as it was observed that RNA was inaccurately and more toxic upon delivery to cells if the isopropanol precipitation was included in the method.

Then 2.5µg of in vitro transcribed guide RNA from both purification streams was transfected into HEK293T.Cas9 cells at 70% confluency using a 3:1 ratio of Lipofectamine Messenger Max (µl): mRNA (µg). HEK293T.Cas9 culture genomic DNA was harvested 48 hours later, PCR amplified with 1960F and 2741R primers and 400ng of amplicon subject to the T7E1 assay and resolved on a 2% agarose gel (Figure 4-25). Notably, both guide purification methods produced comparable percentage gene-modification estimates of $5.50\% \pm 0.36$ and $5.28\% \pm 0.82$ SEM (N=3) with and without isopropanol precipitation respectively. This indicated transcribed guide 1 RNA was functional and able to direct Cas9 nuclease cleavage.

4.6.3 Production of In-Vitro Transcribed Cas9 mRNA and Genome Editing Efficacy in Immortalised Δ45-52 Patient Myoblasts:

To facilitate the extension of an RNA-based CRISPR/Cas delivery platform to cell lines that do not have a stable integration of Cas9, production of Cas9 mRNA was examined. In pursuit of this 10µg digests of HC-Cas9 plasmid (Figure 4-26) containing a T7 RNA polymerase promoter upstream of sequence optimised Cas9 was linearised with the SpeI restriction enzyme (Figure 4-27). The digest was subjected to Qiagen PCR purification, and used for in vitro transcription at 30°C for 4 hours with the T7 Ribomax Large Scale mRNA production kit (Promega, UK). The resulting transcripts were resolved on 3% gel to determine RNA purity and yield (Figure 4-25).

Cas9 mRNA and transcribed guide 1 RNA were then transfected into immortalised Δ 45-52 DMD patient myoblasts in two dose conditions: (i) 200ng guide 1 RNA with 1 μ g Cas9 RNA, or (ii) 400ng guide 1 RNA with 2 μ g Cas9 RNA. Harvested genomic DNA from the transfected myoblast cultures were PCR amplified, subjected to the T7E1 assay, and cleavage products resolved on 2% agarose electrophoresis gels (Figure 4-28). Striking gene modification was observed with estimates for the lower transfection dose being 17.4%, and the higher being 10.4%.

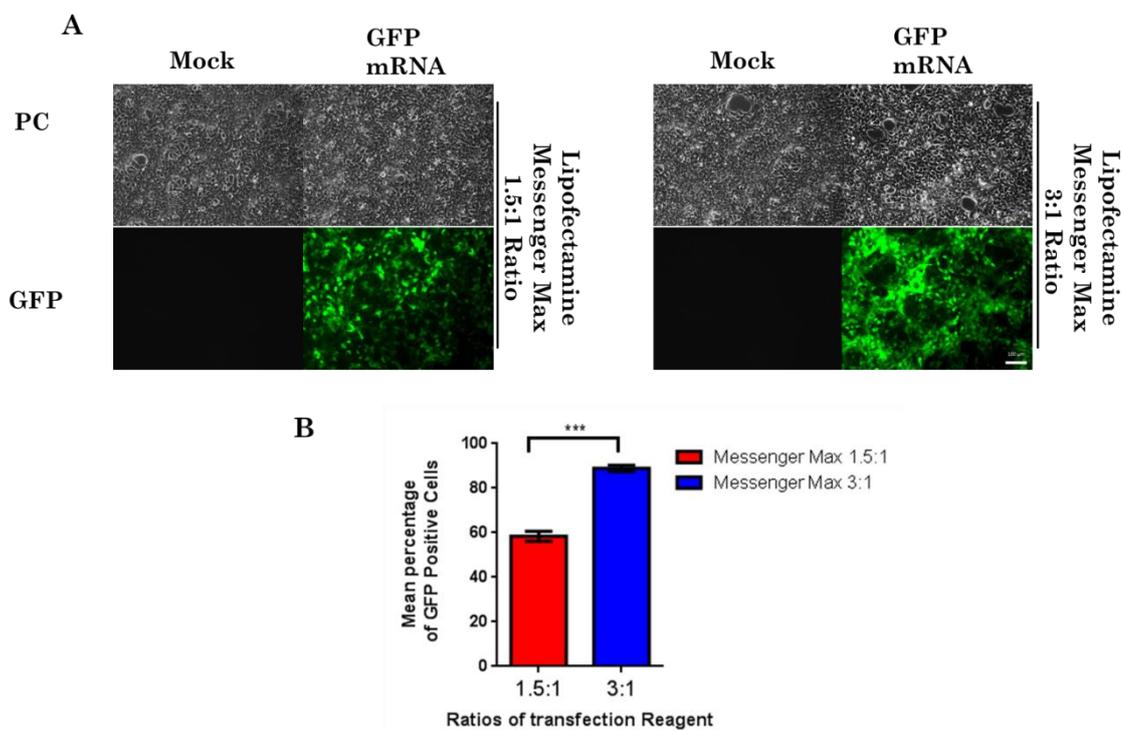


Figure 4-23 Transient transfection of HEK293T with 2.5 μ g of GFP mRNA as a proxy for transfection efficiency.

(A) Phase Contrast (PC) and fluorescent microscopy images were taken of HEK293T cells in the FITC (GFP) channel. Cells were transiently transfected with 2.5 μ g GFP mRNA at both 1.5:1 and 3:1 Lipofectamine Messenger Max reagent μ l to RNA μ g ratios. Transient transfections of 2.5 μ g GFP mRNA were undertaken alongside mock (M) samples, in which cells were treated with lipofectamine Messenger max reagent with: 3.75 μ l in 1.5:1 mock and 7.5 μ l in 3:1 mock respectively. All images were taken with the Zeiss microscope 48 hours post transfection at 100x magnification, phase contrast images were taken at 30ms exposure and GFP images taken at 800ms exposure. **(B)** A bar chart showing the mean percentage of GFP positive cells as assessed using flow cytometric analysis of treated cells fixed in 2% PFA-FACs buffer 48hours post transient transfection, with differing ratios of Messenger Max reagent to GFP mRNA. The resultant GFP positive population was determined to be significantly different between the 1.5:1 ration and 3:1 ratio by unpaired t-test ($p=0.0003$).

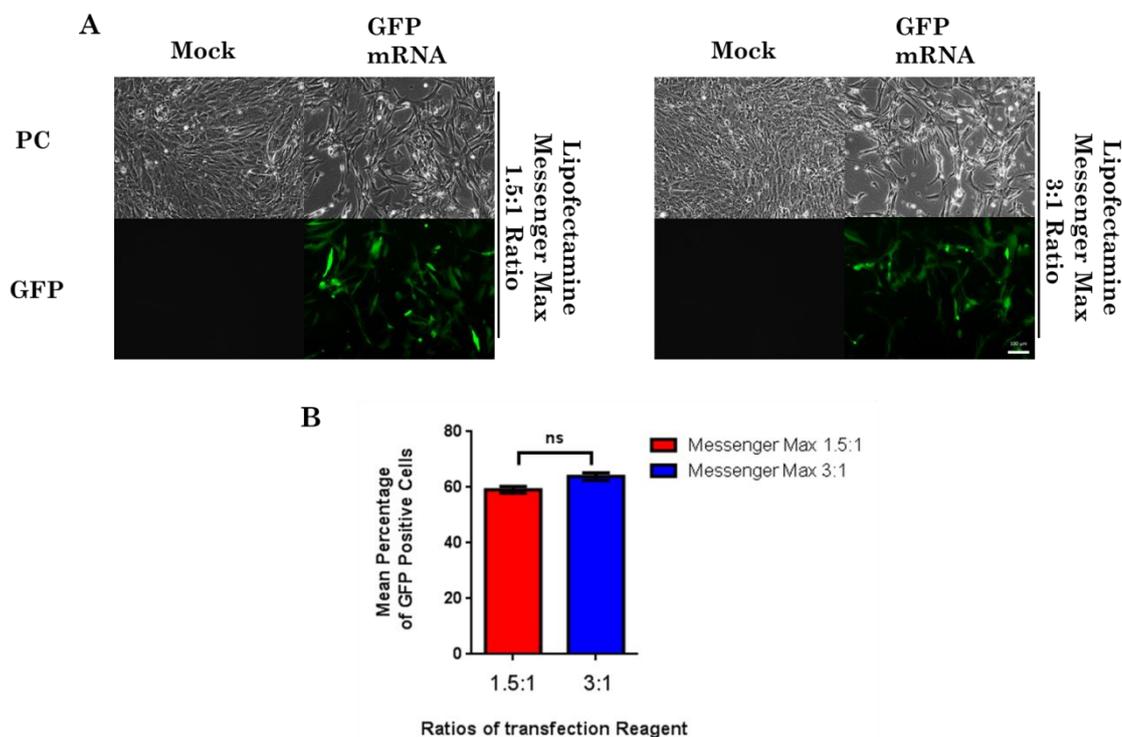


Figure 4-24 RNA Transient Transfection of Immortalised $\Delta 45-52$ patient myoblasts, using GFP mRNA expression as a proxy for transfection efficiency

A) Phase Contrast (PC) and fluorescent microscopy images were taken of the $\Delta 45-55$ patient myoblasts in the FITC (GFP) channel. Cells were transiently transfected, with 2.5 μ g GFP mRNA at both 1.5:1 and 3:1 Lipofectamine Messenger Max reagent μ l to RNA μ g ratios. Transient transfections of 2.5 μ g GFP mRNA were undertaken alongside mock (M) samples, in which cells were treated with lipofectamine Messenger max reagent with: 3.75 μ l in 1.5:1 mock and 7.5 μ l in 3:1 mock respectively. All images were taken with the Zeiss microscope 48 hours post transfection at 100x magnification, phase contrast images were taken at 30ms exposure and GFP images taken at 800ms exposure. **B)** A bar chart showing the mean percentage of GFP positive cells, determined using flow cytometric analysis of treated cells fixed in 2% PFA-FACs buffer 48hours post transient transfection, with differing ratios of messenger max reagent to GFP mRNA. There was no significant difference was found between the two transiently transfected populations.

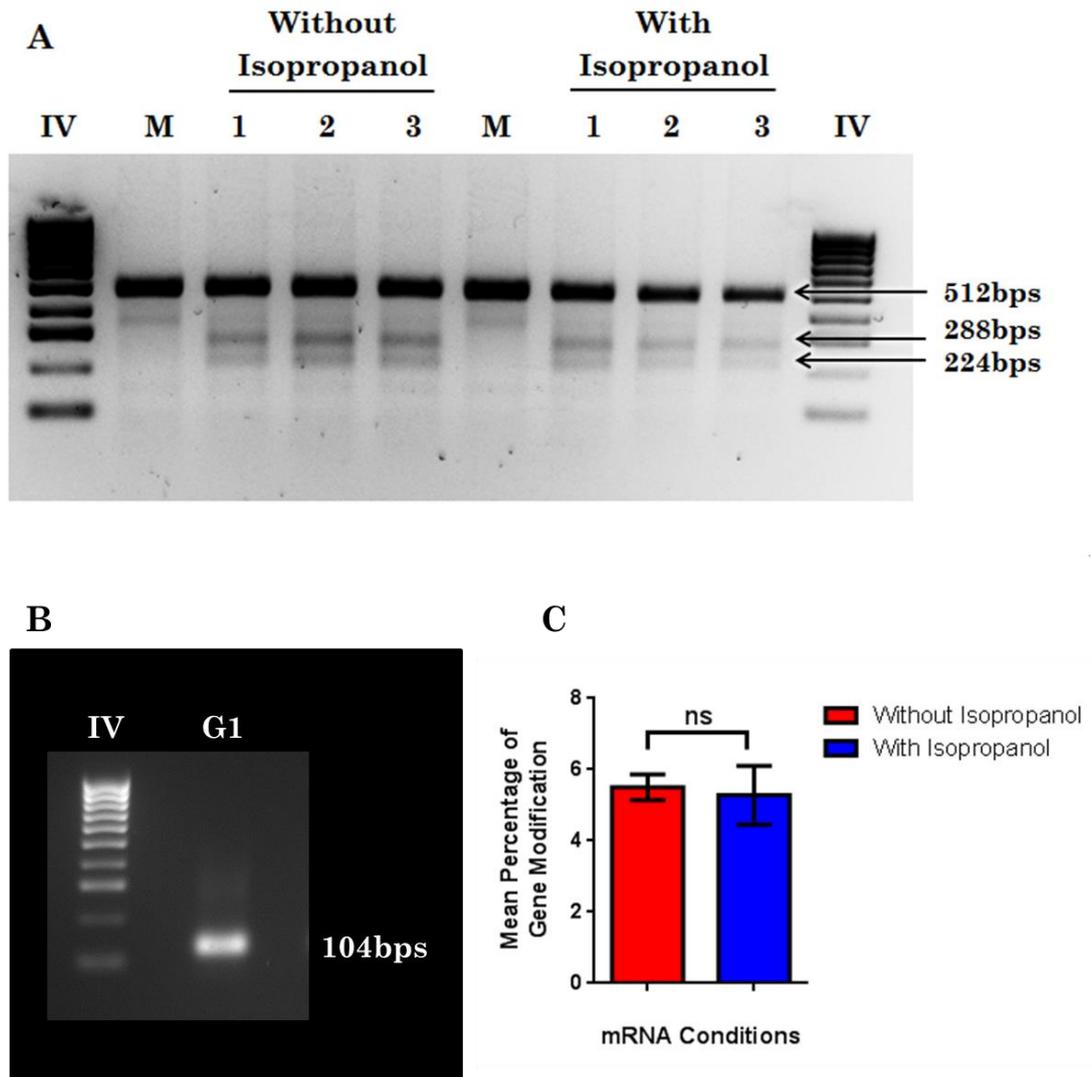


Figure 4-25 Production of *in vitro* transcribed guide 1 and subsequent T7E1 assay:

A) A T7E1 Assay performed on 400ng of DNA amplicon; amplified from 250ng of HEK293T.Cas9 genomic DNA, harvested 48 hours post transfection using 1960F and 2471R, with the proofreading Hotstart Phusion polymerase. Treatments included Mock (M) with 7.5 μ l of Lipofectamine Messenger Max alone and 2.5 μ g transcribed guide 1 RNA (purified with and without an isopropanol precipitation step. All samples were resolved on a 2% agarose gel, alongside hyperladder IV. Note all transfections were carried out using a 3:1 Lipofectamine messenger max transfection reagent: mRNA ratio. **B)** A 1 μ g aliquot of In Vitro transcribed Guide 1 RNA loaded onto a 2% agarose, alongside hyperladder IV. **C)** A graphical summary of gene modification estimates obtained through Image J analysis of the gel image shown in A. The difference between the gene modification estimates was deemed non-significant by an unpaired t-test (N=3).

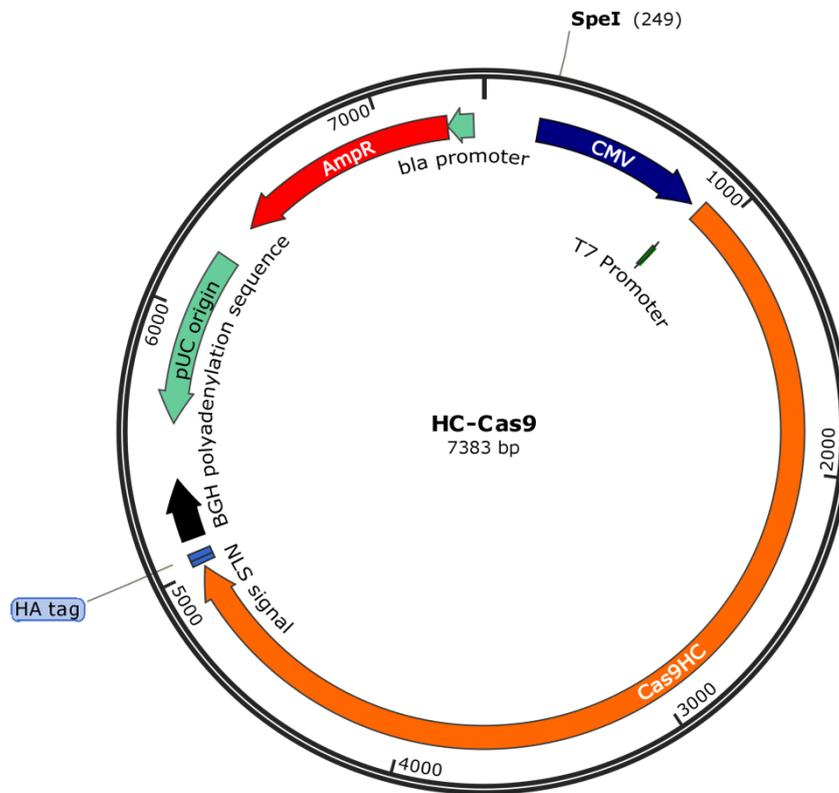


Figure 4-26 Ps3-Cas9-NLS-HA, A Cas9 Expression Cassette with an upstream T7 promoter amenable to In vitro transcription processes.

This contains human sequence optimised *Streptococcus Pyogenes* Cas9 downstream of cytomegalovirus (CMV) and a T7 RNA polymerase promoter (T7). Finally, the 3' end of Cas9 has a HA tag and nuclear localisation signal (NLS), prior to terminating in a bovine growth hormone polyA. This plasmid was used for the T7 RNA polymerase and was used for Cas9 in vitro transcription experiments. In addition it harbours an ampicillin resistance cassette, used during bacterial heat shock transformation and to select for plasmid in starter and maxi cultures. The plasmid was kindly gifted by Dr. Taeyoung Koo, from Professor Jin Soo Kim's lab.

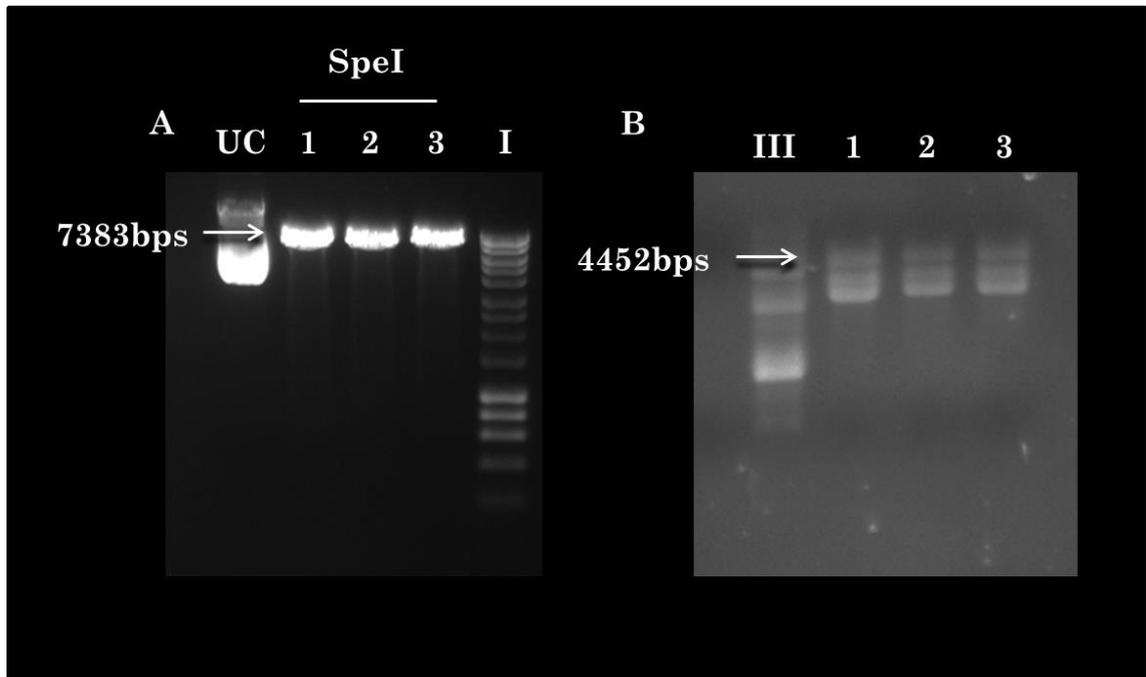
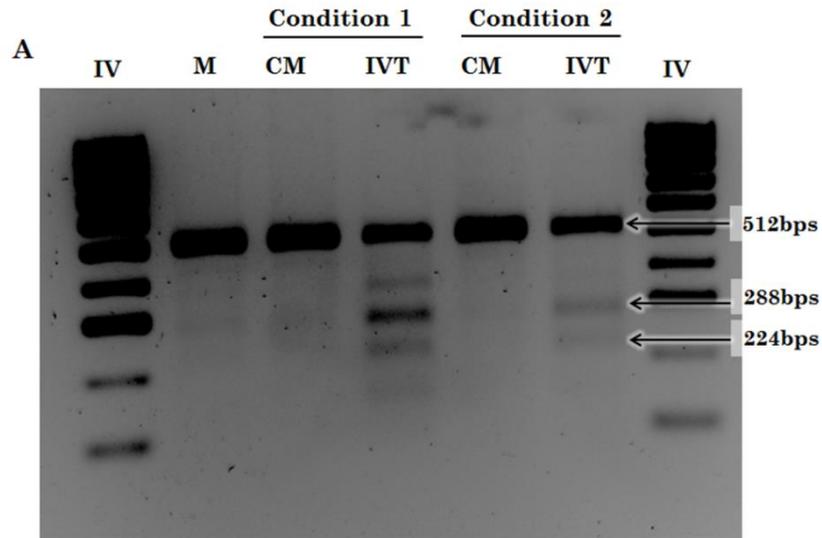


Figure 4-27 Images of agarose gels showing linearised HC-Cas9 and resultant mRNA production from *in vitro* transcription processes

A) Gel electrophoresis of Linearised Human sequence optimised Cas9 plasmid (HC-Cas9), digested with the SpeI restriction endonuclease in triplicate as denoted by 1, 2 and 3, alongside Uncut (UC) controls resolved on a 1% agarose gel alongside hyperladder I in 1X TAE buffer. This enzyme was selected as the restriction site upstream of the T7 promoter site. **B)** A 2 μ l aliquot of Cas9 mRNA, from the triplicate reactions undertaken, resolved on a 2.5% agarose gel alongside hyperladder III. Note all gels were resolved in 1x TAE buffer.



A bar chart showing the gene modification estimates of dual guide 1 and Cas9 mRNA delivery to immortalised $\Delta 45-52$ patient myoblasts

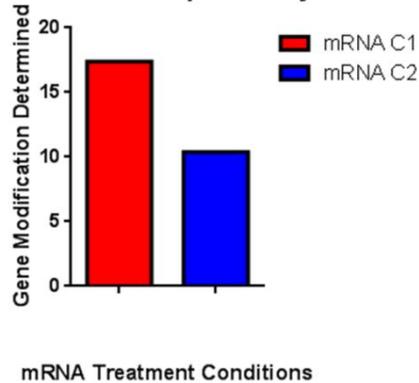


Figure 4-28 Examination of DNA cleavage following a dual transient transfection of in vitro transcribed guide 1 RNA and Cas9 mRNA, into $\Delta 45-52$ immortalised DMD patient myoblasts.

A) A T7E1 Assay performed on 400ng of DNA amplicon; amplified from 250ng of Immortalised $\Delta 45-52$ genomic DNA, harvested 48 hours post transfection using 1960F and 2471R, with the proofreading Hotstart Phusion polymerase. Dual transient transfection of transcribed guide 1 and Cas9 mRNA were undertaken at two conditions. Condition 1: 200ng transcribed guide 1 RNA and 1 μ g Cas9 mRNA and Condition 2: 400ng transcribed guide 1 RNA and 2 μ g Cas9 mRNA, with both Commercial Cas9 mRNA (CM) (using Cas9 mRNA from Trilink) and in-house *in vitro* transcribed Cas9 mRNA (IVT), the cleavage products following the T7E1 Assay were resolved on a 2% agarose gel alongside hyperladder IV in 1x TAE buffer. **B)** A graphical summary of gene modification estimated for the transient transfection of in house *in vitro* transcribed RNA using Image J analysis of condition 1 and 2 depicted on the gel image shown in part A. Unfortunately, the co-transfection of guide 1 RNA alongside commercial cas9 mRNA did not produce quantifiable gene modification.

4.6.4 Production and Assessment of Cas9 Protein

Finally, Cas9 protein was produced with the outlook to generate ribonucleoprotein (RNP) complexes by hybridising the protein to transcribed guide RNA, as first described by Kim et al (2014). This form of Cas9 delivery is becoming increasingly favoured in the gene-editing field owing to its ability to be introduced to hard-to-transfect cells lines (Kim et al. 2014), low toxicity profiles (Kim et al. 2014; Zuris et al. 2014; Yu et al. 2016) and finally transient expression profile (Kim et al. 2014) which reduces the propensity for off-target effects (Kim et al. 2014; Zuris et al. 2014; Tycko et al. 2016; Cho et al. 2014). These parameters could prove pertinent to the successful implementation of a gene editing strategy in the context of DMD patient myoblasts.

4.6.4.1 Production, Concentration and Purification Processes for Cas9 Protein.

The production of Cas9 protein was undertaken in Paris with the support and technical expertise of Dr. Anne DeCian in the lab of Dr. Jean-Paul Concordet, following the protocol of (Jinek et al. 2012). A pET-based expression plasmid encoding Cas9 protein (residues 1 -1368) was modified from the variant described in Jinek et al (2012), to include two further Nuclear Localisation Signals (NLS) this was to improve the shuttling of Cas9 protein to the nucleus of the cell, and subsequently enhance DNA modification achieved by the nuclease. The modified plasmid was used to produce Cas9 protein (Figure 4-29). This plasmid encodes a fusion protein with N- and C- terminal hexahistidine (His) tags, and an N-terminal maltose binding protein (MBP) tags. The MBP tag functions to improve the solubility of recombinant Cas9 in *E. coli* and is separated from Cas9 by a tobacco etch virus protease cleavage site (TEV).

Briefly, the pET-plasmid expressing Cas9 was introduced to chemically competent Rosetta 2 BL21 *E.Coli* derivatives and the expression of Cas9 protein was induced with 0.2mM IPTG a mimic of the allolactose that triggers the lac operon. The bacterial culture was pelleted, re-suspended in lysis buffer and sonicated. This crude lysate was loaded and immobilised on a HisTrapCrude FF 5ml affinity column prior to being competitively eluted in 3ml fractions with increasing concentrations of imidazole. A 2 μ l aliquot of each of the fractions was visualised on 8% SDS-PAGE and visualized with Instant Blue staining. This SDS-PAGE combined with absorption peaks at 260nm and 280nm was used to ascertain which column fractions contained Cas9 protein for subsequent enrichment processes (Figure 4-30).

Following HisTrap enrichment the Cas9 was concentrated, liberated from the MBP tag, and further purified using an SP sepharose ion exchange and size exclusion column chromatography. In each case 2 μ l aliquots of elutes and 260:280 ratios were used to assess Cas9 protein containing fractions (Figure 4-31 and Figure 4-32).

At this stage fractions identified as containing pure Cas9 were concentrated to 30 μ M and 90 μ M and are denoted P108 and the P109. The P108 fraction of Cas9 protein at 30 μ M or 4.7mg/ml concentration was pipetted into 20 μ l aliquots and snap-frozen in liquid nitrogen. This P108 aliquot is used in throughout this investigation.

4.6.4.2 Validation of Cas9 Nuclease Activity via an In-Vitro Cleavage Assay:

Once Cas9 protein was produced and purified an assessment of its activity to introduce DSBs was undertaken. Cas9 protein was complexed with an in vitro transcribed test guide RNA 'GFP-T2' which targets a GFP gene sequence (Mali, Yang, et al. 2013) and the RNP complex incubated with an Xmn1-linearised Vgej-AAV plasmid which contains a target GFP sequence (Figure 4-33-A). Using this assay the Cas9 nuclease was shown to be functional (Figure 4-33-B) by the emergence of the expected cleavage products of 3245bps and 2637bps.

4.6.4.3 Delivery of Cas9 Protein to HEK293T Cells:

Once the Cas9 protein was produced and validated by the *in-vitro* cleavage assay, the next milestone for the development of this system as a suitable gene editing strategy was the delivery of the Cas9 RNP to a mammalian cell culture. To facilitate this, Dr. DeCian kindly provided GFP- fused Cas9 protein which was transfected into HEK293T cells with the use of a 3:1 Lipofectamine messenger max (μ l) : protein (μ g) using doses of 2.5, 5, 10 and 20 μ g Cas9-GFP protein (Figure 4-34). These doses were selected due to the wide range of doses used throughout literature (S. Kim et al. 2014; Lin et al. 2014; X. Liang et al. 2015; Yu et al. 2016). Fluorescence microscopy images of the cells 3 hours after treatment showed GFP positive cells were present at all doses of Cas9-GFP protein with the signal dissipated by the 12 hours.

4.6.4.4 Assessment of Gene Editing in HEK293T Cells Mediated by Cas9 RNP complexes:

Once successful delivery had been demonstrated, RNP complexes were produced, comprised of Cas9 protein and transcribed guide 1 RNA. RNPs were formed by heating the guide 1 RNA to 98°C for two minutes, cooling on ice, and then incubation at room temperature for 10 minutes with Cas9 protein at 4:1 and 2:1 ratios (ug: ug), comprising of 2.5 μ g:0.625 μ g and 2.5 μ g:1.25 μ g mixtures respectively. The RNP complexes were introduced into HEK293T cells using Lipofectamine CRISPRMax (Yu et al. 2016). The human *DMD* intron 1 site was assessed for gene modification using genomic DNA harvested 24 hours post treatment using PCR with 1960F and 2741R primers, and the standard T7E1 assay. The T7E1 assay with the 2:1 and 4:1 ratio RNPs gave mean percentage gene modification values of 10.21% \pm 2.81 and 7.85% \pm 0.10 SEM respectively (Figure 4-35). In addition RNP-treated and control amplicons from the mock conditions were sequenced and assessed by TIDE which yielded mean percentage InDels of

31.5% \pm 2.36 and 25.7% \pm 3.71 SEM at 2:1 and 4:1 ratios respectively (Figure 4-35). This provides a demonstration of high efficiency gene-editing in mammalian cells using Cas9 RNPs as the delivery format.

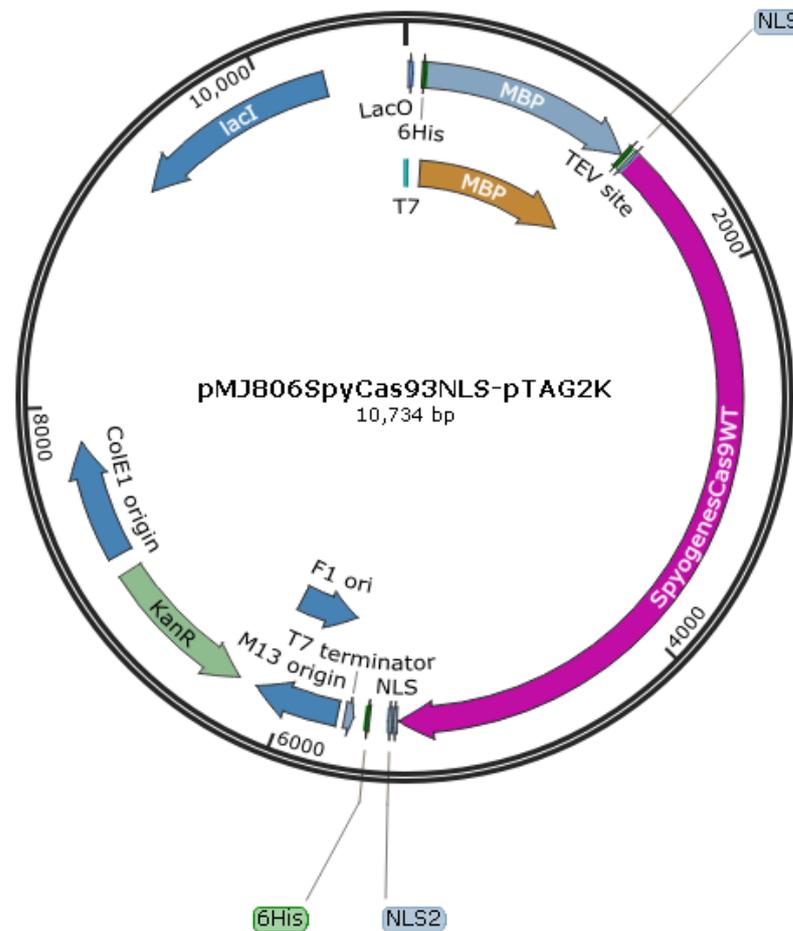


Figure 4-29 A schematic map of the custom Pet plasmid expressing SpCas9 protein used in protein production processes.

Custom Pet plasmid used to transform Rosetta BL2 and induced to produce Cas9 protein, with an N terminal hexahistidine-maltose binding protein (His-MBP) separated from Cas9 protein by a tobacco etch virus protease cleavage site. In addition, at the 3' terminal of the Cas9, the plasmid harbours 6 x histidine tags for the process of purification using a nickel column. Note, this plasmid was modified from the PMJ806 plasmid from Addgene, which was used in the original production of Cas9 protein (Jinek et al. 2012), by cloning additional nuclear localisation signals in the 5' and 3' ends of the protein. Finally, the plasmid harbours a kanamycin resistance cassette, used during bacterial heat shock transformation and to select for plasmid in starter and maxi cultures.

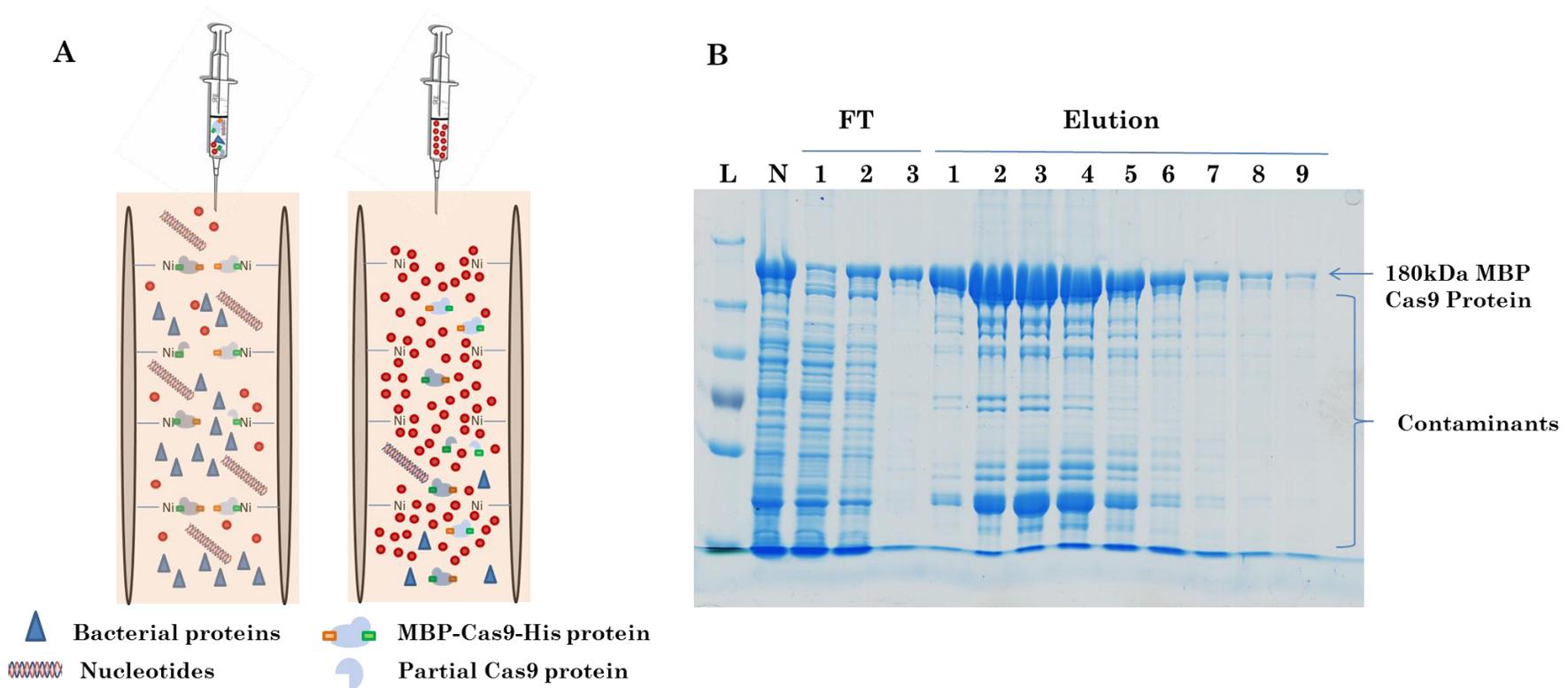


Figure 4-30 Purification of MBP-Cas9 protein, by loading washing and eluting the samples from a crude lysate histrap column:

A) A schematic of histrap crude FF 5ml column purification of Cas9 protein. The left column illustrates the loading process of Rosetta BL2 crude lysate post transformation with a custom Pet-plasmid encoding Cas9 protein and induction with 0.2mM IPTG. The hexahistidine tags on the C and N terminals of the Cas9 protein have affinity to and are immobilised by nickel present in the column. The column on the right illustrates the elution process in which Cas9 protein – nickel binding is out competed using a buffer with a high concentration of Imidazole, which has a higher affinity for nickel. **B)** A 4-8% SDS PAGE gradient gel loaded with 2µl of flow through and sequential elution aliquots, these aliquots were selected on the basis of the Akter 260:280 absorption traces (Not shown). The samples were run alongside Page Ruler Plus prestained protein ladder and visualised with instant blue.

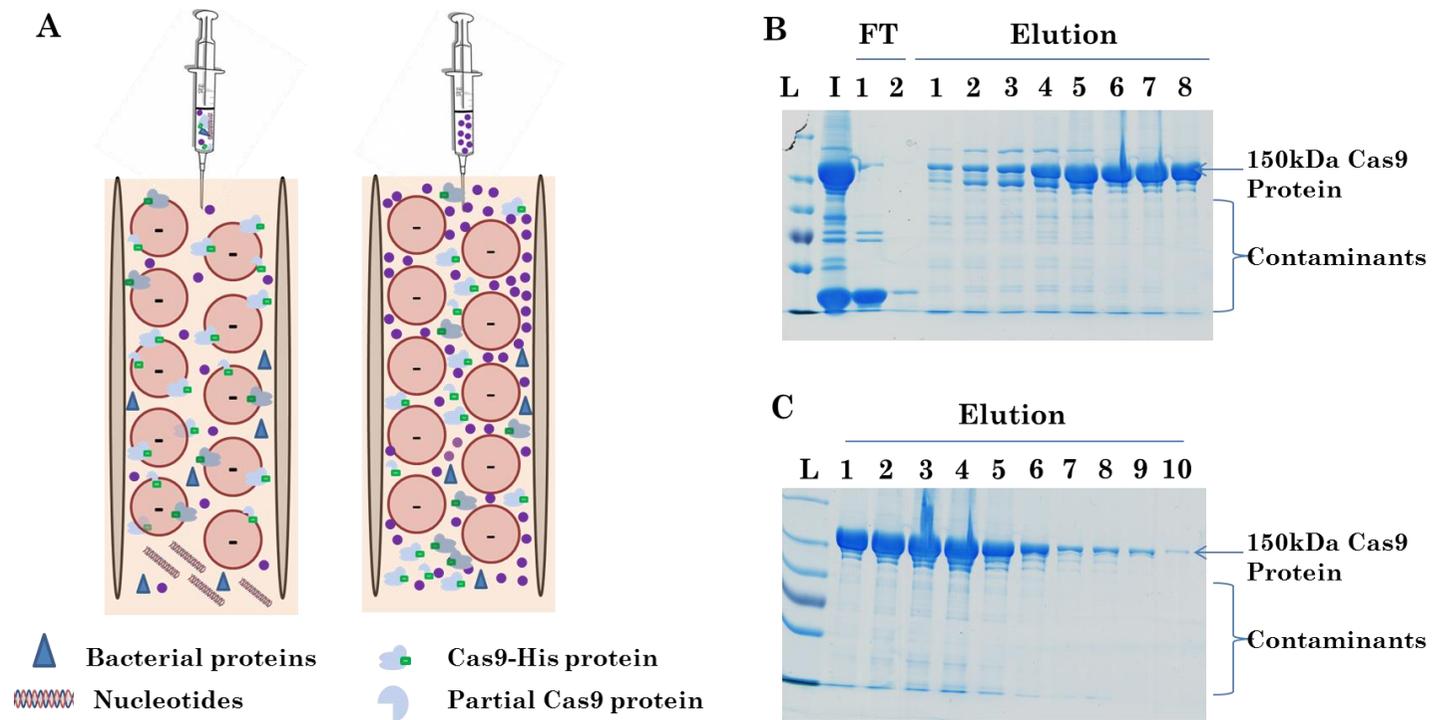


Figure 4-31 Purification of Cas9 protein, liberated from the MBP tag, on an ion-exchange Sepharose Column

A) A schematic of Sepharose column purification of Cas9 protein. The left column illustrates the loading process of concentrated TEV cleaved Cas9 protein post-HisTrap purification and dialysis. The positive charge of the Cas9 protein means it will have affinity to the negatively charged beads within this column and thus will be immobilised. The right column demonstrates the elution process which was undertaken by using a high concentration of 1M KCl; the ions dissociate and out compete Cas9 protein for the negatively charged beads. **B)** A 4-8% SDS PAGE gradient gel loaded with 2µl of the initial (I) protein, flow through (FT) and sequential elution aliquots for pool 1, which was kept separate to ascertain purity based upon indication from the Akter 260:280 traces. **C)** The same as B, except for pool 2, which was implicated to be slightly purer than the former. The samples were run alongside Page Ruler Plus prestained protein ladder (L) and visualised with Instant Blue.

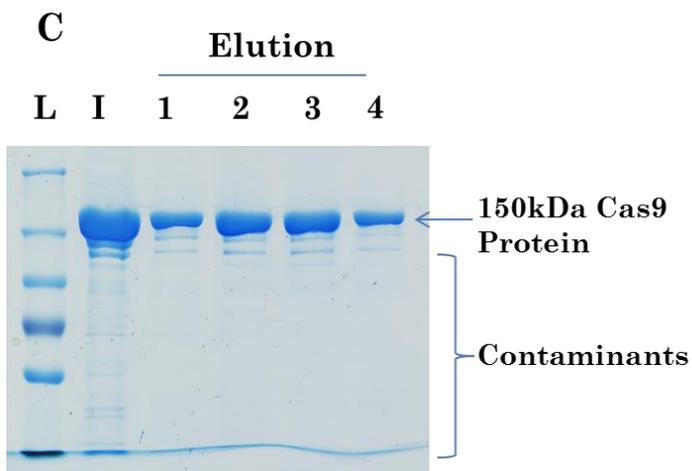
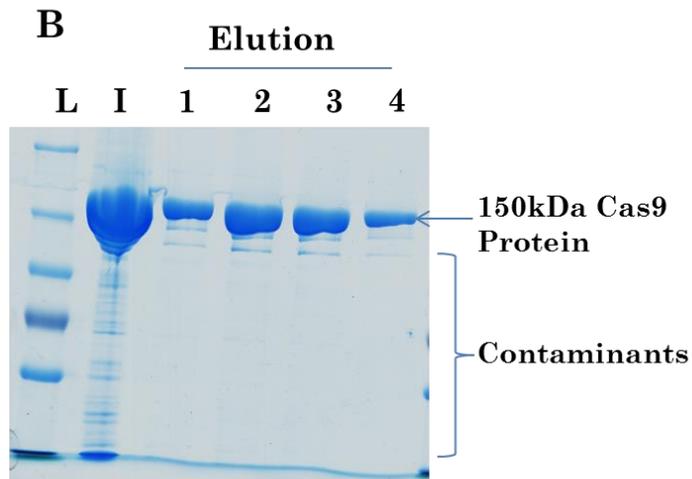
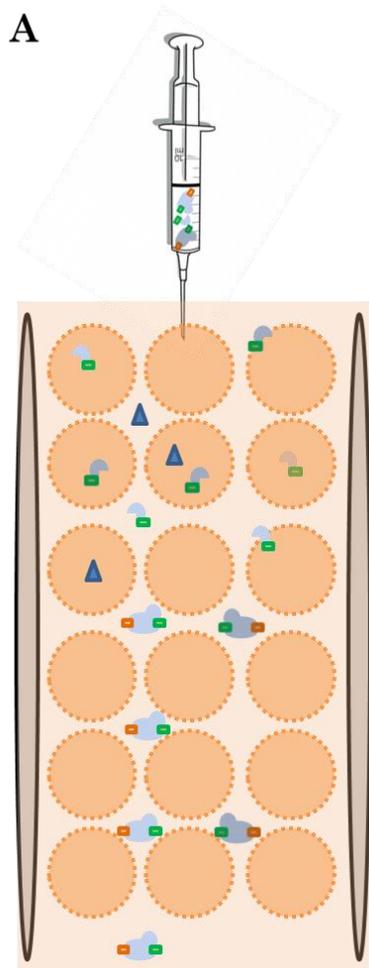
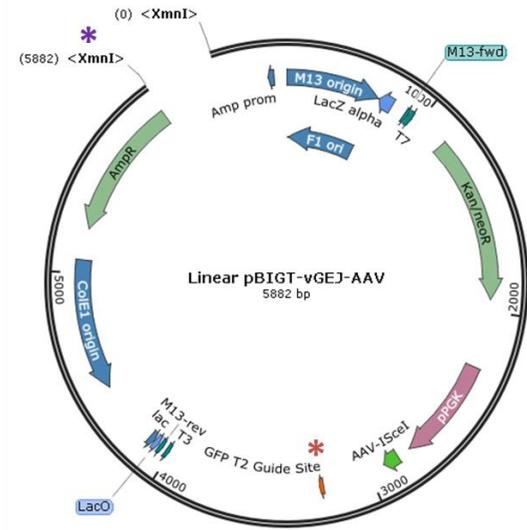


Figure 4-32 Purification of Cas9 protein using a Superdex size-exclusion column:

A) A schematic of Superdex Size exclusion of Cas9 protein. The larger 150kDa Cas9 protein will be excluded from the porous beads and thus eluted in the void volume. In contrast smaller contaminants will be able to permeate the beads and thus will be eluted later. **B)** A 4-8% SDS PAGE gradient gel loaded with 2 μ l of the Sepharose purified initial (I) products and sequential elution aliquots for pool 1, based upon indication from the Akter 260:280 traces. **C)** The same as B except for pool 2. The samples were run alongside Page Ruler Plus prestained protein ladder and visualised with Instant Blue. Notably Pool 1 became P109 and Pool 2 P108, in accordance with the production numbers of Dr. Anne DeCian. Both samples in this process had similar purification.

A



B

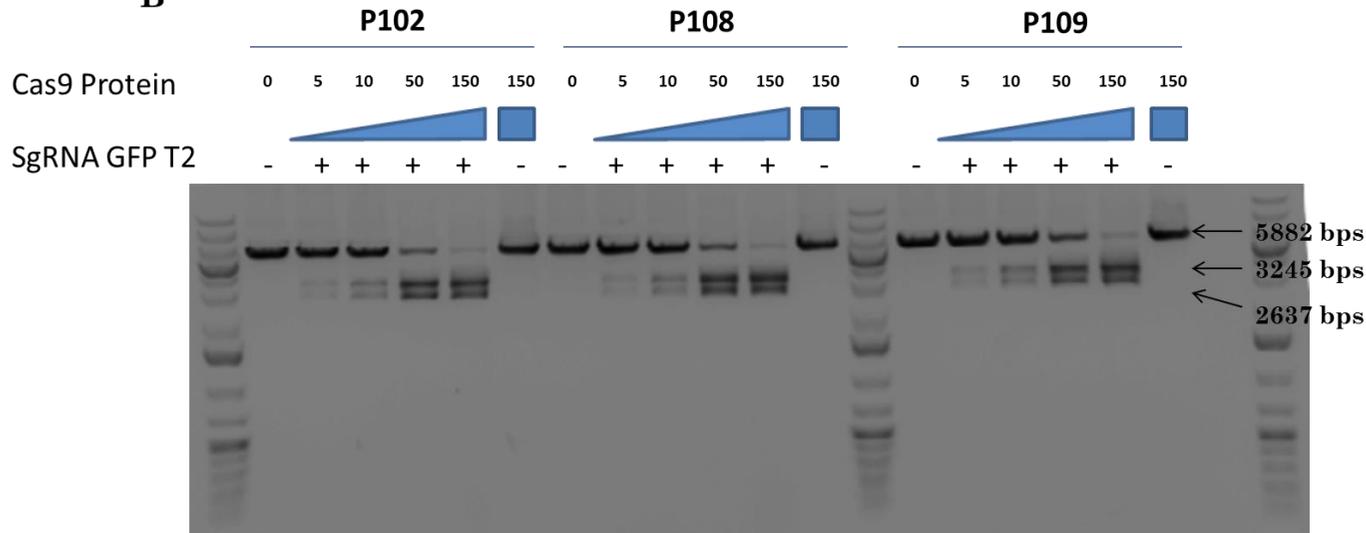


Figure 4-33 In Vitro Cleavage Assay to determine Cas9 protein functionality.

A) A schematic of pBIGT- vGEJ-AAV plasmid used in the *in vitro* cleavage process, linearized at the XmnI restriction site denoted by *. Importantly, this plasmid also has the site required for sgRNA GFP T2, a control guide RNA sequence (Mali et al. 2013). **B)** An assay of Cas9 protein activity of P108 and P109, alongside P102 a previously characterised Cas9 control resolved on a 0.75% agarose gel, cast in ethidium bromide and resolved in 0.5X TBE buffer. Varying nM concentrations of ribonucleoprotein complexes (RNPs) comprised of Cas9 protein and guide RNA were used in this investigation including 5, 10, 50 and 150nM. RNPs were made by using Cas9: sgRNA at a 2:1 ratio respectively. In addition 5nM of a XmnI linearized pBIGT-vGEJ-AAV plasmid (200ng), which contained a suitable guide target site within the eGFP was digested with the resultant RNP for 30 minutes at 37°C, prior to being halted with proteases and subjected to electrophoresis. The image was inverted to improve resolution of cleavage products.

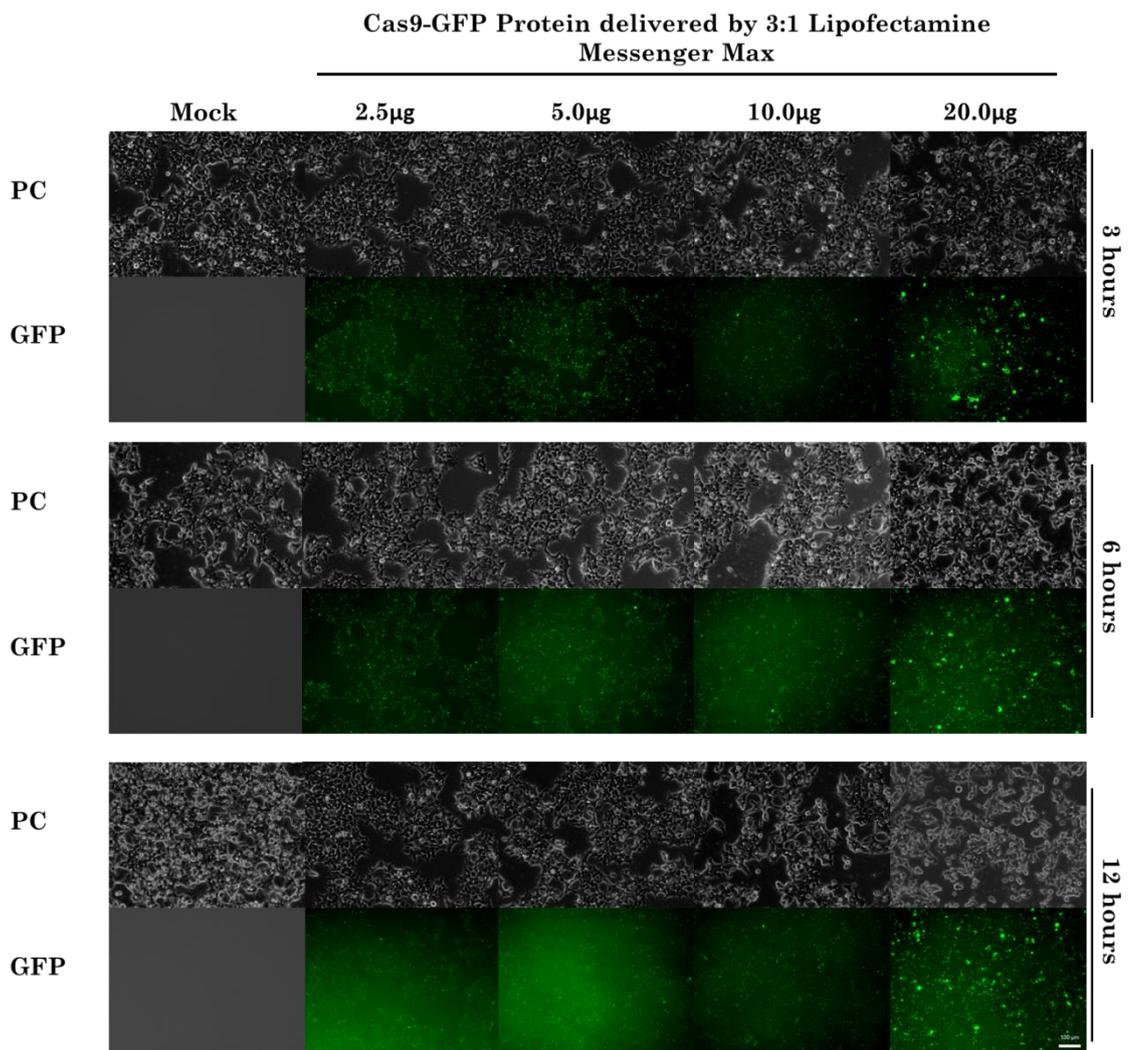


Figure 4-34 Introduction of GFP-fused Cas9 protein in HEK293T using Lipofectamine CRISPRMax reagent.

HEK293T cells were seeded at 3×10^5 in a six well plate and transfected with a variety of Cas9-GFP protein doses 24 hours later when they reached at a confluency of approximately 50%, consistent with the 30-70% indicated in the protocol. Doses of the Cas9 protein included: 2.5 μ g, 5 μ g, 10 μ g and 20 μ g. These were transfected at a 3:1 transfection reagent μ l: Cas9 protein μ g ratio. In consideration of high Cas9 protein turnover, cells were imaged at 3, 6 and 12 hour time points. Images were taken at 100x magnification; phase contrast exposure of 30ms and GFP exposure of 4s was used. Cells appeared green consistent with the uptake of the Cas9-GFP protein, the cells exhibited low mean fluorescence, with a FITC signal observed most prominent with the 2.5 μ g dose at the 3 hour time point. Interestingly, over time cells are more challenging to discern likely in accordance with the degradation of the low level fluorescence. Moreover, a high background makes the determination of cells challenging; this could be due to Cas9-GFP protein being present in the media. Finally, the morphology of the cells at the higher transfection amount in the mock sample and the 20 μ g treated sample appears to change over the time course suggesting the reagent is associated with some toxicity.

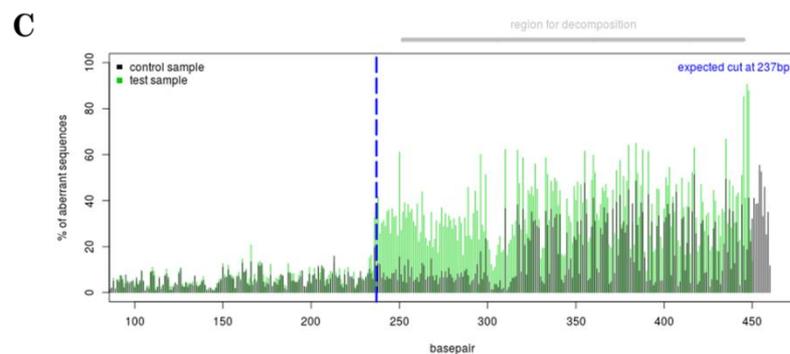
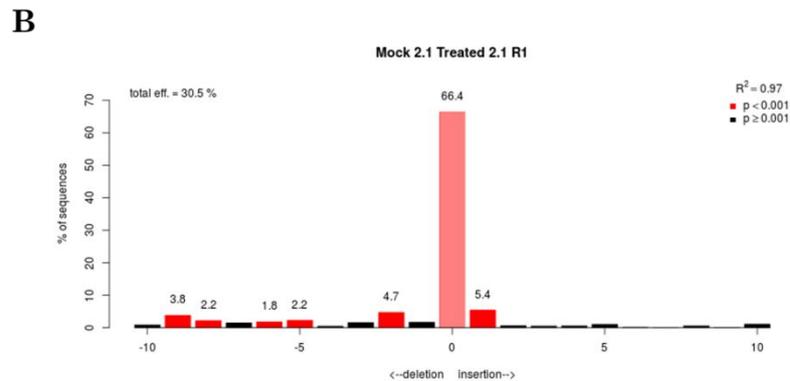
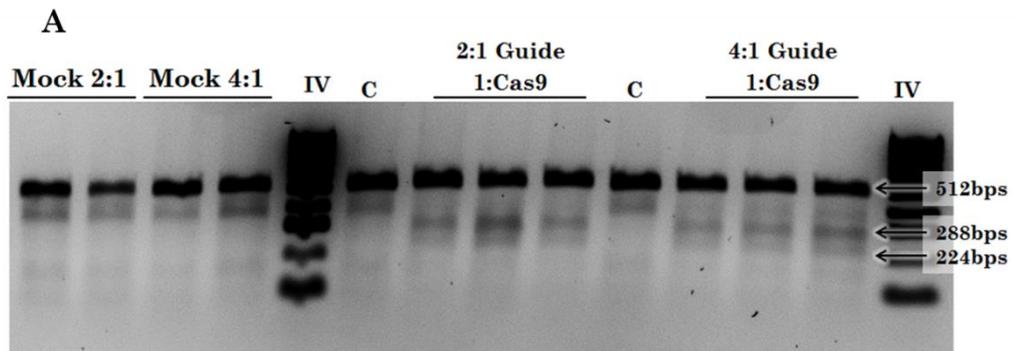


Figure 4-35 Gene- modification determined for Cas9 protein HEK293T treated samples using TIDE.

A) A T7E1 assay performed on 400ng of purified amplicons produced by PCR of genomic DNA harvested from HEK293T transfected with 2:1 and 4:1 Cas9 protein: guide RNA RNP mixtures 24 hours post treatment. Treatments comprised of 2.5 μ g of Cas9:1.25 μ g guide 1 RNA and 2.5 μ g Cas9: 0.625 μ g Guide 1 RNA. Amplifications were performed using 1960F and 2741R primers with the proofreading Hotstart Phusion polymerase, hybridised and digested with T7E1, prior to being resolved on a 2% agarose gel in 1X TAE buffer alongside Hyperladder IV. Notably, gene-modification is absent across mock conditions, in which cells were treated with 7.5 μ l Lipofectamine CRISPR Max and in Cas9 alone samples (C) in which 2.5 μ g of Cas9 protein was transiently transfected in the absence of in vitro transcribed RNA.. Mean gene-modification was determined to be 10.2% \pm SD and 7.8% \pm SD from 2.1 and 4.1 conditions. **B)** An example bar chart generated by TIDE analysis software to indicate the INDEL spectrum present in an amplicon from a 2:1 RNP transiently transfected HEK293T genomic sample from decomposition determined relative to a mock baseline. **C)** A decomposition trace for the sample described in B. Dark green indicates comparable reads, light green indicates trace decomposition. The gene modification estimated for this sample is approximately 30.1%.

4.7 Discussion

The repurposed CRISPR/Cas9 system was originally transiently transfected in a plasmid format to achieve gene-modification in mammalian cells (Mali, Yang, et al. 2013; Jinek et al. 2013; Cong et al. 2013; Fu et al. 2014). Since these early experiments the Cas9 plasmid constructs have been further developed, a chimeric single guide RNA has been generated to replace the dual crRNA and tracrRNA (sgRNA where the two components crRNA and TracrRNA are fused by a GAAA tetraloop), the Cas9 cDNA has been subject to sequence optimisation and finally the tethering of Cas9 to GFP (Jinek et al. 2013; Ran et al. 2013). The established protocols for the transient transfection of gene-editing plasmids and the aforementioned advances, that have served to simplify delivery, improve Cas9 protein expression and simplify the assessment of transient transfection efficacy made this delivery strategy attractive for this investigation. As such, plasmid based gene-editing techniques were initially trialled with HEK293T cells and immortalised Δ45-55 patient myoblasts.

4.7.1 Establishing Mismatch Detection Assays

Once western blotting confirmed that both Px165 and Px458 plasmids express Cas9 protein following transient transfection into HEK293T cells, an examination of the efficacy of in silico guide designs was warranted. In pursuit of this co-transfection of U6 guide plasmids encoding guides 1-4 alongside the Cas9 expressing Px165 plasmid was first trialled at doses of 1µg:1µg or 2µg:2µg. DNA was harvested and analysed with established mismatch detection assays, this being either the Surveyor (Qiu et al. 2004; Guschin et al. 2010) or T7E1 (de Massy et al. 1987; White et al. 1997) assay.

Initially cleavage products from these assessments were resolved on agarose and proved challenging to visualise and quantify. When agarose was used to resolve cleavage products, quantifiable gene modification could not be quantified across either 1µg Guide: 1µg Px165 for either the Surveyor or T7E1 assay. In contrast 2µg of guide 1, 2 or 4 co-transfected with 2µg of Px165 could not be quantified in the surveyor assay, but mean gene modification estimates of 11, 5 and 12% were estimated in the T7E1 assay respectively. This finding could be due to differing resolution of the Surveyor and T7E1 agarose gels. However, a recent study by Vouillot et al 2015, comparing the Surveyor and T7E1 mismatch detection systems highlighted that the T7E1 enzyme was sensitive relative to the Surveyor nuclease (Vouillot et al. 2015). In line with study, current estimates suggests that the T7E1 assay can detect in the range of 0.5- 5% InDels (Zhu et al. 2015) relative to the 3 -5% detected by the Surveyor assay (Qiu et al. 2004; Fu et al. 2013). This was pertinent to the decision to use T7E1 for future gene modification estimates.

Strikingly, when the same 1µg: 1µg and 2µg: 2µg were treated with T7E1 and resolved on a 10% TBE polyacrylamide gel, there was an apparent increase in gene-modification estimates. Guide RNAs 1 and 2 gave gene modification estimates in the 1µg:1µg conditions of 4 and 7% respectively. Furthermore, at in the 2µg:2µg condition an apparent increase in gene modification was observed for all guide RNAs, with a striking 2 fold difference in the case of guides 1 and 4 giving estimates of 27.1 and 24.9%. An important implication of this is that whilst mismatch detection cleavage products are resolved both on agarose (Qiu et al. 2004) and polyacrylamide gels (Guschin et al. 2010) and either is recommended in the manufacturer's protocol (Transgenomic 2010), the estimates of gene editing will differ in accordance with the medium used to resolve the T7E1 cleavage products. This stated, the cost effectiveness of agarose meant this was used in provisional screens of guide RNA efficacy.

4.7.2 Assessing the Efficacy of In Silico Guide Designs

Continued examination of guide efficacy was undertaken in HEK293T.Cas9, in which 4µg, 2µg and 1µg doses of U6 plasmids encoding guide RNA 1, 2, 3 or 4 were transiently transfected. This cell-line constitutively expresses Cas9, thereby removing any variations in Cas9 expression that could occur on transient transfection of Px165. Notably, this investigation demonstrated that guide1 produced higher mean gene-modification relative to all other in silico guide designs; this was deemed statistically significant at the 95% confidence interval by a two way ANOVA. This could also be seen visually on inspection of the gel images where mean gene modification in the range of 10-14% was observed for guide 1, 1-4% for guide 4 and 2% or less for guides 2 and 3.

The finding that guide1 produces the best gene-modification is consistent with in silico predictions produced by CRISPOR (Haeussler et al. 2016). Notably, scores assigned to this guide based upon specificity (Hsu et al. 2013b) and efficacy (Doench et al. 2014) algorithms, which were developed after the initial guide designs of this study, were determined to be 69 and 54 respectively. These scores are both above the threshold score of 50, which is used to assess whether a guide is favourable for development. In addition, the increase in efficacy associated with this guide could be due it harbouring sequence determinant deemed beneficial to introduce DSBs at a locus, including: centrally positioned adenine nucleotides, a cytosine in position 20 and an adenine in position 1 (Doench et al. 2014). Note all numbering is PAM distal to proximal.

4.7.3 Px458-Guide 1 Efficacy

Guide 1 was selected for future experiments based on its higher efficacy relative to other in silico guide designs and examined using an all in one px458 plasmid backbone. Transient transfection of Px458- guide 1, demonstrated that gene-editing could be achieved in both

HEK293T and Δ 45-52 immortalised DMD patient myoblasts using an all-in-one plasmid based approach.

Importantly, Ran et al 2013 highlighted that the GFP tethered to the Cas9 does not affect the propensity of the Cas9 nuclease to cleave DNA. In consideration of this it is likely that the mean gene modification estimated at 4.0% and 7.4% in 2 μ g and 4 μ g transient transfections of Px458-guide 1 transfected HEK293T, is likely under reported. This is based upon the observation that co-transfection of 2 μ g U6-guide 1 and Px165 to HEK293T culture produced estimated gene modification of 27.1 % when resolved on 10% TBE polyacrylamide gel. Notably, this was a single reading established using T7E1 assay; nonetheless this disparity in estimation was unanticipated. Further examination of gene modification resulting from U6-guide 1 and Px165 and Px458-guide 1 transfections would be beneficial to assess the most robust system for introducing DSBs at the DMD intron 1 locus.

In exploration of determining robust estimates of gene modification, three methods were used to assess amplicons from the Px458 transient transfections in HEK293T and Immortalised Δ 45-52 Patient myoblasts. These included: T7E1 digest products resolved either on a 10% TBE polyacrylamide gel, or the Agilent bioanalyser and sequencing analysis of treated amplicons using TIDE analysis. Interestingly, the gene modification estimated by the 3 methods showed similar trends across the two cell lines. In both instances, the Agilent bioanalyzer showed an apparent but non-significant reduction in gene-modification estimates relative to that observed with 10% TBE polyacrylamide gels as established by 2 way ANOVA. Notably, in the case of HEK293T this resulted in almost a 2 fold reduction of estimates from 4% - 2.3% and from 7.4% - 4.8%. In contrast the reduction was subtle in the case of patient myoblasts from 1.1% - 1% and 5.2% - 4.4%.

Finally, both samples illustrated an increase in apparent gene-modification estimated by TIDE analysis. In the instance of HEK293T the mean gene modification was increased more than 2 fold in the 2 μ g sample from 4% to 10.9% deemed statistically significant and a more subtle difference in the 4 μ g sample from 7.4% to 10.3% deemed non-significant at the 95% confidence interval by two way ANOVA was observed. In the case of patient myoblasts the effect was more striking with an increase from 1.1 to 6.8% and 5.2% to 12.3% in 2 μ g and 4 μ g treated samples, both of which were deemed statistically significant by two way ANOVA at the 95% confidence level respectively.

The results obtained serve to highlight that the methods used to determine gene-modification can influence the estimates obtained from a single sample. TIDE analysis may have illustrated an apparent increase in gene modification on the basis that a multitude of

INDEL types are detected, including SNPs, which may not be susceptible to cleavage on the T7E1 assay (Brinkman et al. 2014; Vouillot et al. 2015).

An important implication of this finding is that where possible multiple methods to screen guide RNA efficacy should be undertaken. This is suggested on the basis of increasing evidence that InDel patterns following gene-modification is a non-random process. As evidenced by an extensive study that assessed 223 CRISPR guide RNAs and noted from sequencing that InDel patterns were maintained across multiple biological replicates (van Overbeek et al. 2016). Thus high efficacy guides that produce a high proportion of single nucleotide polymorphisms, could be excluded from further assessment if gene modification is only estimated by T7E1, which is unable to detect such subtle sequence alterations (Vouillot et al. 2015).

Interestingly, a recent review by Tycko et al highlighted the requirement to standardise InDel detection methods at off-target sites by using stringent Next generation sequencing, a method with high sensitivity estimated to detect 0.1% of InDels (Tycko et al. 2016; Kleinstiver et al. 2016). Arguably, such a standardisation of gene-modification detection methods would also be helpful to characterise on-target gene modifications. In this manner, guide RNAs short-listed for development could be assessed using the existing accessible platforms, prior to being subject to defined delivery conditions standardised next-generation sequencing. This would provide a system by which guide RNAs between studies could potentially be bench-marked for activity.

4.7.4 Alternative Cas9 Delivery Approaches

In an attempt to improve upon the plasmid transient transfection delivery strategy in the context of patient myoblasts, alternative formats of Cas9 nuclease were examined, particularly as a transition in the field away from plasmid based gene-editing has become apparent (Zuris et al. 2014; Lin et al. 2014; S. Kim et al. 2014; X. Liang et al. 2015; Yu et al. 2016). Functional assessment of guide 1 RNA, produced by in vitro transcription, delivered to HEK293T.Cas9 cells in a six well plate highlighted the presence of DSB. Interestingly the quantification of mean gene modification in this instance asserted to be 5.5% was half the activity observed with DNA. Strikingly, however, when guide 1 RNA and Cas9 mRNA were co-delivered to patient cells in a 12 well plate the gene-modification estimates attained were 17.4 and 10.4% for a lower and higher doses of mRNA delivered. This modification estimate is 3 times that achieved on 10% TBE gels with a plasmid based transient transfection. One reason for this disparity could be that the guide RNA and Cas9 mRNA was delivered in a six well plate in one instance and a 12 well plate in another. This could serve to validate a recent argument that toxicity allowing, gene-editing reagents should be delivered to fewer cells at a 40-70% confluency (Yu et al.

2016). The premise being that this increases the amount of reagent relative to the pool of cells and functions much like a dose escalation.

Unfortunately reproducing this result has proved challenging, which is likely indicative of variability within the mRNA quality. This serves to highlight the requirement for more stringent assessment of RNA quality; this could be achieved by resolving the RNA on a 3% agarose gel or on an Agilent Bioanalyser RNA microchip. Furthermore recent work in the field of gene editing has highlighted that extension of the polyA tail serves to increase mRNA stability (Yoshimi et al. 2016). Moreover, in the course of this work a Cas9 GFP-tethered mRNA was produced (Yoshimi et al. 2016). These constructs could definitely facilitate the development of more robust RNA based gene-editing strategies in the context of the patient cells.

The ribonucleotide protein format of the nuclease delivery was the last to be investigated. Initially, Cas9-GFP kindly provided by Dr. DeCian and Dr. Concordet, was delivered alone to assert identify a range of Cas9 protein doses that could be delivered to HEK293T cell culture with observable expression. This was an important milestone as a range of doses of Cas9 protein from 1.5 μ g - 45 μ g were reported in literature, to be delivered alongside guide RNA (X. Liang et al. 2015; S. Kim et al. 2014; Lin et al. 2014). Using Lipofectamine mRNA messenger max 2.5 μ g and 5.0 μ g Cas9 protein appeared to be well tolerated in HEK293T cells. Whereas in comparison with untreated controls (not shown), the amount of transfection reagent required for the delivery of 10 μ g and 20 μ g Cas9 protein appeared to cause toxicity. The reason for this toxicity being observed in this instance but not reported in other investigations could be due to the route of delivery. Cas9 protein has been largely delivered by electroporation in gene editing studies (S. Kim et al. 2014; Lin et al. 2014; X. Liang et al. 2015).

Notably, the field of cationic protein delivery continues to evolve, thus when a newer delivery reagent with reduced toxicity Lipofectamine CRISPR Max was released, this was used to trial lower doses of Cas9 protein with in vitro transcribed guide 1 RNA (Yu et al. 2016; Zuris et al. 2014). When 2:1 and 4:1, doses of Cas9 protein: transcribed guide RNA were trialed mean gene modification was estimated to be 10.2% and 7.8% at the two ratio's trialed when resolved on an agarose gel. The highest estimate is comparable to estimates obtained for guide 1 plasmid delivered into HEK293T.Cas9 cells and dual transfection of 2 μ g U6 guide 1 and 2 μ g Px165 experiments, which provided gene modification estimates of 11.1% and 12.9% respectively. Furthermore, when the resultant intron 1 amplicon was subjected to TIDE analysis this showed nearly threefold increase in gene modification estimates, with 31.5% and 25.7% being estimated, relative to 11% estimated for Px458-guide 1. Once again the different gene modification estimates serve to reinforce findings that multiple gene-editing assays

should be used to estimate mean gene-modification. Intriguingly, the InDel spectrum, this being the number of insertions and deletions at a DSB point differed between the Px458-guide 1 and RNP analysis. This was unexpected as CRISPR guides are indicated to cause reproducible InDel patterns at a DSB point (van Overbeek et al. 2016). Interestingly, the increased number of larger deletions 3+ nucleotides observed with the Cas9 protein, could be attributed to either the Cas9 protein being more efficacious or due to the platform of delivery. This being protein as opposed to plasmid, further investigation is warranted.

The work undertaken serves to expand the gene-editing repertoire of the lab and may prove beneficial to achieving a homology directed repair outcome in the lab. One constraint of this methodology however, is that it relies on efficient delivery of the MD1 repair template. However, recent work has demonstrated that Cas9 protein can be delivered alongside an Adenoviral Associated Viral vector. This virus with the packaging capacity of 4.5kb could be used to deliver the repair template whilst the Cas9 protein is able to introduce DSBs in cell culture. Furthermore, a dual strategy of this nature may also be supported by an alternative approach to introduce a transgene relating to NHEJ-mediated knock-in (Suzuki et al. 2016).

4.7.5 Strategies to Enhance Gene-modification:

In addition to the in-silico analysis for improved guide design (Chapter 3) and alternative delivery format of Cas9 nuclease, gene-modification at the human *DMD* intron 1 locus could be enhanced via a number of strategies, including: i) Improved single guide RNA architecture and composition to make it resilient to RNase attack and subsequent degradation ii) Modified culturing conditions relating to the temperature the nuclease is incubated and iii) Examination of alternative delivery strategies, with electroporation appearing in literature to be the most efficient format.

The original study fusing the crRNA and tracrRNA performed by Jinek et al. (2012), served to identify the minimum sequence architecture required to mediate a DSB. This was deemed to be 42 nucleotides in length significantly shorter than the native dual RNA guide, comprised of crRNA and tracrRNA observed in the endogenous CRISPR-Cas9 system. However, since this numerous alterations to the single guide RNA architecture have been undertaken; predominantly these have served to examine the importance of duplex length and of the contiguous poly T stretch immediately downstream of the crRNA region, which serves endogenously to modulate the RNA III polymerase activity (Nielsen et al. 2013). Interestingly, a longer duplex than that used by Jinek et al. (2012) was deemed optimal for mediating gene modification (Hsu et al. 2013b). Thus an sgRNA of approximately 85 nucleotides became the mainstay in the gene editing field, with extension beyond this deemed detrimental by Hsu et

al. (2013). In addition to this visualisation of telomeres undertaken using a dCas9, a Cas9 with ablated catalytic activity and an extended single guide RNA architecture harbouring mutations in the poly T region demonstrated increased target binding relative to the wild-type (Chen et al. 2013). Further assessment of these parameters in the context of knock-out generation, demonstrated that an increase of 5nt and the mutation of the 4th thymine to a cytosine adjacent to the cRNA, resulting in improved gene-editing outcome. Specifically, an increase guide efficacy when tested alone that was significant in a third of cases and an 10 fold increase when combined in knock-out generation with a multiplex NHEJ-mediated deletion strategy (Dang et al. 2015). These sequence modifications have subsequently been adhered to in the gene-editing field. An implication of these investigations is that alternative single guide RNA scaffolds have varying propensity to introduce DSBs, further investigation and optimisation of the structure could serve to improve guide efficacy and subsequently the exploitation of NHEJ and HDR DNA repair pathways.

In addition to an modified architecture of a single guide RNA, recent work has highlighted that modification of the encoding RNA could render it more resilient to RNase activity (using a phosphorothiorate backbone) and facilitate it attaining a conformation more favourable to RNA-DNA Watson Crick binding (the addition of 2' O-Methyl or 2-Fluoro groups) (Rahdar et al. 2015). Whilst investigations thus far indicate that the gene-editing achieved is sub-optimal relative to the single guide RNA in a plasmid delivery format, this may be attributable to the expression profile of the RNA delivered relative to plasmid. Furthermore, the increasing use of Cas9 nuclease highlights the requirement for more work in this area to be undertaken.

It has been demonstrated previously that gene-editing with customisable nucleases, may be influenced by the temperature. Both ZFNs and TALENs have been demonstrated to show a higher level of efficacy in orchestrating a DSB, when cell cultures are placed under transiently hypothermic 30°C conditions (Doyon et al. 2010; Miller et al. 2011), an effect also demonstrated to extend to *in vivo* modification of the porcine and rat genes, upon TALEN mRNA injection into zygotes (Carlson et al. 2012; Remy et al. 2014). Interestingly however, a recent paper has demonstrated that the Cas9 nuclease responds to temperature in a distinct manner from its predecessors, with increased gene editing efficacy at hyperthermic (39°C) as opposed to hypothermic (30°C) conditions (Xiang et al. 2017). This stated the effect was deemed to be subtle and not significant in many cell lines; nonetheless it may warrant investigation to improve gene modification. Finally, it was theorised that the increased gene editing observed was attributable to an increase in the expression of the guide RNA delivered and the activity of the Cas9 nuclease. In light of this it could be speculated that increasing the

amount of guide delivered, in addition to an increase in the temperature may serve to improve gene-editing outcomes.

4.8 Conclusion:

In conclusion this chapter assessed the efficacy of in silico guide designs targeted to the DMD intron 1 locus; it was found that guide1 produced DSBs with the highest efficacy. Thus this guide RNA will be used in future experiments to affect HDR-mediated integration of the MD1ΔEx1 transgene downstream of the full length dystrophin isoform promoters.

In addition, plasmid based systems were explored in the introduction of DSBs to the DMD Intron 1 locus, including U6-guide plasmid alone in HEK293T.Cas9 cells and the U6-guide and Px165 Cas9 protein expressing plasmid co-transfection system or the Px458 all-in-one plasmid in HEK293T. The Px458 represents a particularly attractive for system for development as the GFP linked to Cas9 by 2A, can also be used to provide estimates of transfection efficiency.

Finally, alternative systems of in vitro transcribed guide and Cas9 delivery and RNP formats were explored. These systems were also shown to have the propensity to introduce DSBs at the DMD intron 1 locus. Further research is required to establish these systems further. Nonetheless this represents an expansion to the gene-editing repertoire.

Chapter 5 Dystrophin Transgene Sequence Optimisation and Putative Design of cDNA-based Repair Templates:

5.1 Introduction:

As described previously, a CRISPR guide RNA, denoted as guide 1 was identified that directed high levels of DSB introduction at the selected human *DMD* Intron 1 site. Focus of the research then shifted to the design of a suitable exogenous repair template which would potentially allow insertion of recombinant cDNA sequences by HDR to mediate expression of dystrophin or microdystrophin. The exogenous repair template designed, will function as a cDNA donor during G2 and S phases of the cell cycle (Heyer et al. 2010). In the context of the therapeutic strategy under investigation, this repair template will serve to deliver a cDNA transgene encoding dystrophin variants from exon 2 onwards, utilising the full length endogenous promoters and *DMD* exon 1 to reconstitute the 5' end of the dystrophin cDNA and, by extension the resultant protein. The primary emphasis will be to introduce microdystrophin cDNA, a functional albeit truncated dystrophin transgene, bearing deletion of dystrophin protein domains from spectrin-like repeat 4-23 and truncation of the C-terminal (MD1: $\Delta R4-23-\Delta CT$) (Foster et al. 2008; Athanasopoulos et al. 2011). This strategy could also be applied with other dystrophin cDNAs but the MD1-based approach was initially investigated since its smaller size was deemed more favourable to integration via the HDR pathway. However, repair template plasmid design was developed with the intention to utilise this strategy, time permitting, to introduce the more extended dystrophin variants which may have functional advantages. These extended dystrophins variants included minidystrophin bearing an internal deletion of exons 45-55, a variant identified to result in mild Becker muscular dystrophy (BMD), which could be used to slow disease progression (Taglia et al. 2015) and full length dystrophin (Romero et al. 2004). Hence, the exogenous repair template was constructed to enable cloning and potentially the delivery of each of the dystrophin transgenes. Upon successful integration into intron 1, it is hypothesised that the endogenous full length promoters will transcribe exon 1 appropriately and this will splice to exon 2 of the MD1 cDNA and drive functional dystrophin protein expression, with spatial and temporal patterns consistent with that observed physiologically.

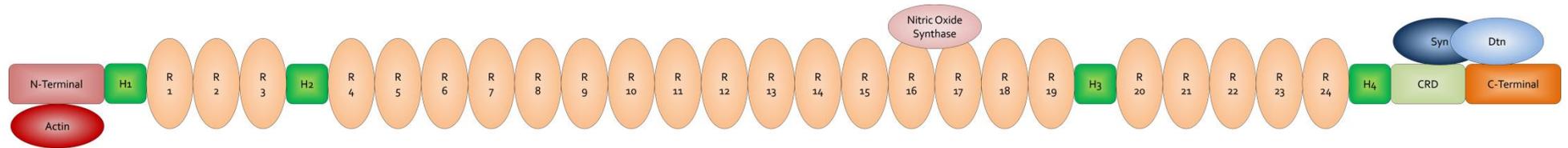
5.1.1 Minidystrophin and Microdystrophin Transgenes

Dystrophin has a modular architecture, comprising a N-terminal domain binding to cytoskeletal F-actin, a centrally located rod domain comprised of 24 coiled-coil spectrin-like repeats, interrupted by 4 proline rich hinges into 3 sub-regions, prior to a cysteine-rich region and C-terminal domains which associate with dystroglycan, syntrophin and dystrobrevin, and

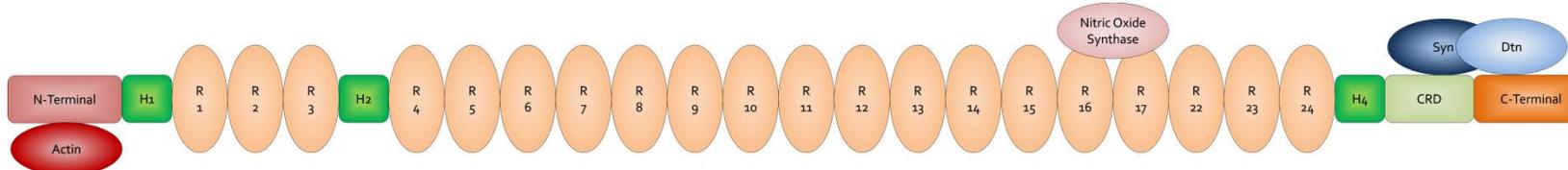
other crucial components for the formation of the DAPC (Harper et al. 2002; Allen et al. 2016). In addition, it should also be noted that large in-frame deletions spanning the central rod region of dystrophin protein have been reported to result in mild symptoms. As exemplified by the identification of BMD patient who patient lived into his 60s, despite of having a large deletion of central rod domain, equating to 46% of the dystrophin gene (England et al. 1990). A further recent example is provided by BMD deletions spanning exons 45-55 of the DMD gene (Taglia et al. 2015; Ousterout et al. 2015), which gives rise to a relatively benign BMD phenotype. These observations in combination prompted the development of shorter transgenes encoding the minimal components of dystrophin protein necessary to enable structural linkage between the intercellular cytoskeleton and the extracellular matrix. These smaller transgenes have the advantage that they can be delivered through packaging into lentiviral and AAV vectors, whose capacity is exceeded in the case of the native full-length dystrophin cDNA.

Microdystrophins represent a compromise between the requirements to shorten the dystrophin cDNA, whilst retaining dystrophin protein functionality. It has been demonstrated at a biochemical level that microdystrophins enable the restoration of the DAPC (Bostick et al. 2008; Koo et al. 2011), and at a physiological level they serve to reduce dystrophic pathology in both mouse and canine models of DMD (Koo et al. 2011; Le Guiner et al. 2017). Numerous microdystrophins have been designed by multiple groups (Wang et al. 2000; Harper et al. 2002; Foster et al. 2008; Banks et al. 2010; Athanasopoulos et al. 2011; Reza et al. 2016). One well characterised microdystrophin is the $\Delta R4-23-\Delta CT$ variant in which spectrin-like repeats 4 - 23 and the C-terminal region are removed. It has been demonstrated the removal of the C-terminal of dystrophin is not detrimental to its structural function on the basis that syntrophins and dystrobrevins associate with the cysteine rich region of dystrophin protein (Yue et al. 2006). DNA constructs corresponding to microdystrophin (MD1, $\Delta R4-23-\Delta CT$), a minidystrophin harbouring deletion of exons 45-55 (minidystrophin $\Delta 45-55$), and the full length dystrophin itself were all considered as transgenes of interest for integration experiments of this investigation (Figure 5-1). Continued research has highlighted that several signalling molecules associate with the spectrin-like repeats of the dystrophin protein contributing to its function. Neuronal NOS associates with repeats 16-17 and is implicated in modulating the blood flow to muscles (Lai et al. 2009; Meng et al. 2016), and Par1b associates with repeats 6 and 7 and plays a role asymmetric division of satellite cells (Dumont et al. 2015). This has led to additional configurations of microdystrophin being designed and tested (Reza et al. 2016). However, in the current study we have focussed in particular on the sequence-

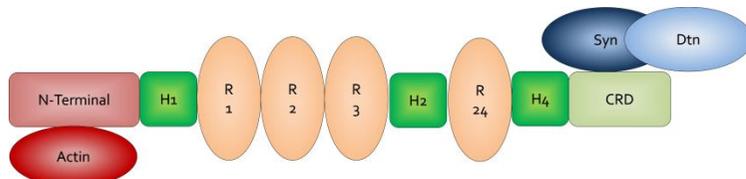
optimised microdystrophin $\Delta R4-23-\Delta CT$ transgene termed MD1 (Foster et al. 2008; Le Guiner et al. 2017) which has been shown to be highly functional in mouse and dog models of DMD.



Full Length Dystrophin Protein 427kDA



Δ45-55 Dystrophin Protein 382kDA



MD1 Protein 137kDA

Figure 5-1 Schematic of the full length, Quasi Δ45-55 Dystrophin and Microdystrophin (R4-23 ΔCT) protein structure.

N-terminal shown in pink, proline rich hinge regions (H₁-H₄) shown as dark green squares, 24 spectrin-like repeats orange ovals (R₁-R₂₄), Cysteine rich region (CRD) light green rectangle and the C-terminal orange rectangle. In addition to the dystrophin protein structure, signalling proteins that associate with dystrophin are also shown including actin (red oval), nitric oxide synthase (pink oval) and syntrophin (Syn) and dystroglycan (DTN) (dark blue and light blue ovals respectively). These are the resultant protein structures that will be formed from the potential exogenous repair templates.

5.1.2 Sequence Optimisation of Dystrophin Transgenes

Sequence optimisation broadly refers to the direct modification of a native DNA sequence, addressing parameters such as codon usage, DNA motifs, GC content and sequence repetitions, to increase resultant mRNA levels and stability and protein translation. This by extension serves to increase protein expression (Raab et al. 2010). This first became of interest in molecular science as generation of native and recombinant human proteins was limiting to both basic research and therapeutic development. Early attempts to increase the production of human proteins focused upon expression in *E.Coli*, thereby using the bacterial transcriptional machinery and quick doubling time to improve protein yield (Makrides 1996). Whilst the fast growth and easy culturing of this bacterium proved favourable to protein production and accessibility, the resultant protein exhibited some fundamental differences from the anticipated native protein. Notably, this included: aberrant folding of proteins, a lack of post translational modifications and evidence of protein aggregation. Furthermore, some constructs could not be expressed in *E.Coli* likely due to toxicity and/or size limitations (Fath et al. 2011). The alternative approach of autologous expression of the protein, this being expression in the native host cell culture, provided the expected protein but at sub-optimal levels (Fath et al. 2011).

To overcome these limitations, sequence optimisation of DNA aimed to increase gene expression without the introduction of detrimental alterations to the downstream products was performed. Several strategies were employed in pursuit of this goal and it should be noted that sequence optimisation is an on-going area of research that is still evolving. Two early strategies employed were focused upon DNA-motifs and back-translation. The first strategy was founded on the premise that certain DNA motifs regulate gene expression, some pertinent examples include: UpA critical for mRNA stability, CpG content correlated to de novo transcription and AU rich elements related to mRNA stability (Fath et al. 2011). Once this was understood conventional genetic engineering approaches were employed to increase the occurrence of these motifs. In parallel, it was shown that codon composition, namely the use of codons favoured from a degenerate pool by the particular organism, could increase protein production by an order of magnitude (Graf et al. 2000; Graf et al. 2004). Hence the latter approach of back-translation was soon employed, causing a shift from genetic engineering to more de novo synthesis (Raab et al. 2010). The former being modification of the nature of the genetic sequence and the latter being the commercial synthesis of an optimised product.

The back-translation strategy is based upon the degeneracy of the genetic code. Degeneracy refers to the fact that a singular amino acid can be coded for by more than one codon. The multiple codon combinations that could code for a singular amino acid are called synonymous

codons. However, the frequency of codon usage varies in accordance with organism and even tissue type (Plotkin et al. 2004). The central doctrine of back translation is to start the optimisation with an amino acid sequence and construct a DNA sequence comprised of codons most frequently used by the organism in which expression is desired. This underlying codon preference is posited to be due to variable tRNA pools within the given organism and/or tissue type, with larger pools of tRNA correlating to greater levels of translation (Cannarozzi et al. 2010; Fath et al. 2011). Although superficially this would appear to provide an optimal solution for expression, there are potential drawbacks that for this approach. Including the inadvertent introduction of undesired features such as restriction sites, splice sites and repetitive sequences (Raab et al. 2010).

Thus the 'optimal sequence' for a particular transcript is not obtained by using the back-translation approach in isolation, but rather by compromising between optimal expression and the exclusion of certain undesirable sequence elements (Raab et al. 2010). Upon this realisation algorithms were employed to provide the optimal sequence solution by addressing these multiple parameters. Whilst it was universally recognised that the gold-standard to this approach would be to generate all of the possible combinations and test them, this is simply not a viable option as a great many combinations are possible. Raab uses the case in point that a 100 amino acid protein would have 10^{47} possible combinations (Raab et al. 2010). Thus 'optimal' sequence results currently can only be calculated within the remit of the parameters of a given algorithm. The Monte Carlo and GeneOptimizer Sliding window algorithms are two prominent algorithms (Gallagher et al. 1991; Raab et al. 2010). Although, the mathematical formulae, assertions and derivations are beyond the remit of this investigation, the conceptual elements of the algorithms harbour importance to sequence optimisation and thus will be briefly examined.

The Monte Carlo algorithm uses a short sequence window and performs an exhaustive screen across a limited number of codons. The findings within this local region are then applied to the whole gene (Gallagher et al. 1991). This has been found in many instances to provide a protein where the local region within the window appears optimised, but globally a sub-optimal result is achieved. The current gold-standard in sequence optimisation is the GeneOptimizer algorithm. Since an analysis of all the possible combinations is unachievable, this algorithm reduces the sequence space to a "variation window" of a few codons. An exhaustive search is performed in which all iterations of the sequence are trialled and analysed against desired parameters. If one of the iterations increases the overall quality of the protein optimisation then it is retained if not, it is disregarded. In this manner a candidate sequence is built. Upon achieving an optimal solution within the "variation window" the sequence is fixed and the

window shifts to the next region of the sequence in a unilateral manner, analogous to how it is transcribed in the cell (Raab et al. 2010). This continues along the entire length of the gene and until an optimal solution is achieved. A singular multi-gene trial has been used to examine the GeneOptimizer algorithm extending to 50 genes (Fath et al. 2011), the results of which were striking with 96% of the sequence optimised constructs performing equal to or better than native counterparts. Moreover, 86% of the genes expressing in this subgroup showed a two-fold or greater increase in expression relative to the native counterpart (Fath et al. 2011).

In terms of this investigation, the sequence optimised transgenes that are being characterised, include full-length dystrophin, $\Delta 45-55$ minidystrophin and MD1 microdystrophin which have previously been optimised using GeneOptimizer algorithm, the MD1 construct referring to (R4-23 Δ CT) as previously described in the literature (Foster et al. 2008; Le Guiner et al. 2017). The sequence optimisation was undertaken using input parameters to optimise human codon usage and avoid detrimental cis-acting motifs including internal TATA boxes, chi-sites and ribosomal entry, AT- or GC-rich sequence stretches, RNA instability motifs, repeat sequences and RNA secondary structures, and cryptic splice donor and acceptor sites in higher eukaryotes. This sequence optimisation and de novo synthesis was undertaken by GeneArt prior to the initiation of this project. These constructs were subsequently examined to assess whether sequence optimisation increased expression levels of the corresponding recombinant protein. Optimised dystrophin variants hitherto will be referred to as Opt-HuDys (SO), Opt- $\Delta 45-55$ HuDys (SO) and Opt-HuMD1 (SO). As described the GeneOptimizer algorithm uses a sliding window and accounts for optimisation in both a local and global manner. Thus owing to the varying lengths of all three constructs, subtle differences in sequence optimisation may be present.

5.1.3 Chapter Aims:

This chapter aimed to assess the expression of recombinant dystrophin protein resulting from both native and sequence optimised cDNA encoding microdystrophin, $\Delta 45-55$ minidystrophin and full length dystrophin. Subsequently, the design and molecular cloning of a generic exogenous repair template backbone was undertaken. This generic exogenous repair template was designed to unique endonuclease restriction sites to enable the sub-cloning and subsequent delivery of the 3 dystrophin cDNA variants. Finally, the microdystrophin variant was sub-cloned into the generic repair template plasmid.

5.2 Results:

5.2.1 Assessment of Dystrophin Expression from Native and Optimised Fused Dystrophin-eGFP Expression Constructs.

Initially, a microscopic indication of whether sequence optimisation improves transcription and subsequently synthesis of recombinant dystrophin protein was sought. Plasmids containing native and optimised $\Delta 45-55$ or full length dystrophin cDNAs directly fused to eGFP were transiently transfected into HEK293T cells. Expression was driven by the cytomegalovirus (CMV) promoter, a strong viral promoter to ensure adequate levels of protein expression. The fusion of eGFP to dystrophin enables fluorescence microscopy to be used as an indicator of dystrophin expression, particularly as the stoichiometry of eGFP to dystrophin is equivalent in the resultant fusion protein.

Full length native and optimised dystrophin-eGFP fusion plasmids, Nat-hDys-eGFP and Opt-HuDys-eGFP respectively, were transfected at a 4 μ g dose to HEK293T cells using a 5:1 ratio of Viafect transfection reagent (μ l): DNA (μ g). Cultures were then subject to microscopy imaging at 24, 48 and 72 hours post transfection. There was a dramatic difference in eGFP expression observed between cultures transfected with native and optimised plasmids across all time points examined. GFP positive cells were observed in all cultures indicating successful transfection with both native and optimised constructs, and in all samples the intensity of GFP fluorescence increased from 24-72 hours irrespective of optimisation, likely attributable to the accumulation of dystrophin-eGFP transcript and protein (Figure 5-2). The dramatic difference observed in eGFP expression between native and optimised constructs is speculated to be due to the sequence optimisation of the plasmid. The proposed implication being that the optimised construct enhances mRNA levels and/ or translational efficiency, and subsequently dystrophin-eGFP protein synthesis.

Native and optimised $\Delta 45-55$ minidystrophin-eGFP fusion constructs, Nat- $\Delta 45-55$ hDys-eGFP and Opt- $\Delta 45-55$ HuDys-eGFP (SO) were next examined in a similar way. The optimised $\Delta 45-55$ minidystrophin-eGFP fusion construct also showed a dramatic increase in eGFP expression relative to the native construct (Figure 5-3). Thus the overall trend was retained across both studies.

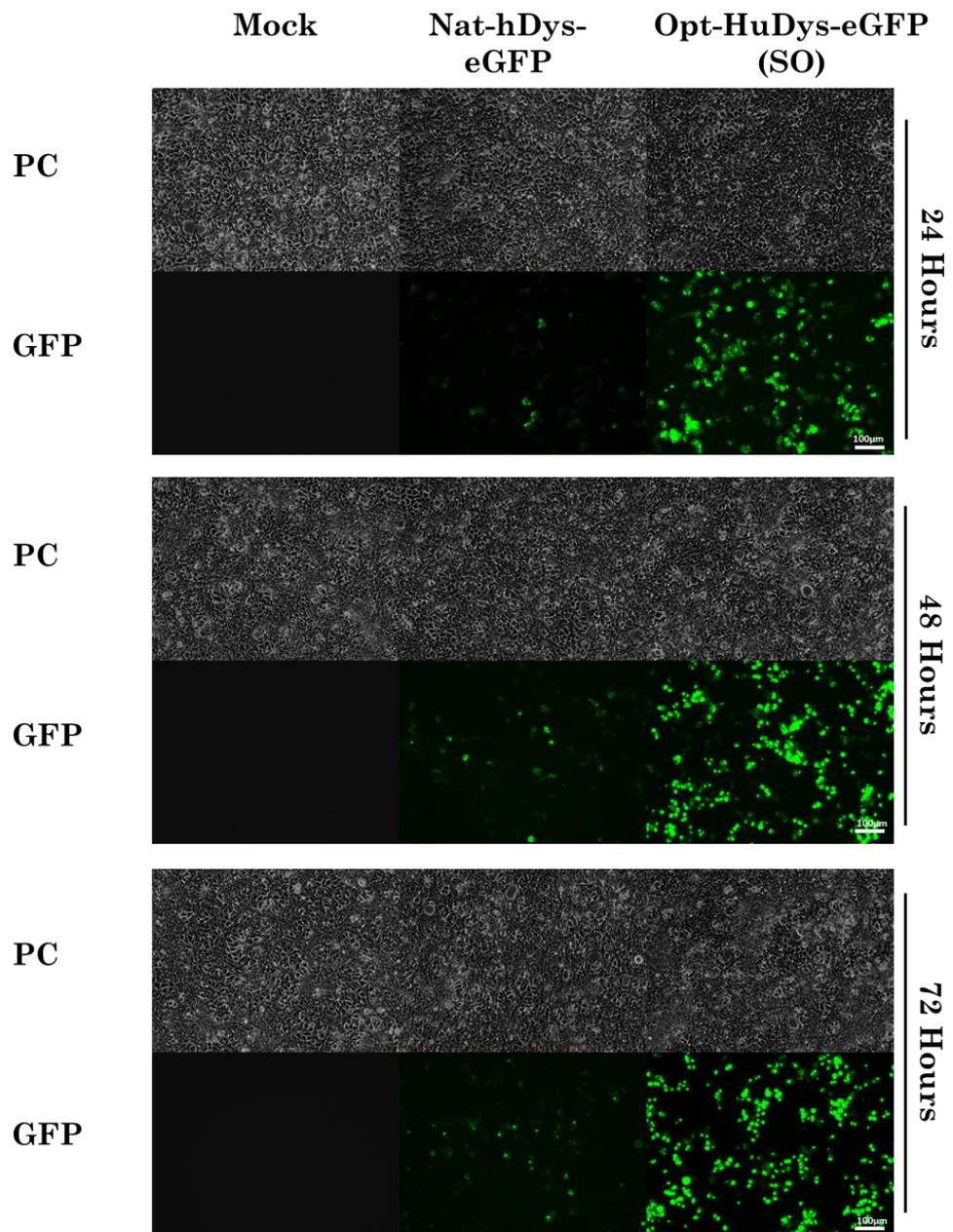


Figure 5-2 Visual demonstration that sequence optimisation of cDNA improves recombinant protein expression as demonstrated by using constructs encoding native and optimised, full-length dystrophin fused to eGFP:

HEK293T cells transfected with native and sequence optimised full length dystrophin-eGFP constructs, Nat-hDys-eGFP and Opt-HuDys-eGFP (SO) respectively. Transient transfections were undertaken at 4 μ g with a 5:1 viafect reagent: DNA ratio. Subsequent eGFP expression was imaged at sequential time points of 24, 48 and 72 hours with the Zeiss microscope at 100x magnification. Phase contrast (PC) images were undertaken at 30ms exposure and GFP images at 800ms exposure. A noticeable accumulation of eGFP expression can be seen in sequential time points for both constructs; in addition a difference in fluorescence intensity is seen between native and optimised dystrophin encoding constructs.

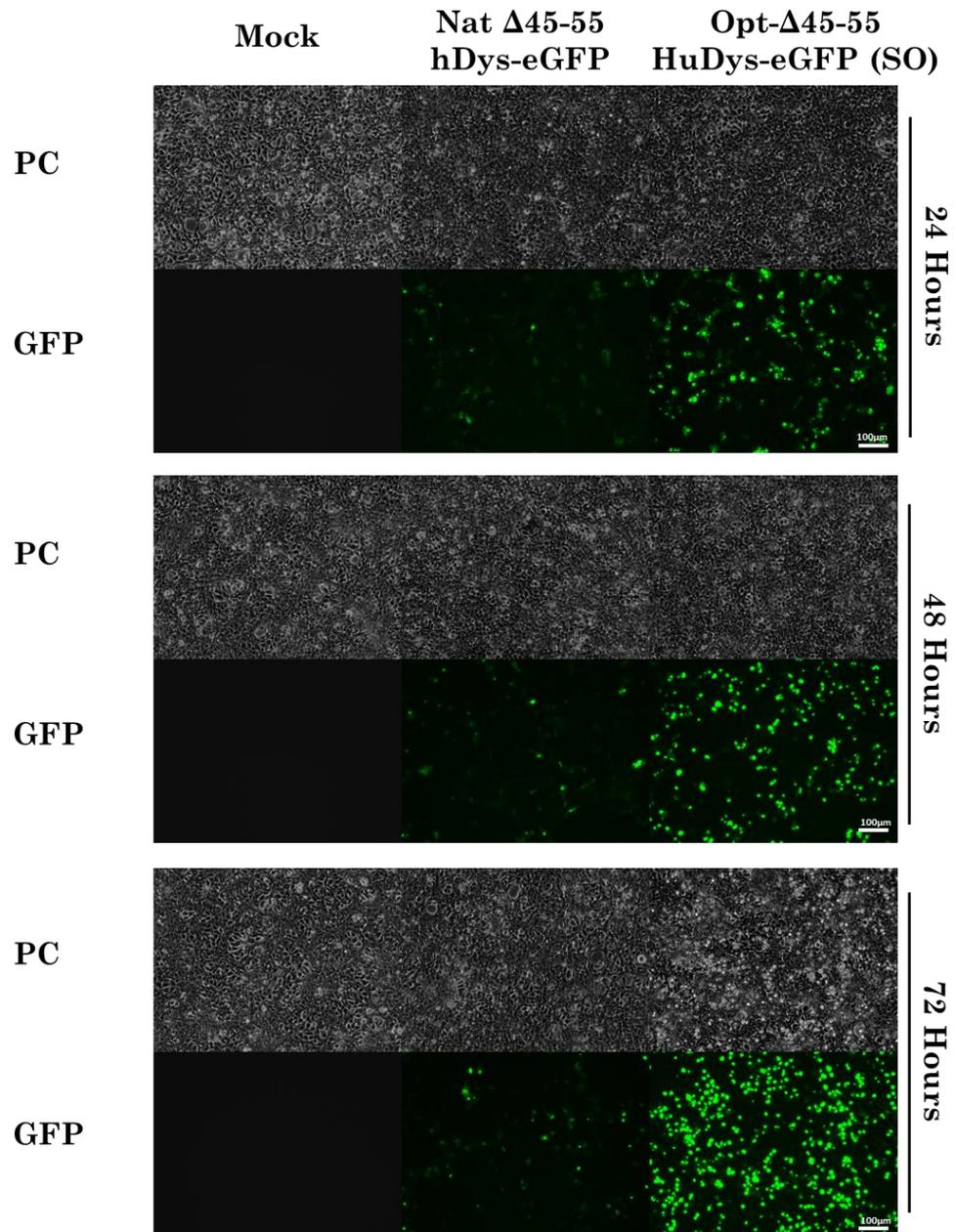


Figure 5-3 Visual demonstration that sequence optimisation of cDNA improves recombinant protein expression as demonstrated by using constructs encoding native and optimised Quasi $\Delta 45-55$ dystrophin fused to eGFP

HEK293T cells transfected with native and sequence optimised quasi $\Delta 45-55$ dystrophin-eGFP constructs, Nat- $\Delta 45-55$ hDys-eGFP and Opt- $\Delta 45-55$ HuDys-eGFP (SO) respectively. Transient transfections were undertaken at $4\mu\text{g}$ with a 5:1 viafect reagent: DNA ratio. Subsequent GFP expression was imaged at sequential time points of 24, 48 and 72 hours with the Zeiss microscope at 100x magnification. Phase contrast (PC) images were undertaken at 30ms exposure and GFP images at 800ms exposure. A noticeable accumulation of eGFP can be seen in sequential time points for both constructs; in addition a difference in fluorescence intensity is seen between native and optimised dystrophin encoding constructs.

5.2.2 Quantification of the Difference in Expression of Native and Optimised Dystrophin Constructs Driven by a CMV Promoter:

The increased fluorescent intensity observed prompted direct assessments by Western blot of dystrophin protein expression for both CMV-driven Nat-hDys and Opt-HuDys (SO) constructs. In this case constructs without the eGFP fusion were used. Transient transfections of plasmids were repeated in HEK293T as previously described; cultures were incubated for 72 hours and total protein lysate harvested for Western blot analysis. The protein samples were subsequently quantified and 50µg was examined with two dystrophin antibodies: the 6C5 antibody, which binds to the C-terminus of the dystrophin, and the MannEx1011c, which binds to a dystrophin epitope encoded by exons 10 and 11. Blots were subsequently scanned and dystrophin expression was quantified and normalised against an alpha-tubulin loading control, using the Odyssey Licor system, as described in the Materials and Methods.

Both the Nat-hDys and Opt-HuDys (SO) constructs expressed a 427kDa protein as determined by a HiMark Nupage Ladder, consistent with full length dystrophin. This was detected with both the 6C5 and the MannEx1011c antibodies. Congruent with the previous microscopic evaluations of eGFP fusions, the sequence optimised construct yielded a larger area band with increased intensity relative to the native construct. This trend was retained across 5 samples that were transiently transfected (Figure 5-4 and Figure 5-5).

For the purposes of protein detection secondary-antibodies labelled with near-infrared (NIR) fluorescent dyes, with emission peaks detected within the 800nm and 700nm channels were used for dystrophin and alpha-tubulin respectively. This enabled the normalisation of dystrophin expression to alpha tubulin, to provide dystrophin / alpha tubulin expression ratios for both native and optimised dystrophin constructs. The mean ratio for optimised dystrophin was then divided the mean ratio of native dystrophin to provide a fold difference in expression. This form of analysis indicated a striking 57-fold difference with 6C5 antibody and a 22-fold difference with the MannEx101c antibody. This was deemed statistically significant in both cases giving a $p < 0.001$ by a two tailed unpaired t-test.

Additionally, it should be emphasised that on examination of the western blot an alternative banding pattern between the two antibodies was observed. The 6C5 antibody appeared to produce a characteristic laddering pattern with the Opt-HuDys (SO) construct that was not observed in the case of the Nat-hDys construct. The additional banding seen had moderate intensity comparable to the full length band (Figure 5-1). In contrast the Mannex1011c antibody resulted in a singular sharp band for both Nat-hDys and Opt-HuDys (SO) constructs, with few additional bands at low intensity. In the case of the latter, this is likely result of

increasing the brightness to visualise the Nat-hDys bands. This could indicate that the 6C6, C-terminal antibody detects further dystrophin splice variants or indeed degradation products.

A parallel western blot series was extended to the Nat- Δ 45-55hDys and Opt- Δ 45-55HuDys (SO) constructs. These constructs yielded a 380kDa protein as determined by HiMark Nupage Ladder, consistent with the internal deletion of exons 45-55 present in this minidystrophin. Promisingly, this yielded similar results with an apparent increase in the area and intensity of the protein bands resolved in the sequence optimised condition (Figure 5-6 and Figure 5-7). In this experiment, no attempt was made to quantify the difference in expression resulting from native and sequence optimised constructs. This was not undertaken as the native Δ 45-55 Dystrophin band was barely visible relative to the optimised conditions, whilst the loading controls detected with alpha tubulin remained consistent. This suggested that the difference observed was significant, likely greater than that observed with full length dystrophin and would be over-estimated during quantification processes.

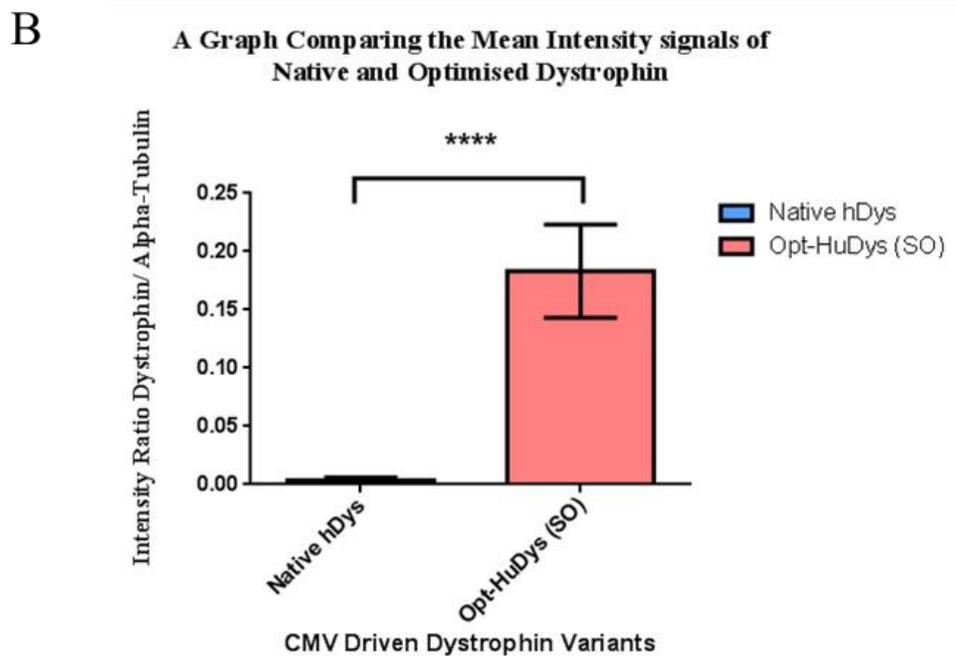
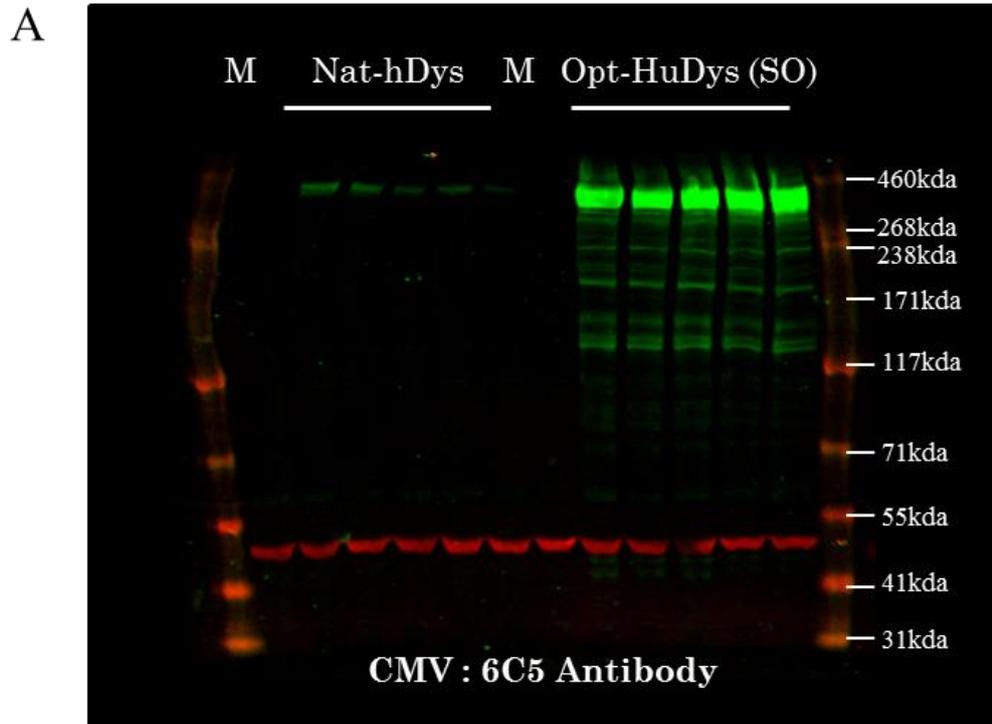


Figure 5-4 Optimisation of cDNA sequence increases recombinant dystrophin expression.

Plasmids expressing native (Nat-hDys) and sequence-optimised dystrophin (Opt-HuDys-SO) cDNAs from the CMV promoter were transfected into HEK293T cells (n=5). After 72h protein were harvested using RIPA buffer and processed for Western Blotting (A) 50µg total protein lysate was analysed with antibodies to dystrophin (6C5) and alpha-tubulin. Dystrophin bands were then quantified relative to the alpha-tubulin loading controls and mean intensity ratios plotted (B: mean±SEM: ****8 p<0.001 (unpaired t-test). Sequence optimisation produced a 57-fold increase in expression. Note that full length dystrophin corresponds to a 427kDa protein.

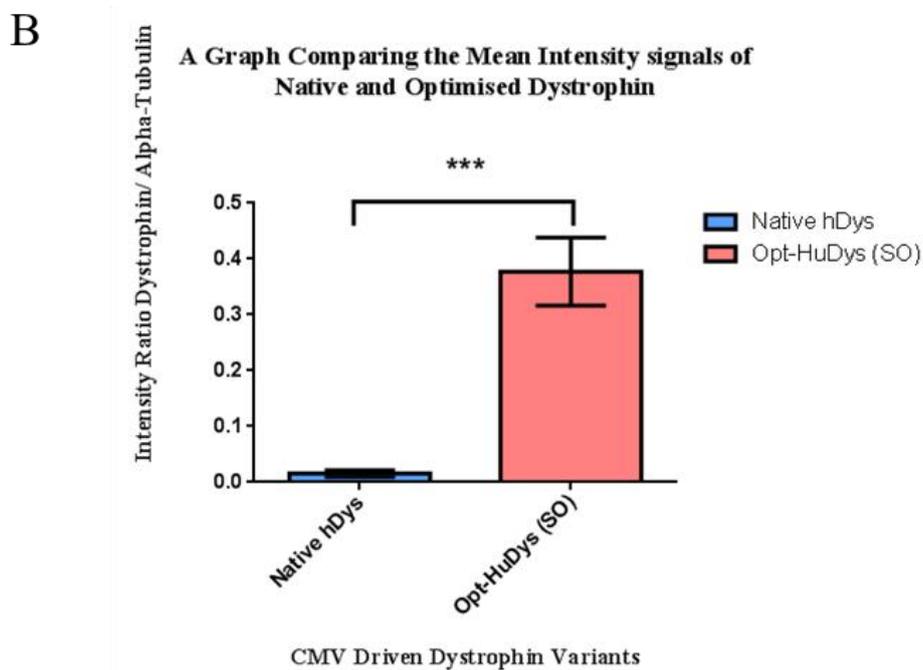
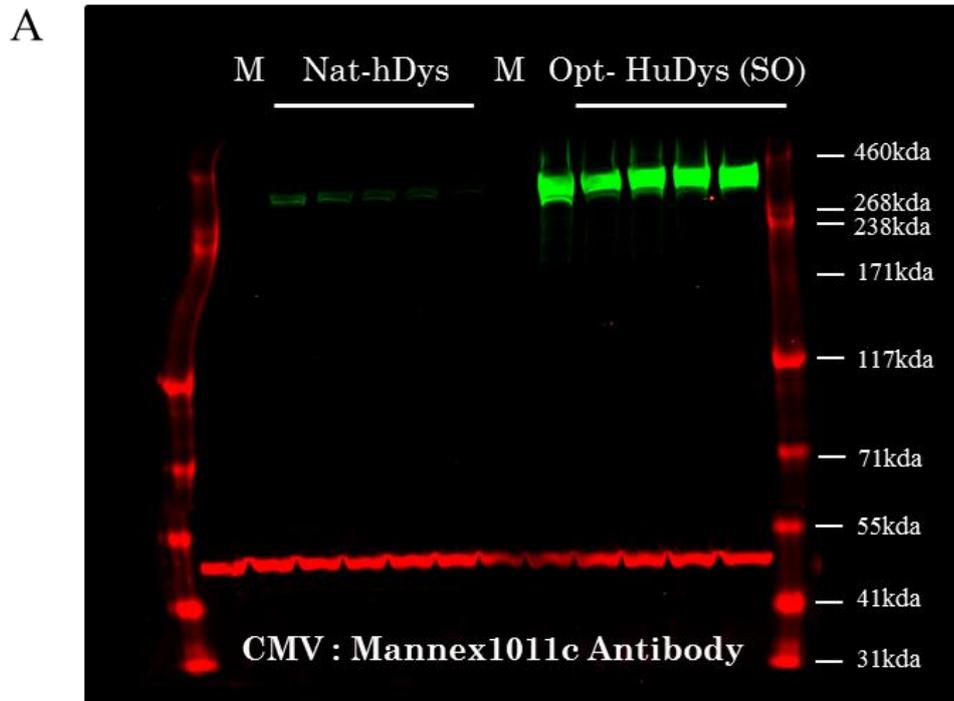


Figure 5-5 Optimisation of cDNA sequence increases recombinant dystrophin expression.

Plasmids expressing native (Nat-hDys) and sequence-optimised dystrophin (Opt-HuDys-SO) cDNAs from the CMV promoter were transfected into HEK293T cells (n=5). After 72h cultures were harvested and processed for Western Blotting (A) 50µg total protein lysate was analysed with antibodies to dystrophin (ManEx1011c) and alpha-tubulin. Dystrophin bands were then quantified relative to the alpha-tubulin loading controls and mean intensity ratios plotted (B: mean±SEM: *** p<0.001 (unpaired t-test). Sequence optimisation produced a 22-fold increase in expression. Note that full length dystrophin corresponds to a 427kDa protein.

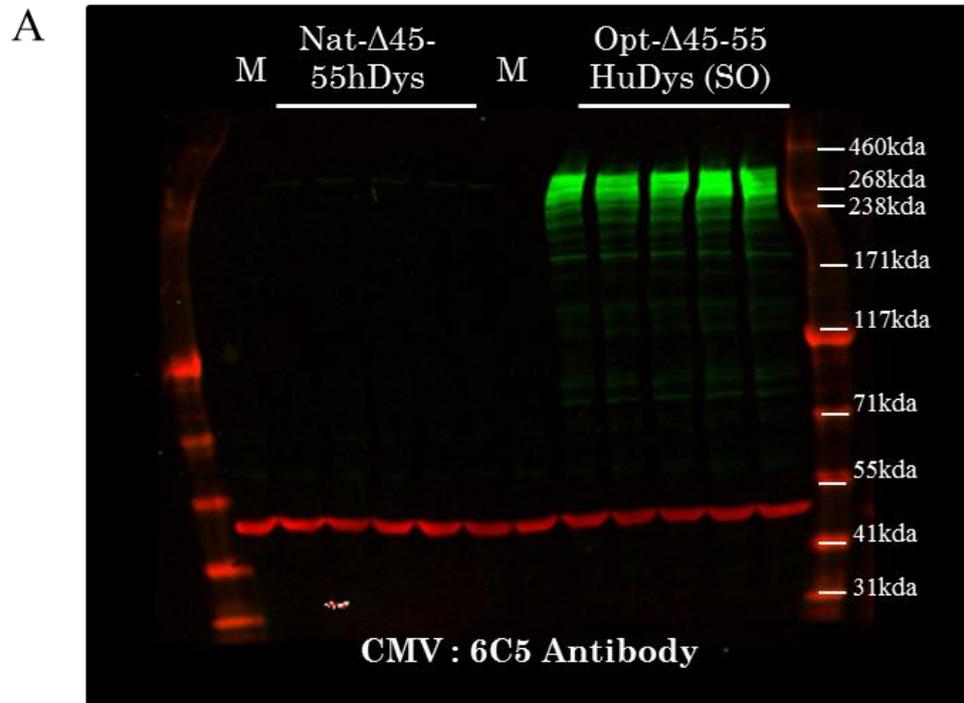


Figure 5-6 Optimisation of cDNA sequence increases recombinant $\Delta 45-55$ minidystrophin expression.

Plasmids expressing native (Nat- $\Delta 45-55$ hDys) and sequence-optimised dystrophin (Opt- $\Delta 45-55$ HuDys-SO) cDNAs from the CMV promoter were transfected into HEK293T cells (n=5). After 72h cultures were harvested and processed for Western Blotting (A) 50 μ g total protein lysate was analysed with antibodies to dystrophin (6C5) and alpha-tubulin. In this instance the samples could not be accurately quantified as the native quasi dystrophin had sub-optimal visualisation; banding in addition to the full-length band appeared apparent in the Opt- $\Delta 45-55$ HuDys-SO lanes. Note $\Delta 45-55$ minidystrophin corresponds to a 380kDa protein.

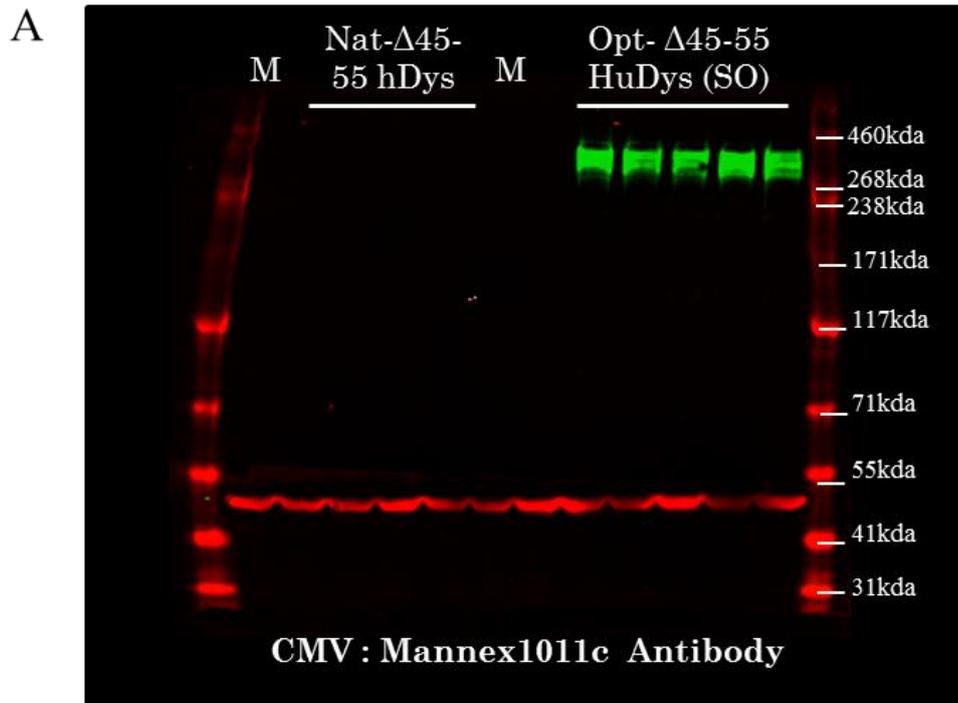


Figure 5-7 Optimisation of cDNA sequence increases recombinant $\Delta 45-55$ minidystrophin expression.

Plasmids expressing native (Nat- $\Delta 45-55$ hDys) and sequence-optimised dystrophin (Opt- $\Delta 45-55$ HuDys-SO) cDNAs from the CMV promoter were transfected into HEK293T cells (n=5). After 72h cultures were harvested and processed for Western Blotting (A) 50 μ g total protein lysate was analysed with antibodies to dystrophin (MannEx1011c) and alpha-tubulin. In this instance the samples could not be accurately quantified as the native quasi dystrophin had sub-optimal visualisation; banding in addition to the full-length band appeared apparent in the Opt- $\Delta 45-55$ HuDys-SO lanes. Note $\Delta 45-55$ minidystrophin corresponds to a 380kDa protein.

5.2.3 Quantifying the Difference in Expression of Native and Optimised Dystrophin Constructs Driven by a Spc512 Promoter:

CMV driven constructs were used in the series of investigations described above. However, in the context of translational application a muscle specific promoter would be advantageous in providing expression restricted to muscle (Meng et al. 2016). Nat-hDys and Opt-HuDys-SO constructs driven by the muscle specific Spc512 promoter (Li et al. 1999) were thus developed. In pursuit of this, 4µg of the constructs were transfected into HEK293T cells and total protein harvested at 72 hours for Western blot analysis. Protein samples were prepared, subjected to western blot and quantified in a manner consistent with the CMV based experiments. It should be noted that while HEK293T cells are not muscle derived, they do unexpectedly express constructs using the spc512 promoter

The Opt-HuDys-SO construct once again resulted in a protein band of a greater area and fluorescent intensity, relative to Nat-hDys with both the 6C5 and ManEx1011c antibodies with 15-fold and 17-fold increases respectively (Figure 5-8 and Figure 5-9). The laddering effect previously observed with the 6C5 antibody is retained even with the use of an alternative promoter.

5.2.4 An Examination of Native and Optimised Microdystrophin Constructs Driven by the Spc512 Promoter:

The ΔR4-23-ΔCT microdystrophin (MD1) is the shortest dystrophin variant in this investigation with a cDNA size of ~3.6Kb. Owing to its small size it was speculated that this transgene would be most likely to successfully integrate at the human *DMD* intron 1 site in a genome editing context. A sequence optimised variant of this dystrophin transgene was previously synthesised by the Dickson lab and placed downstream of the a Spc512 muscle specific promoter (Athanasopoulos et al. 2011). It was synthesised on the basis that this would be most favourable for translational gene augmentation studies, however direct comparison with native non-optimised forms had not been performed. The previous examination of Spc512 driven constructs encoding full length dystrophin cDNA, highlighted that sequence optimisation improved protein expression, even in HEK293T cell culture. This finding indicated that differences in MD1 protein expression of native and optimised cDNA under the control of the Spc512 promoter could be examined in the same manner.

A native microdystrophin sequence counter-part of MD1 was synthesised, characterised and sequenced by Dr Susan Jarmin. This construct was kindly provided to enable a comparison in protein expression resulting from native and optimised MD1 cDNA driven by the Spc512 promoter. As the small size of the MD1 cDNA prompted this to become the primary dystrophin

variant utilised during genome editing and HDR integration experiments, this arguably made an assessment of protein expression of these construct more pertinent.

Importantly, both Nat-hMD1 and Opt-HuMD1 (SO) expressed a 137kDa protein upon Western Blotting with the MannEx101c antibody consistent with the size anticipated. Once more the sequence optimised condition, produced a band with a greater area and much increased intensity relative to the native construct (Figure 5-10). The subsequent analysis indicated a 23-fold difference, which was asserted to be statistically significant by unpaired T-Test $p=0.0001$.

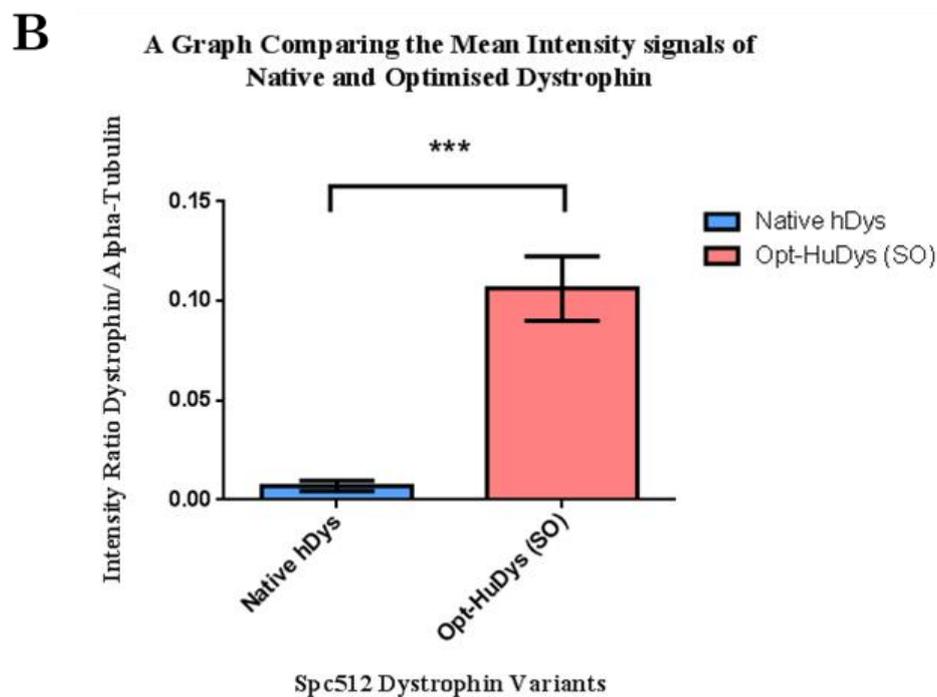
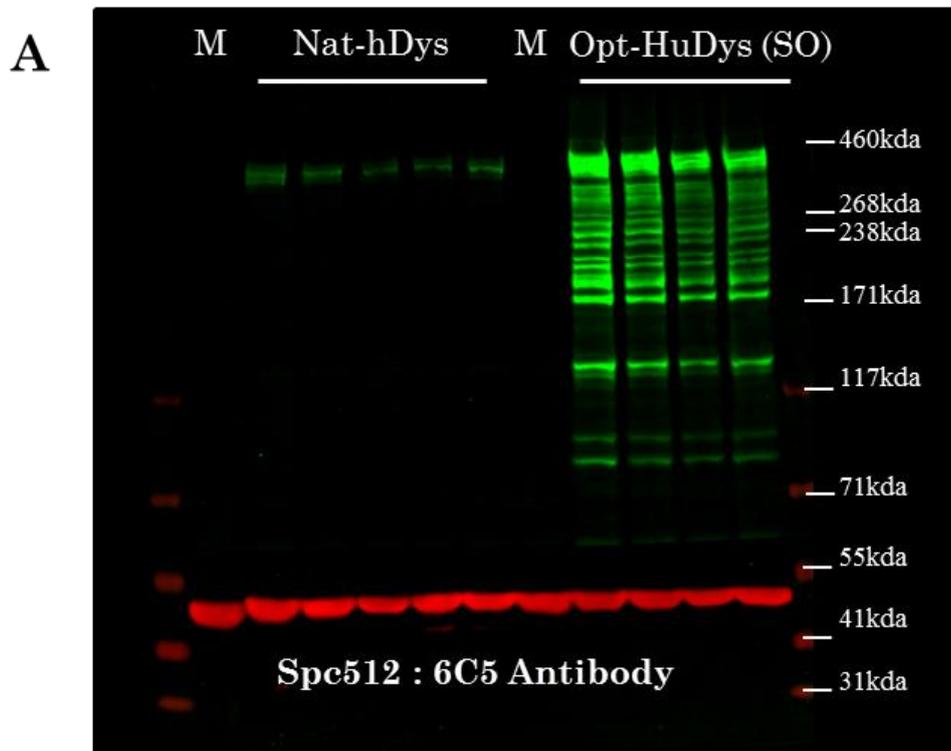
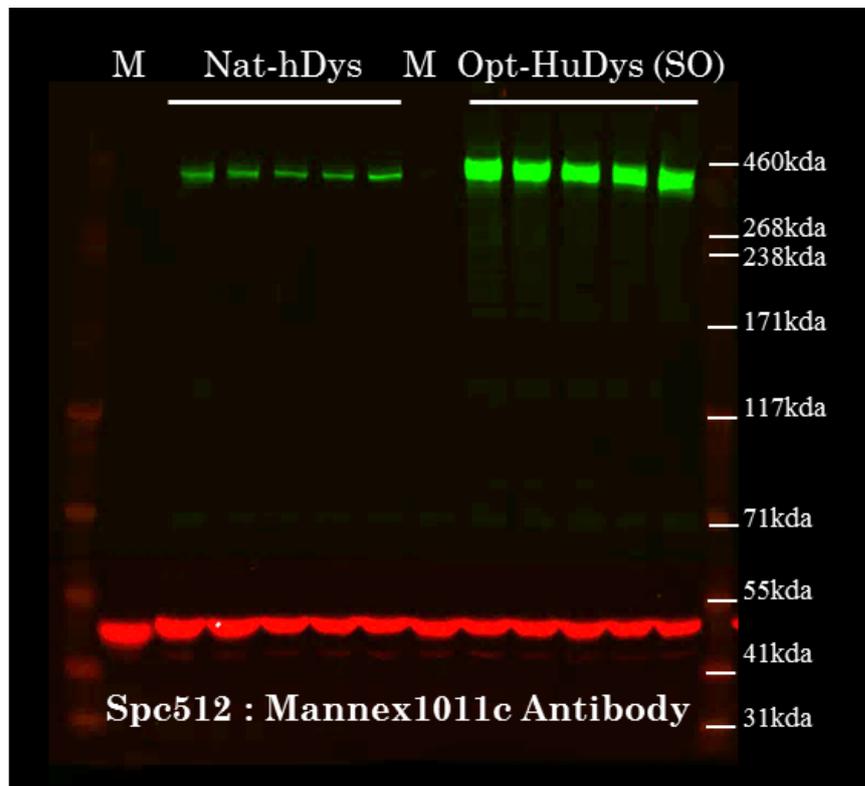


Figure 5-8 Optimisation of cDNA sequence increases recombinant dystrophin expression.

Plasmids expressing native (Nat-hDys) and sequence-optimised dystrophin (Opt-HuDys-SO) cDNAs from the Spc512 promoter were transfected into HEK293T cells (n=5). After 72h cultures were harvested and processed for Western Blotting (A) 50µg total protein lysate was analysed with antibodies to dystrophin (6C5) and alpha-tubulin. Dystrophin bands were then quantified relative to the alpha-tubulin loading controls and mean intensity ratios plotted (B: mean±SEM: ***8 p<0.001 (unpaired t-test). Sequence optimisation produced a 15-fold increase in expression. Note that full length dystrophin corresponds to a 427kDa protein.

A



B

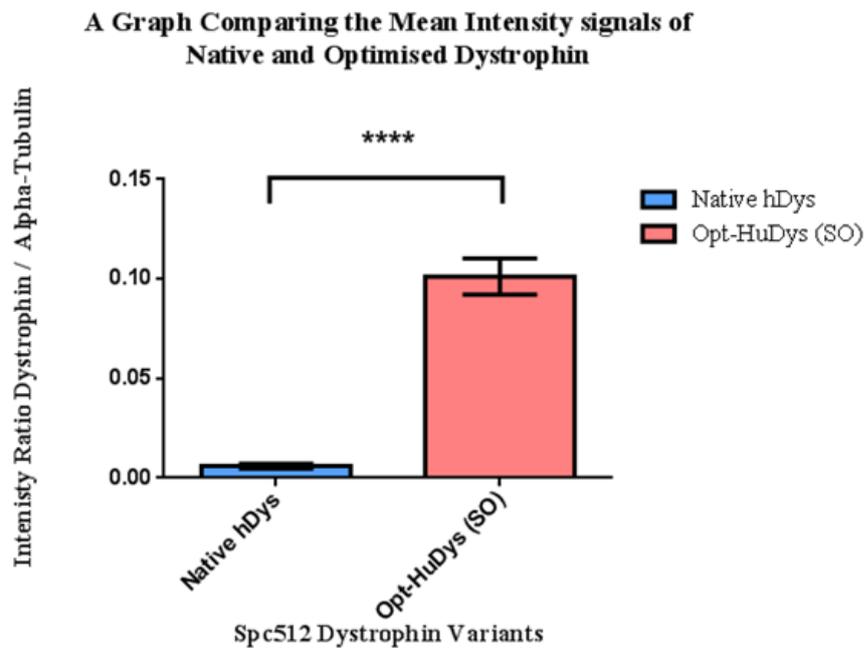


Figure 5-9 Optimisation of cDNA sequence increases recombinant dystrophin expression.

Plasmids expressing native (Nat-hDys) and sequence-optimised dystrophin (HuDys-CO) cDNAs from the Spc512 promoter were transfected into HEK293T cells (n=5). After 72h cultures were harvested and processed for Western Blotting (A) 50µg total protein lysate was analysed with antibodies to dystrophin (ManEx1011c) and alpha-tubulin. Dystrophin bands were then quantified relative to the alpha-tubulin loading controls, and mean intensity ratios plotted (B: mean±SEM: **** p<0.0001 (unpaired t-test). Sequence optimisation produced a 17-fold increase in expression. Note that full length dystrophin corresponds to a 427kDa protein.

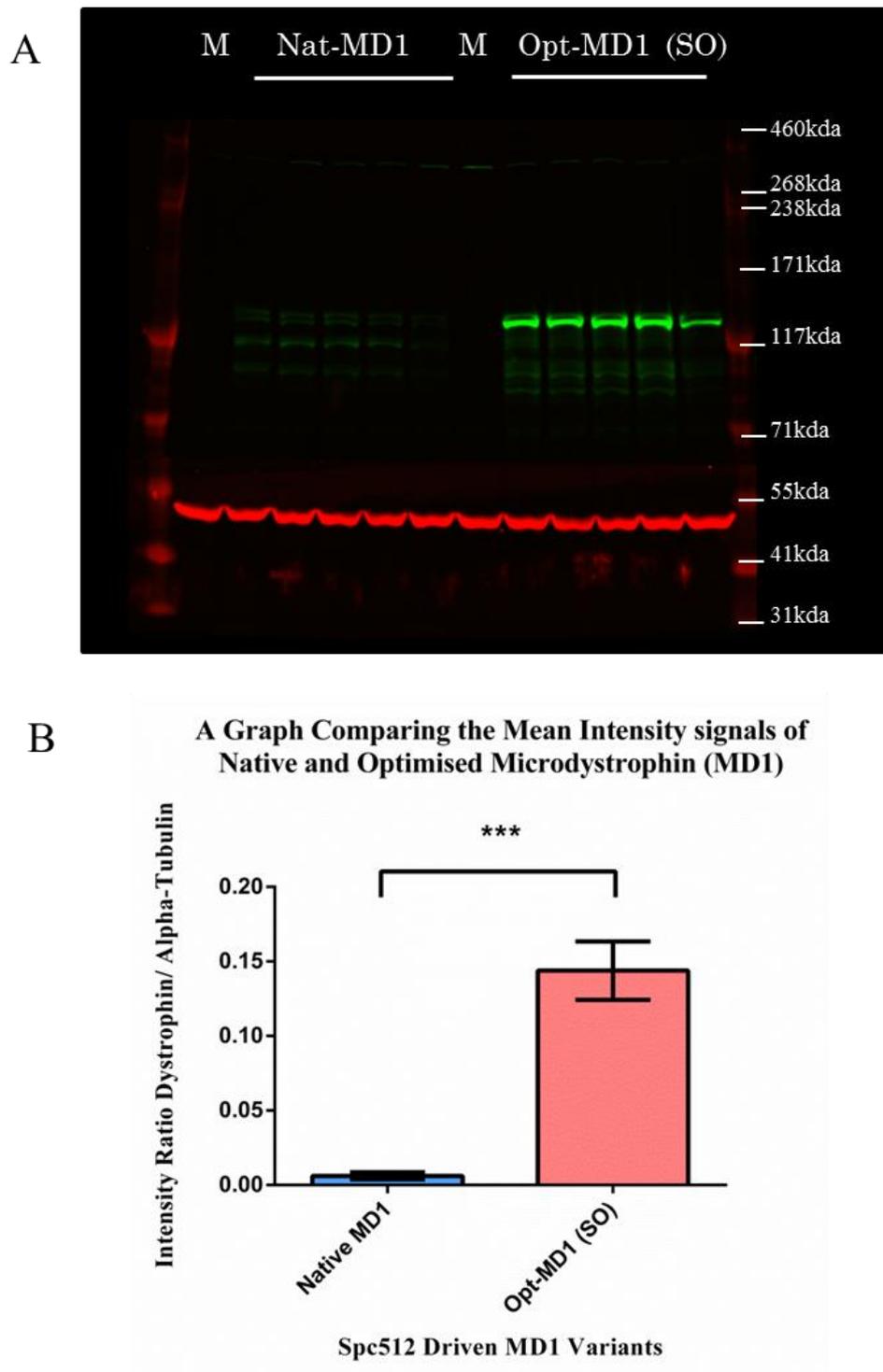


Figure 5-10 Optimisation of cDNA sequence increases recombinant microdystrophin expression.

Plasmids expressing native (Nat-MD1) and sequence-optimised dystrophin (Opt-MD1-SO) cDNAs from the Spc512 promoter were transfected into HEK293T cells (n=5). After 72h cultures were harvested and processed for Western Blotting (A) 50µg total protein lysate was analysed with antibodies to microdystrophin (ManEx1011c) and alpha-tubulin. Microdystrophin bands were then quantified relative to the alpha-tubulin loading controls, and mean intensity ratios plotted (B: mean±SEM: *** p<0.0001 (unpaired t-test). Sequence optimisation produced a 23-fold increase in expression. Note the microdystrophin construct refers to a 137kDa protein.

5.2.5 Design of an Exogenous Repair Template Plasmid Backbone Compatible with multiple dystrophin cDNA Variants:

The exogenous repair template was designed with numerous features to make it optimal for this investigation. Firstly, it included a backbone sequence amenable to the cloning of multiple dystrophin cDNA variants including full-length, $\Delta 45-55$ and MD1, all SO on the basis of expression studies performed above. Secondly, a splice acceptor was appended at the 5' end, in place of a promoter. The presence of the splice acceptor would enable the endogenous dystrophin full-length promoters Dp427c, Dp427m, Dp427p1, Dp427p2 and unique exon 1 elements to splice to the exogenous repair template cDNA. As a consequence, the resultant protein will have the correct spatial and temporal expression patterns. It is hypothesised that this will ameliorate a range of pathogenic disease causing mutations across the *DMD* gene, and has near universal applicability to DMD patients.

To attain a backbone sequence amenable to the cloning of a variety of dystrophin cDNA transgenes all unique restriction endonuclease sites were identified across full length Opt-HuDys (SO), Opt- $\Delta 45-55$ HuDys (SO) and Opt-HuMD1 (SO). This served to identify two unique restriction sites which were present and consistent across all of the above dystrophin plasmids. At the 5' end this was FseI, present 30 nucleotides into exon 6 and at the 3' end NotI situated at the terminus of exon 79. Thus a sequence was constructed with the following features:

- DMD cDNA of exons 2-6 upstream of an FseI site
- Intervening cyan fluorescent marker sequence
- 3'-located NotI site.

It was determined that this backbone construction would enable directional cloning of all sequence optimised dystrophin variants outlined previously, and this was the central premise around which the exogenous repair template was developed.

In construction of an Exon 2-6 DMD cDNA block, consensus sequences of exons 1-6 of the Dp427m isoform were aligned against full length sequence optimised dystrophin cDNA (Sequence unpublished, Dickson Lab). In this manner, exons 1-6 of sequence optimised dystrophin cDNA was identified, and exon 1 determined as the first 31 nucleotides of the sequence removed. It was anticipated that these nucleotides would be reconstituted by the endogenous Dp427m promoter-exon 1 sequence, upstream of the intron 1 *DMD* target site on successful RNA splicing, to integrated dystrophin transgenes. To this purpose, exon 2 was flanked at the 5' end with a human β -globin synthetic splice acceptor and other regulatory sequences to facilitate splicing, including a polypyrimidine tract and synthetic branch points

(Seth et al. 2008; Popplewell et al. 2013). This splice acceptor was selected to facilitate the splicing of the endogenous promoter to the integrated dystrophin transgene, on the basis of being ranked as a strong splice acceptor site by Human Splice finder (Desmet et al. 2009) and previously being used successfully by our lab (Popplewell et al. 2013). To provide the repair template with more flexibility, this sequence was modified to include a silent mutation, to generate a HpaI restriction site. This would enable the β -globin synthetic splice acceptor to be replaced with a splice sequence native to the *DMD* gene with ease should this be required.

The 3' end of this cDNA block, downstream of the NotI site, was also flanked by a sequence indicated to improve transcription in lentiviral vectors, a mutated woodchuck hepatitis virus post-transcriptional regulatory element (mWPRE) fused to a polyadenylation site polyA (Ranzani et al. 2013). This mutated WPRE and polyA tail sequence was kindly provided by Prof. Rafael Yanez.

Between these FseI and NotI sites, a cyan fluorescent protein (CFP) was encoded. The sequence was gifted by Michael Davidson, (Addgene plasmid #54595, unpublished) and modified to remove the initiating methionine and append the first two nucleotides of DMD exon 2. This was undertaken to retain the open reading frame of partial Dys-CFP fusion protein and reduce background fluorescence that may arise from aberrant firing of the methionine. This was anticipated to provide the benefit of a visual blue-fluorescent output, which could be used to indicate Dys-CFP transgene integration and track subsequent enrichment processes. It was anticipated that this would streamline the development of integration methods in HEK293T cells and DMD patient myoblasts. The presence of a 491bp band removed upon FseI and NotI double digest could be used to indicate successful cleavage of the repair template and facilitate sub-cloning of dystrophin cDNA variants.

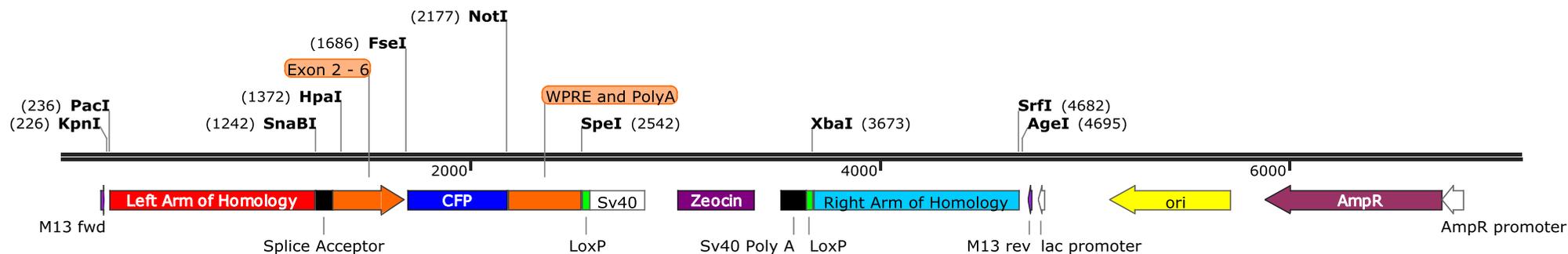
Due to the HDR DNA pathway utilised in gene-editing, occurring with relatively low efficiency, an antibiotic selection cassette, SV40-Zeocin-PolyA, was placed downstream of the WPRE and PolyA sequences (Heyer et al. 2010). This encodes the *She Ble* protein, which when expressed renders the zeocin antibiotic inactive. This sequence was gifted by Rudolf Jaenisch (Addgene Plasmid #20725) and modified to include a silent point mutation to remove a FseI site present. This ensured the unique core FseI and NotI sites required for directional sub-cloning of dystrophin variants was retained.

The decision to use a zeocin selection cassette was made because during the immortalisation process of DMD patient myoblasts harbouring the deletion of exons 45-52 ($\Delta 45-52$), resistance to puromycin and neomycin antibiotics was introduced (Mamchaoui et al. 2011). Importantly, the zeocin cassette was floxed with LoxP sites that can conditionally remove

intermediate sequences in the presence of Cre-recombinase. Continued expression of the SheBle protein is undesirable in any future clinical context (Marie et al. 2010) and Cre-mediated excision can provide a solution.

Once the sequences of all constituents were assembled, they were flanked with 1kb arms of homology. These were derived from human DMD Intron 1 consensus sequence from NCBI. The two 1kb arms initiated either 6 nucleotides upstream of guide 3 or 6 nucleotides downstream of guide 4, these being the most upstream and downstream of the CRISPR guides identified. Consequently, none of the selected guide RNA target sequences were encoded within the exogenous repair template. This feature circumvents the risk that the repair template plasmid DNA may be cleaved by the Cas9 system in situ, or indeed that dystrophin cDNA variants might re-targeted following integration into the genome.

Once the full sequence of the human DMD intron 1 repair template was compiled, it was assessed in parallel with dystrophin variants and the ISceit-Lentiviral vector for the absence of common restriction endonucleases. The list of common non-cutter restriction endonucleases was examined for compatibility in a double digest setting and used to flank all components of the exogenous repair template. This sequence was then submitted to GenScript, synthesised and sequenced in a pUC57 backbone (Figure 5-11). Finally, the sequence was inserted in reverse orientation into the ISceit lentiviral backbone, to maintain the viral mRNA structure and prevent aberrant splicing or termination occurring (Figure 5-12).



Puc57-hINT1-RT
7129 bp

Figure 5-11 A linear schematic of the pUC57- human *DMD* intron 1 plasmid repair template, synthesised by GenSys.

A linear schematic of pUC57-hINT1-RT empty vector. The FseI and NotI sites flanking the Cyan fluorescent marker, serve as the directional cloning sites for all optimised dystrophin variants: MD1, Quasi-Dystrophin and Full length dystrophin. Upstream of this cloning site is a 1kb left arm of homology, synthetic beta globin splice acceptor and Exons 2-6 of optimised *DMD* cDNA. Downstream of this cloning sequence is the WPRE and polyA to enhance expression followed thereafter by a floxed zeocin cassette to facilitated selection and a 1kb right arm of homology. Where possible this constituents are flanked by unique restriction sites that are annotated, ensuring that the repair template is not only common to all optimised forms of dystrophin but also amenable to changing constituents if required.

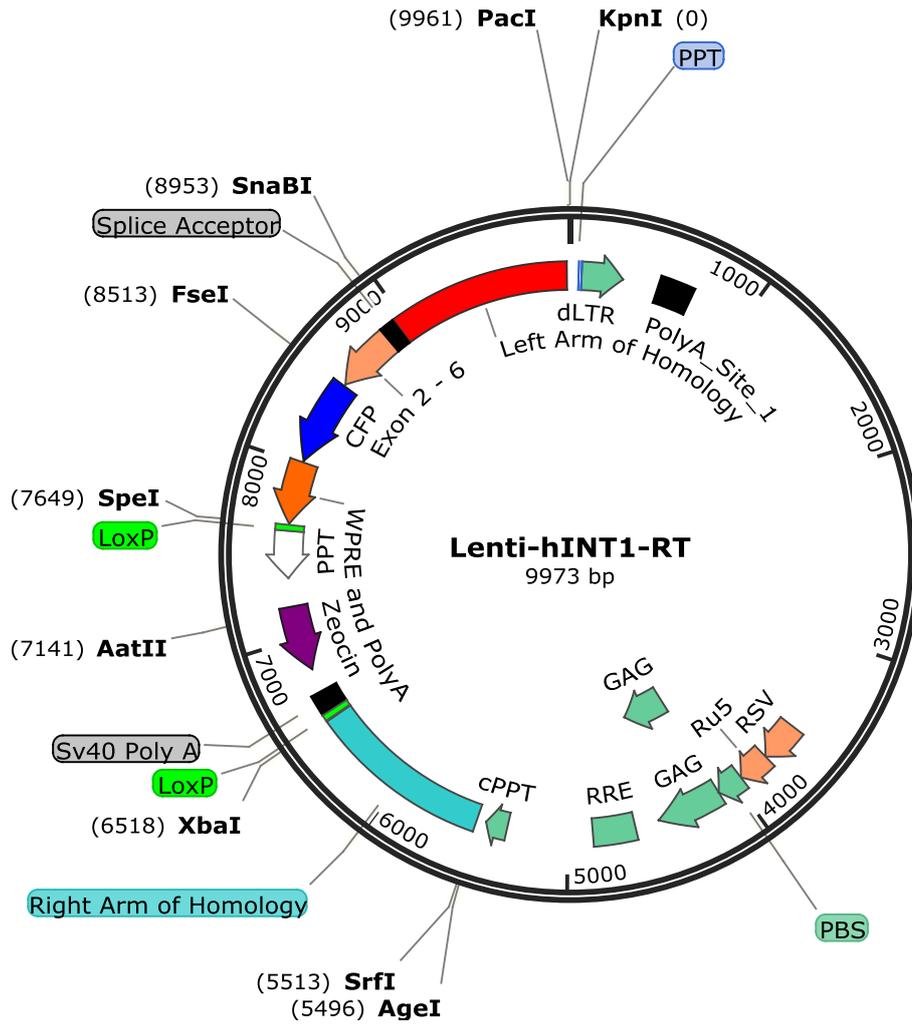


Figure 5-12 A schematic of the Lentiviral human *DMD* intron 1 plasmid repair template:

A lentiviral vector encoding the human *DMD* Intron 1 repair template; notably, the constituents of the human *DMD* intron 1 plasmid run in reverse orientation to the transcription of the lentiviral plasmid; this is to prevent aberrant splicing from occurring that may be detrimental to viral mRNA transcription and subsequently viral production. This includes the *FseI* and *NotI* cloning sites downstream of optimised *DMD* cDNA exons 2-6, CFP protein and Woodchuck hepatitis virus regulatory element (WPRE). In addition the upstream and downstream 1kb arms of homology isogenic to the genomic sequence adjacent to the guide sites in intron 1 and the floxed zeocin cassette are also present.

5.2.6 Sub-cloning MD1 into the pUC57 Intron 1 Exogenous Repair Template:

Following the synthesis of the Intron 1 exogenous repair template, focus was shifted to the sub-cloning of the MD1 from exon 6 to exon 79 between the FseI and NotI endonuclease restriction sites. To this purpose MD1 was subject to a series of diagnostic digests to assert the identity of the construct (Figure 5-13). Once assured the construct digested in a manner consistent with that anticipated, both the destination vector pUC57-hINT1-RT and MD1 plasmids were subject to double digest with FseI and NotI restriction endonucleases. In the case of the destination vector the double digest served to remove the CFP marker and leave a 6638bps backbone, whereas in the case of the MD1 it enabled the cDNA insert of interest to be liberated resulting in a 3213bp band (Figure 5-14). Samples were resolved by electrophoresis and gel purification, prior to being subjected to overnight ligation.

Subsequent ligation mixtures of insert and destination vector and controls were then subject to standard heat shock transformation in Top-10 cells. The E.coli suspension was allowed to recover and plated onto ampicillin plates for an overnight incubation at 30°C. The vector control yielded no colonies, indicating that no self-ligation had occurred, likely attributable to the incompatible DNA termini resulting from the double digest. In contrast, the ligation mixtures yielded a high number of single colonies that could be picked and characterised. These results taken together serve to indicate that the MD1 fragment from exon 6-79 was likely to have ligated correctly into the destination vector backbone.

To assess whether ligation was successful a colony PCR was undertaken. The primer pairs were designed for the colony PCR so that the forward primer was present within the Exon 2-6 region of the pUC57-hINT1-RT backbone upstream of the FseI cleavage site, whereas the reverse primer with the MD1 transgene, would only be present if the MD1 insert was ligated (Figure 5-15). Thus the 221bp resultant amplicon was only anticipated to occur in instances where the MD1 transgene had been successfully inserted into the pUC57-hINT1-RT destination vector in the correct orientation. Amplification of the Spc512-HuMD1 (SO) plasmid was used as a positive control, and the pUC57-hINT1-RT destination vector pre-digest and ligation served as a negative control. Strikingly, all of the 32 colonies screened were positive for the presence of a 221bp amplicon. This indicated that the MD1 transgene was successfully inserted into the pUC57-hINT1-RT destination backbone in all colonies screened.

The first 4 colonies from the colony PCR series were selected and grown as a starter culture, the resulting plasmid isolated and subject to restriction digest confirm presence and identity of the MD1 transgene insert. Initially, FseI and NotI double digestion was performed on putative pUC57-hINT1-MD1-RT clones alongside the parental pSpc512-HuMD1-SO plasmid, which provided the insert. If ligation was successful a 3213bp fragment would be liberated from the

pUC57-hINT1-RT backbone as opposed to the former 491bps CFP fragment. All four colonies gave the anticipated digest profile (Figure 5-16A).

Additionally, a diagnostic digest using the *ScaI* restriction endonuclease was undertaken. This produces distinct banding patterns for the parental Spc512-HuMD1 (SO) and the pUC57-hINT-MD1-RT, resolving as two bands (5569bp, 2664bp) and three bands (5422bp, 2517bp, 1912bp) respectively. Once again the four colonies gave the anticipated digest profiles, consistent with MD1 being inserted into the pUC57-hINT-RT. The above screens indicate a repair template carrying MD1 exons 2-79 and an independent selection cassette, all flanked by arms of homology had been successfully produced (Figure 5-16B).

A



B

Enzyme	Restriction Fragments
HpaI	8233bps
MfeI	8233bps
NcoI	7600, 633bps
NdeI	7771, 462bps
ScaI	5569, 2664bps
SnaBI	4308, 3925bps
XbaI	7865, 368bps

Figure 5-13 Digest profile of Spc512-MD1 plasmid, prior to cloning into the pUC57-hINT1-RT destination vector:

A series of standard 20µl digests were undertaken for 3 hours at 37°C on 1µg of Spc512-MD1 plasmid, alongside an uncut (UC) control. A 10µl aliquot of these digests was resolved on a 1% TAE agarose gel alongside bioline hyperladder I. Notably, the fainter smaller size bands are indicated by a star *. B) A table of the anticipated DNA profiles, showing all of the anticipated fragments for each of the restriction digests undertaken on the Spc512-MD1 plasmid. All of the bands were present and consistent with those anticipated enabling the sub-cloning of exon 6-79 to proceed as planned.

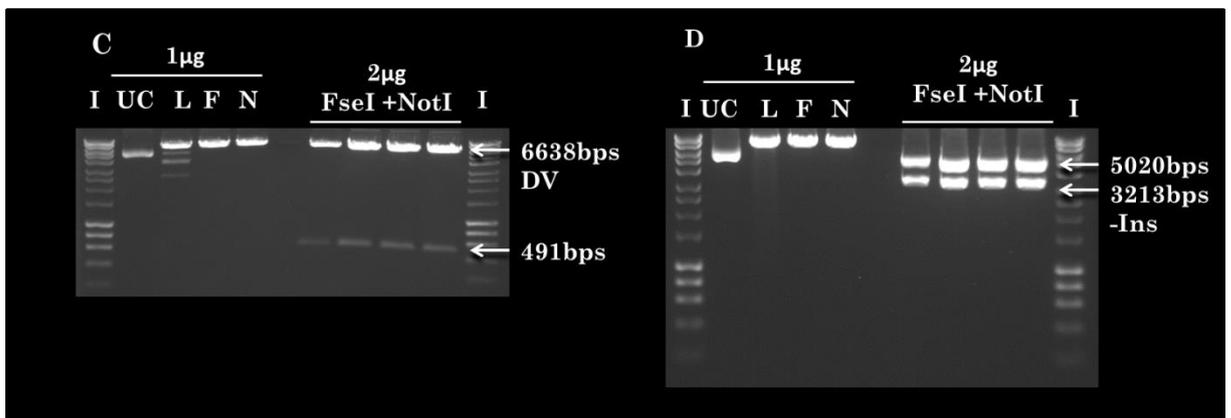
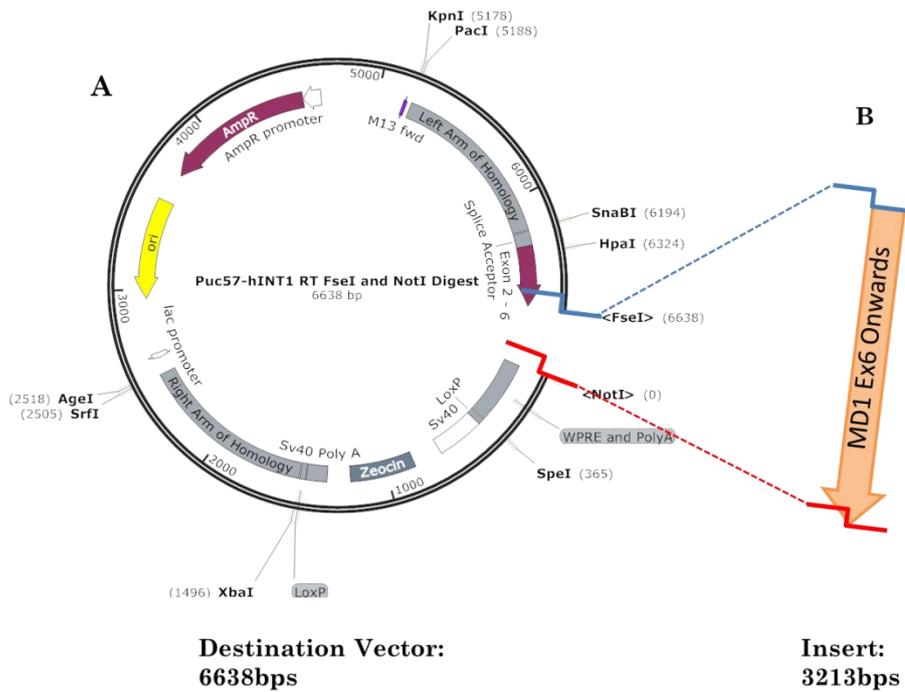


Figure 5-14 Parallel double digest of the pUC57- hINT1-RT and Spc512-MD1 plasmids for directional cloning of MD1 to reconstitute the MD1 transgene spanning exon2-79 in an exogenous repair template.

A) A schematic of the pUC57-hINT1-RT, post FseI and NotI double digest, removing the Cyan Fluorescent protein, thereby serving as a destination vector enabling the insertion of MD1 exon 6-79. B) A schematic of the MD1 exon 6-79 insert, liberated with FseI and NotI double digest. In both instances the FseI cleavage site is denoted in blue and the NotI cleavage site is denoted in red. C) A digest series of pUC57-hINT1-RT digests included: Uncut plasmid (UC), Linearised control (L) –single digest undertaken with SpeI, FseI single digest (F), NotI single digest (N) and FseI and NotI double digest. All single and double digests were undertaken on 1µg and 2µg of plasmid at 37°C for 2 hours respectively. D) The same digest series described in C undertaken on the Spc512-MD1 plasmid. All digests were resolved on a 1% agarose, in 1X TAE buffer; alongside bioline hyperladder I. Sizes of digest fragments are indicated.

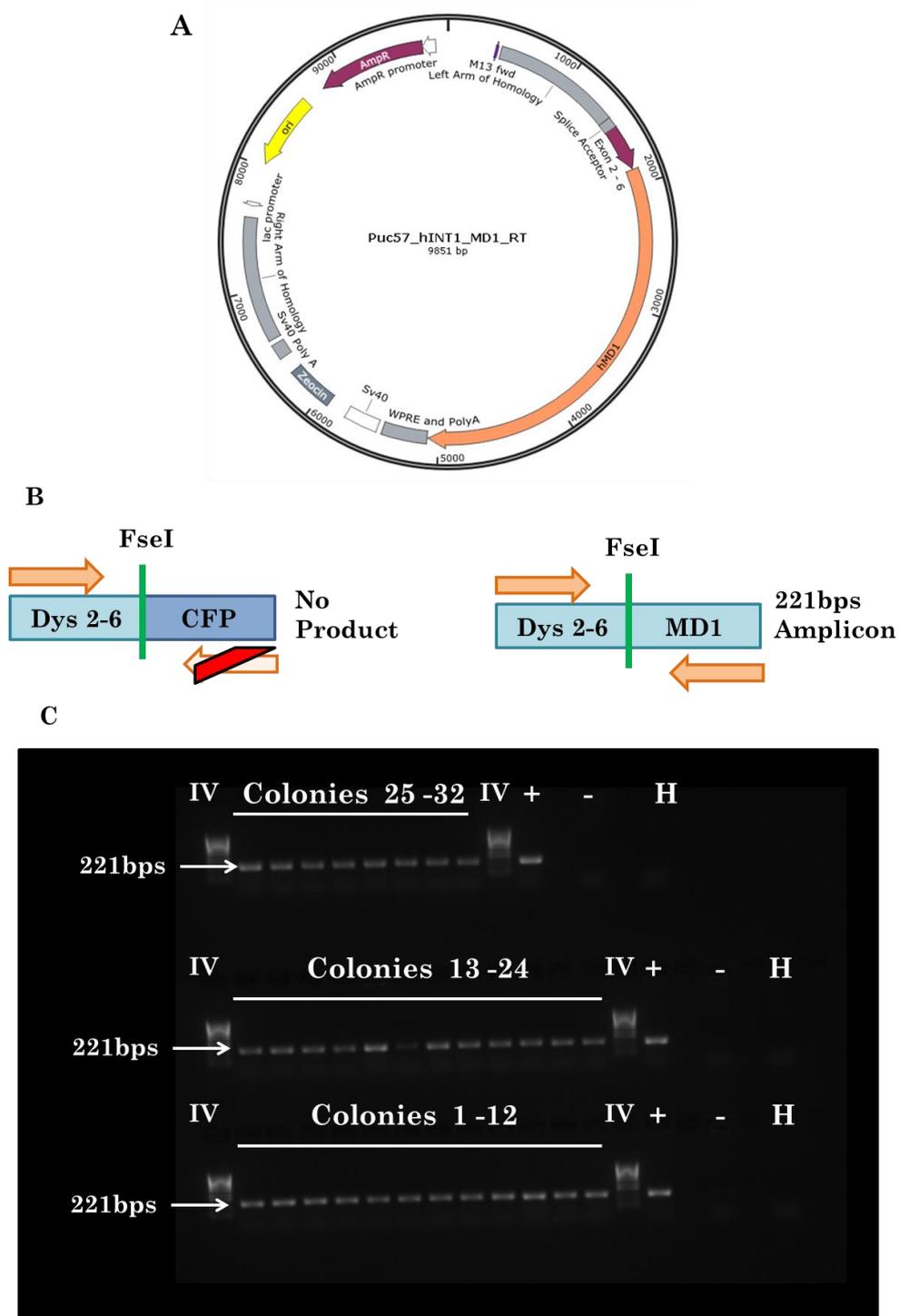


Figure 5-15 Complete pUC57-hINT1-MD1 Δ Ex1 plasmid, characterised by colony PCR.

A) A schematic plasmid map of the pUC57-hINT1-MD1 Δ Ex1-RT, notice that in this instance the MD1 exon 6-79 region has been sub-cloned directionally into the plasmid. B) A schematic of the colony PCR designed: the left indicates the outcome if the MD1 exons 6-79 insert is absent, the right indicates the outcome if the MD1 insert is correctly ligated. Notably, the forward primer binds upstream of the FseI site regardless of whether MD1 transgene insertion has occurred; in contrast the reverse primer only binds in the presence of the MD1 insert. Thus amplification will only be successful in the presence of the MD1 exon 6-79 insert. C) A 10 μ l aliquot of PCRs of 1 μ l DNA from lysed bacterial colonies 1-32, Spc512-MD1 (+), empty pUC57-hINT1-RT (-) and finally a water control (H). All samples were run on a 1% agarose gel in 1XTAE buffer alongside bioline hyperladder I. Sizes of PCR products are indicated.

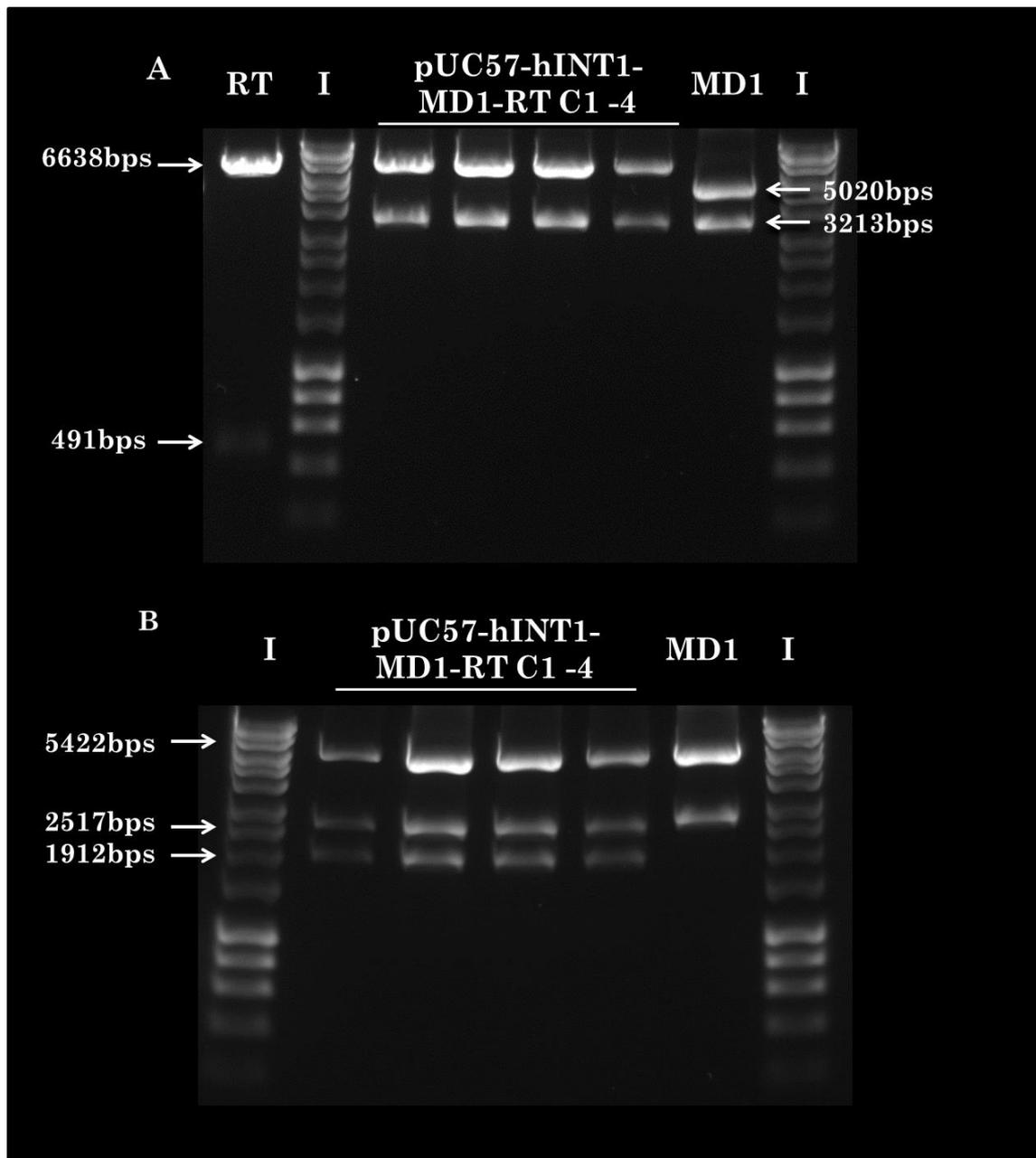


Figure 5-16 Digest profiles of four pUC57-hINT1-MD1-RT colonies:

A) A standard FseI and NotI double digest undertaken at 37°C for 3 hours on an empty pUC57-hINT1-RT (RT), four pUC57-hINT1-MD1-RT colonies and on Spc512-MD1 (MD1). Notably, the backbone of the empty RT and the four colonies is consistent at 6638bps; furthermore the MD1 insert size was consistent for the MD1 and the four colonies being 3213bps. B) A standard Scal digest undertaken on the four Puc57-hINT1-MD1-RT colonies and on the Spc512-MD1 plasmid (MD1) alone. Interestingly, a distinct DNA profile of 3 digests of 5422bps, 2517bps and 1912bps, was apparent in pUC57-hINT1-MD1-RT colonies consistent with expectations. This differed from the digest profile of the Spc512-MD1 plasmid, which only produces the first two larger bands. A 10µl aliquot of all digests were resolved on a 1% agarose gel and run in 1xTAE buffer alongside Bioline hyperladder I. Sizes of digest fragments are indicated.

5.3 Discussion

5.3.1 Sequence Optimised Dystrophin cDNAs:

Sequence optimisation of full length and $\Delta 45-55$ dystrophin cDNAs was shown to greatly enhance protein expression relative to native controls. This was established by microscopy imaging and Western Blotting. Both studies utilised plasmid constructs under the control of a CMV promoter, with the first investigation using dystrophin variants fused direct to eGFP. Importantly, in the case of full length dystrophin, up to a 57-fold difference in protein expression was observed between native and optimised constructs by Western Blotting. Whilst, such a difference could not be accurately determined for the $\Delta 45-55$ dystrophin constructs, the large increase in protein expression from optimised cDNA from both experiments was striking.

The investigation was then extended to include full-length dystrophin cDNA under the control of the Spc512 muscle restrictive promoter (Li et al. 1999; Athanopoulos et al. 2011). This was performed to examine whether the expression of Spc512 driven constructs could be compared in HEK293T cell culture. Furthermore, it was performed with the outlook of comparing Spc512-Nat-hMD1 and Spc512-Opt-HuMD1 (SO). This cell culture model was used on the basis of more amenable transient transfection profiles relative to the immortalised $\Delta 45-55$ patient myoblasts. Importantly, the Spc512 promoter enabled full-length dystrophin expression in HEK293T cell culture; this is speculated to be due to this cell-line facilitating leaky expression of this promoter. Interestingly, native and optimised full length dystrophin cDNA produced a varying fold difference in protein expression, when constructs were driven by the CMV and Spc512; being 22 fold and 15-17 fold respectively. This effect is attributed to two main parameters, the first being the relative strengths of the viral CMV and the muscle restrictive Spc512 promoters, and secondly, the expression of the full-length dystrophin cDNA with the Spc512 promoter may be reduced in HEK293T cells due to its restrictive expression pattern.

This trend of sequence optimisation enhancing protein expression extended to the Spc512 native and optimised MD1 plasmids, which showed a 23-fold increase in expression. Notably this difference is greater than the highest 17 fold increase determined for full length native and optimised dystrophin cDNA driven by the Spc512 promoter. This is speculated to be due to the smaller size of the transgene. The MD1 cDNA is transcribed in a shorter time period relative to the full length dystrophin counterpart. During the 72 hour incubation period prior to protein harvest, it is speculated that this would enable a greater accumulation of MD1 transcript and an increased overall protein synthesis. To confirm this speculation, a further investigation comparing mRNA levels of native and optimised constructs using quantitative PCR is warranted. Interestingly, the difference between native and optimised $\Delta 45-55$

dystrophin, although not quantified visually appeared to be greater relative to full-length constructs, as indicated by the striking increase protein band area and fluorescent intensity in optimised conditions. This warrants further investigation, but could validate the suggestion that shorter optimised transgenes result in greater fold difference in protein expression.

More accurate determination of the protein fold difference of the native and optimised $\Delta 45-55$ and indeed other dystrophin variants could be achieved with refinement of quantification strategies employed. For example in the comparison of CMV-driven full length and $\Delta 45-55$ dystrophin variants fused to eGFP, the exact difference in fluorescence intensity across native and optimised conditions was difficult to ascertain. This was due to occurrence of cell aggregation, which served to limit the quantification that could be achieved by FACs analysis. This cell adhesion is could result from overexpression of dystrophin transgenes, which in turn may have exhausted transcriptional machinery and lead to cell lysis. Now that an overall trend of increased protein expression from optimised cDNA has been identified, future experiments would aim to employ a dose range and include a Live/Dead FACs stain to identify and/or compensate for reductions in cell viability.

Furthermore, in the context of Western Blotting studies the accurate determination of fold differences in protein expression resulting from native and optimised $\Delta 45-55$ dystrophin cDNA was challenging. In this investigation, dystrophin intensity was normalised to alpha tubulin intensity for both conditions, and subsequent the ratios used to assert a fold difference. However, the large difference in $\Delta 45-55$ dystrophin intensity between native and optimised constructs made this method challenging to employ. If this normalisation strategy is to be continued, serial dilutions of the total protein from optimised conditions, may allow fold differences in protein expression to be determined. It is speculated that a 2-fold or a 4-fold, dilution of total protein harvested from optimised conditions, would allow visualisation of dystrophin band in both series. Particularly bands whilst have a determined value are displayed relative to the most intense band detected, in the associated channel using the Odyssey LI-COR. This would allow calculations of fold differences to be undertaken with the use of a dilution factor.

In contrast, alternative normalisation strategies could be employed, which normalise dystrophin protein to total protein within the lane. This could be facilitated by the commercially available REVERT stain which functions much like Ponceau Red, staining all protein and emitting within the 700nm channel (Blot & Handbook 2017). This normalisation strategy would circumvent possible limitations which could lead to overestimation of differences in protein expression, one example being the potential saturation of the

nitrocellulose membrane with the alpha tubulin protein. This saturation of the nitrocellulose membrane with alpha-tubulin protein during the transfer process can occur when high concentrations of total protein are used (Blot & Handbook 2017) such as the 50µg loading required to visualise dystrophin protein resulting from native cDNA. Interestingly, quantification proved more difficult with cDNA constructs driven by the CMV promoter where differences are suggested to be larger. Notably, in the case of native and optimised full length dystrophin it is speculated that overestimation of a 57-fold difference was achieved with Western Blotting using the 6C5 antibody. This is opposed to 22-fold difference determined with the MannEx1011c. Arguably, this could also be attributable to additional banding present in close proximity to the full length dystrophin in optimised conditions. The exact identity of these smaller dystrophin protein bands are currently unknown, one explanation is that the smaller dystrophin proteins result from the firing of downstream in frame methionine residues encoded by the full length dystrophin variant. The use of alternative translational initiation codons has been observed and confirmed clinically in patients harbouring mutations at the 5' end of the *DMD* gene, focused to exons 1 and 2 (Gurvich et al. 2009). Interestingly in these instances, whilst patients should be anticipated to have a Duchenne muscular dystrophy phenotype based on non-sense mutation present, the translation from downstream in frame initiating codons situated in exons 6 and 8 results in a shortened dystrophin protein being translated and a Becker muscular dystrophy phenotype. A second explanation that is plausible is proteolytic degradation products, that may be more produced due to the increased expression of dystrophin protein. Nonetheless, the observations from both CMV and Spc512 studies together could hold important implications for clinical translation. Numerous studies have provided an insight into the amount of dystrophin protein expression required relative to wild-type endogenous levels to ameliorate dystrophic pathology. Variable estimates have arisen, likely attributable to the nature of investigation from which estimates were derived and how dystrophin levels were quantified. Dystrophin expression of 30% in BMD patients (Neri et al. 2007; Anthony et al. 2011), 15% following antisense therapy in mice (Godfrey et al. 2015) and finally 20% in transgenic mice (Chamberlain 1997) were all shown to confer therapeutic benefit. Whilst estimates of the minimum dystrophin expression levels required to be therapeutic vary between 15-30% among these investigations, the studies agree that a uniform dystrophin expression across the majority of myofibres as opposed to a sporadic distribution provides greater functional improvement. In addition, they indicate that the level of dystrophin correction required to be therapeutic will be influenced by the specific muscle pathology and disease progression of the patient seeking treatment. The striking increase in protein expression observed here with the use of sequence optimised cDNAs encoding a range of dystrophin variants could facilitate the attainment of such expression thresholds. By

extension of this it is speculated they could greatly improve clinical outcomes and functional improvements observed in clinical trials.

The striking fold differences in protein expression observed between native and optimised constructs necessitate the examination of the potential effects of supra-physiological levels of dystrophin proteins. Importantly, others have demonstrated that overexpression of full length dystrophin up to 50 fold higher than endogenous levels was well tolerated (Chamberlain 1997). This level was shown to ameliorate dystrophic pathology with the absence of any detrimental effects. This finding was supported and extended by Yue et al who produced similar overexpression of the mini-dystrophin (Δ H2-R19) in transgenic mdx mice in skeletal and cardiac tissues. Detrimental effects, particularly cardiotoxicity, only became apparent when dystrophin expression of 100 fold greater than physiological levels was attained (Yue et al. 2016). Notably, this mini-dystrophin has an alternative architecture to MD1 (R4-23 Δ CT). Similarly, however, the sequence optimised MD1 has been delivered in many studies to assess its safety and efficacy profile (Koo et al. 2011; Le Guiner et al. 2017; Wang et al. 2000). The most recent study, administered a canine optimised rAAV-MD1, loco-regionally to muscle or intravenously to the GRMD *DMD* model. The findings from this indicated that not only was MD1 safe and efficacious in vivo, but it also showed sustained expression for over two years, with an associated reduction in dystrophic pathology (Le Guiner et al. 2017). The supra-physiological and MD1 in-vivo studies taken together highlight that sequence optimised microdystrophin provides enhanced protein expression that is unlikely to have deleterious effects if used clinically. It should be noted that in-vivo trials of the Δ 45-55 Minidystrophin and full-length optimised dystrophin are warranted to independently assess their safety and efficacy. This has not yet been undertaken as suitable delivery systems need to be developed.

5.3.2 Design and Construction of Exogenous Repair Template Directed to DMD

Intron 1:

The improved protein expression resulting from sequence optimised MD1, Δ 45-55 and full length dystrophin cDNA, served to influence exogenous repair template design. This prompted the generation of a backbone that would enable the directional sub-cloning of each of the aforementioned dystrophin variants, allowing the exogenous repair template to be easily adapted for each to be trialled in integration experiments.

It could be envisaged that future microdystrophin constructions, like the one employed in this study, may include additional spectrin repeats with the outlook to restore the correct localisation of dystrophin associated signalling molecules. This is exemplified by a recent study in which a 5 spectrin repeat microdystrophin generated by Hakim et al, during the production

of this manuscript that promoted the restoration of the nNOS (Hakim et al. 2017). Alternatively, constructions with differing hinge compositions may also be generated, particularly as studies have also served to highlight the influence of hinge regions on MD1 function. Interestingly, it was shown that the inclusion of hinge 3 instead of hinge 2 was found to increase the structural flexibility of MD1 and in mdx mice resulted in slightly larger myofibres and reduced evidence of Achilles myotendinous disruption and aberrant neuromuscular junctions (Banks et al. 2010).

As a consequence of these findings a range of different microdystrophin configurations continue to be developed and, the exogenous repair template described here would be able to accommodate such new emerging microdystrophins, thus providing a useful tool in future integration experiments with a repertoire of microdystrophin cDNAs.

It should be noted that the exogenous repair template presented thus far has retained a human focus. Interestingly, the FseI and NotI sites are conserved within mouse sequence optimised dystrophin cDNA constructs. Thus with alteration of the 1kb arms of homology to contain sequences isogenic to the mouse genome, and in conjunction with a murine dystrophin cDNA spanning exons 2-6, this exogenous repair template could be utilised in the same manner. Thus the common structure could be used to target both human and mouse cell lines during integration studies (Byrne et al. 2015; Song & Stieger 2017).

The exogenous repair template was designed to facilitate an HDR-mediated genome editing outcome at the human DMD intron 1 locus. It was designed with 1kb arms of homology isogenic to the human genome directly upstream and downstream of the CRISPR targeted region of homology identified. This length of arms of homology of isogenic sequence is deemed adequate for successful exploitation of HDR pathways. Extension of homology arms beyond this size only result in marginal increases of transgene integration (Byrne et al. 2015; Zhang et al. 2017). In addition, the repair template also includes a floxed zeocin cassette to facilitate positive selection of gene edited cells (Mulsant et al. 1988; Seth et al. 2008). Zeocin-mediated enrichment of corrected cells is important due to the low efficiency of the HDR process.

Importantly, as scientists' understanding of the DNA damage response (DDR) continues to evolve so too does the manner in which DNA repair pathways are exploited to facilitate the integration of genetic material. Recent investigations have used NHEJ DNA repair pathways to introduce genetic material (Maresca et al. 2013; Suzuki et al. 2016). This strategy is reliant upon genomic target sites of the endonuclease TALEN or CRISPR being encoded in reverse orientation directly adjacent to the transgene for which integration is desirable. The resultant

in-situ cleavage of genome and exogenous repair template, facilitates the integration of the transgene independently of the HDR pathway (Suzuki et al. 2016). Importantly, the exogenous repair template developed here in our investigations was designed so components were flanked with endonuclease restriction sites. Thus it could be easily customised to facilitate exploration of such strategies with dystrophin cDNA.

Finally, Opt-HuMD1 (SO) was successfully cloned into the pUC57-hINT1 repair template to enable HDR investigations to be undertaken in HEK293T and DMD patient myoblasts. This became the primary transgene of interest owing to its small size which was anticipated to provide the most successful outcomes. This stated the exogenous repair template in this investigation was designed to enable the sub-cloning and subsequent delivery of the 3 dystrophin cDNA variants discussed and potentially have compatibility to emerging constructs.

5.4 Conclusions

A novel exogenous repair template was designed with restriction sites enabling the sub-cloning of MD1, $\Delta 45-55$ and full length sequence optimised dystrophin cDNA exonic constructions. The design was founded upon the demonstration that sequence optimisation enhanced recombinant dystrophin protein expression. The repair template was designed with 1Kb arms of homology isogenic to sequences upstream and downstream of the CRISPR MIT guide RNA targets identified within human DMD intron 1. Furthermore, the repair template incorporates a floxed zeocin expression cassette to facilitate positive selection of gene edited cell populations during HDR investigations. The Opt-MD1 (SO) exon 6-79 sequence was successfully cloned into the exogenous repair template. The resultant pUC57-hINT1-MD1-RT plasmid affords an exogenous repair template for HDR-mediated integration studies with the use of CRISPR/Cas9; enabling the feasibility of the gene editing therapeutic strategy presented to be assessed.

Chapter 6 An Examination of CRISPR-mediated Microdystrophin Knock-in Downstream of Full-length Endogenous DMD Promoters.

6.1 Introduction:

Current treatment of DMD relies upon the use of corticosteroids, which are inadequate, treating only the palliative symptoms of the disease rather than the underlying genetic aetiology (Bushby et al. 2010a; Mercuri & Muntoni 2013; NICE 2016). Furthermore, some therapeutic strategies under development addressing the genetic causation are largely mutation specific. These aim to convert an out-of-frame deletion to an in-frame deletion, or induce a stop codon read-through, thereby transitioning DMD to a BMD genotype (Monaco et al. 1988). Such therapeutic approaches aim to slow disease progression rather than provide complete amelioration. This is exemplified by antisense oligonucleotide (AON) exon skipping (Aartsma-Rus et al. 2017) and nonsense mutation read-through platforms such as Ataluren (McDonald et al. 2017). Mutation specific strategies for the treatment of DMD remain broadly challenging to translate, with EU medical approval requiring safety and efficacy testing of the individual compounds used for the treatment of specific mutation types; e.g. AON therapies for 45, 51 and 53 would all require individual approval (Aartsma-Rus et al. 2017). Moreover, even with all AON avenues maximally developed, including multi-exon skipping this approach would be applicable for 90% of patients, leaving 10% without treatment (Verhaart & Aartsma-Rus 2012).

Although universal strategies are under development such as microdystrophin 1 (MD1) gene augmentation (Mendell et al. 2010; Le Guiner et al. 2017), utrophin upregulation (Ricotti et al. 2016) and myostatin inhibition (Bhattacharya et al. 2017); the main limitation of these approaches is the transience of their expression and the requirement of repeated administration. In the case of AAV-MD1 delivery, this carries the inherent risk of an adverse immunogenic response to the AAV capsid (Mendell et al. 2010). In contrast, SMTC1100 (Ezutromid) and myostatin antibodies although well tolerated, have variable serum concentrations on administration and limited efficacy (Ricotti et al. 2016; Wagner et al. 2008), challenges which are currently being addressed with improved formulations (Summit Therapeutics Plc 2016; Bhattacharya et al. 2017). Nonetheless, even if successful, such therapies would likely require life-long adherence.

Gene editing offers an attractive alternative strategy, providing the permanent correction of a range of DMD mutations, and preventing the requirement for repeated administration. A plethora of studies have been undertaken using NHEJ-mediated deletions to promote for

example the excision of exon 51 (Ousterout et al. 2015) and exons 45-55 (Kabadi et al. 2014; Maggio et al. 2016). The studies described in the current thesis chapter aim to introduce a microdystrophin cDNA exon MD1 downstream of the full-length endogenous dystrophin promoters and exon1 elements, an approach that if successful would have applicability to 95% of DMD patients.

6.2 Chapter Aims:

Using both the highly efficacious guide 1 targeted to intron 1 of the *DMD* gene and the exogenous pUC57- hINT1- MD1 repair template (MD1 RT) constructed to encode MD1 Δ Ex1 cDNA and a downstream zeocin selection cassette; this chapter aimed to assess the feasibility of a HDR-mediated MD1 Δ Ex1 integration at the 5' end of the DMD gene, in both HEK293T and Δ 45-55 immortalised patient myoblast cells. Initially, the susceptibility of HEK293T and Δ 45-55 immortalised patient myoblasts to zeocin was analysed by the generation of a kill-curve in which cells were exposed to a defined concentration gradient of zeocin for up to 3 weeks. Concentrations of zeocin identified to cause cell death in untreated cultures will then be applied to cells 48 hours following co-transfection of guide 1 and the MD1 RT at 1:1 and 2:1 ratios.

Cultures co-transfected with guide 1 and the MD1 repair template were assessed initially at 48 hours post transient transfection for DNA integration. Polyclonal cultures generated following exposure to zeocin were then assessed for DNA integration, mRNA and protein expression of microdystrophin, driven by the endogenous Dp427m promoter. Individual clones were then isolated and expanded. The resultant monoclonal cultures were then subject to the same characterisation. It was speculated an in vitro correction of this nature would showcase the near universal therapeutic potential of this approach, highlighting the use of MD1 in gene editing to ameliorate the effects of a diverse range of DMD mutations types. Furthermore, the template has been customised to enable the exploration of Δ 45-55 and full length dystrophin transgenes, enabling expansion if this approach proved successful.

6.3 Results

6.3.1 Determination of Zeocin Sensitivity of HEK293T and Immortalised DMD Patient Myoblasts

As HDR is considered to occur with low efficiency (Heyer et al. 2010), it was speculated that positive selection would be crucial to the identification of genetically modified cells and the subsequent success of this study. Zeocin antibiotic was resolved upon as a positive selection reagent, as puromycin and neomycin resistance cassettes were integrated into patient myoblasts as a by-product of the immortalisation process (Mamchaoui et al. 2011). Thus

zeocin antibiotic would be suitable for use in integration experiments undertaken in both HEK293T and immortalised Δ 45-52 patient myoblasts.

In order to determine whether HEK293T and Δ 45-52 Immortalised patient myoblasts were sensitive to zeocin toxicity, they were subjected to a 50-1000 μ g/ml dose range of the antibiotic. HEK293T cells were seeded at a density of 1.7×10^5 cells in a 6 well plate estimated to provide an estimated 25% confluency in 24 hours, as recommended by the manufacturer's protocols (Invitrogen 2017). In contrast, 8.3×10^4 immortalised patient myoblasts were seeded in a 6 well plate, a cell number estimated to provide 35% confluency in 24 hours. It was decided to seed patient myoblasts at standard cell number due to their poor viability in sparse culturing conditions. Importantly, cells were seeded in growth media supplemented with incremental doses of zeocin. Notably, this dose escalation was performed in duplicate for both cell lines. Furthermore to maintain selection, 1ml of growth media was removed and substituted with 1ml of fresh growth media supplemented with the appropriate volume of zeocin antibiotic every 72 hours.

Once zeocin selection was commenced, phase contrast images were taken of each dose at 3, 7, 10, 14 and 21 day time points. Interestingly, on exposure to zeocin, dying cells do not only detach but exhibit irregular morphologies, including: thin spindle-like projections or appendages, cellular elongation and a thin squamous appearance (Invitrogen 2017). Microscopy imaging allowed visual assessment of these morphological changes and cell viability in HEK293T (Figure 6-1) and Δ 45-55 Immortalised Patient Myoblasts (Figure 6-2).

In HEK293T cultures, there was an estimated 60-70% reduction in viable cells by day 3, apparent in 50 μ g - 600 μ g/ml doses of zeocin antibiotic relative to the untreated control. This affect was heightened at higher doses of 800 μ g

/ml and 1000 μ g/ml where the reduction was closer to 80-90%. By days 7 and 10, >95% cell death was observed across all doses, with few cell clusters present at doses below 400 μ g. Strikingly, by day 7 localised patches of cell growth formation was apparent and cellular detachment started to occur in the mock condition likely attributable to the over-confluency and continued proliferation of the cells, serving to emphasise the result observed. Imaging was halted by day 14, as >95% cell death continued to be present across all zeocin doses.

In contrast, at day 3 there was 20-30% subtle decline in cell viability between 50 μ g-600 μ g/ml in patient myoblasts relative to the untreated mock. Again this affect was heightened in 800 μ g/ml and 1000 μ g/ml doses, following a similar trend to HEK293T, with an estimated 50-60% reduction. By day 7, the untreated patient myoblasts started to undergo differentiation,

due to continued proliferation and myoblasts forming contacts with one another. Interestingly, the number of cells across 50µg-600µg/ml doses at days 7 and 10 appeared visually comparable to the number observed at day 3, which due to the increase of growth in the untreated sample gave an apparent 50-60% reduction in cell viability, at this stage. In contrast a more marked decline in cell viability was apparent in 800µg/ml and 1000µg/ml conditions, with an estimated 70-80% reduction at day 7 and a 90% reduction at day 10. By day 14, noticeable cell death was apparent across all doses: with cell viability reduced by 70% in 50µg-100µg/ml, 80% in 200µg-600µg/ml and 90% in 800µg/ml and 1000µg/ml doses. This trend continued on day 21 with viability reduced by 80% in 50µg-100µg/ml, 90% in 200µg- 600µg/ml and >95% in 800µg/ml and 1000µg/ml doses.

Of note, an MTT cell viability assessment would have been desirable to provide more accurate quantification of cell death; however, the requirement to continuously expose cells to zeocin for up to 3 weeks during selection precluded the use of this assay. Thus visual examinations were used to determine cell death. Nonetheless, this investigation demonstrated that both HEK293T and Δ45-55 Immortalised patient myoblasts show susceptibility to zeocin antibiotic; promoting the use of this positive selection reagent in the identification of stably transfected cell lines.

On the basis of this assessment minimal zeocin doses of 600µg/ml and 800µg/ml were selected for use with HEK293T cell lines, as by day 7 the lowest dose where no cell clusters were observed was 600µg/ml of zeocin antibiotic. In contrast, minimal zeocin doses of 800µg/ml and 1000µg/ml will be used for Δ45-55 patient myoblast, as these doses are required to promote almost complete cell death following 21 days of exposure. These doses will be used, following transient transfection to select for cells with stable integration of the exogenous repair template.

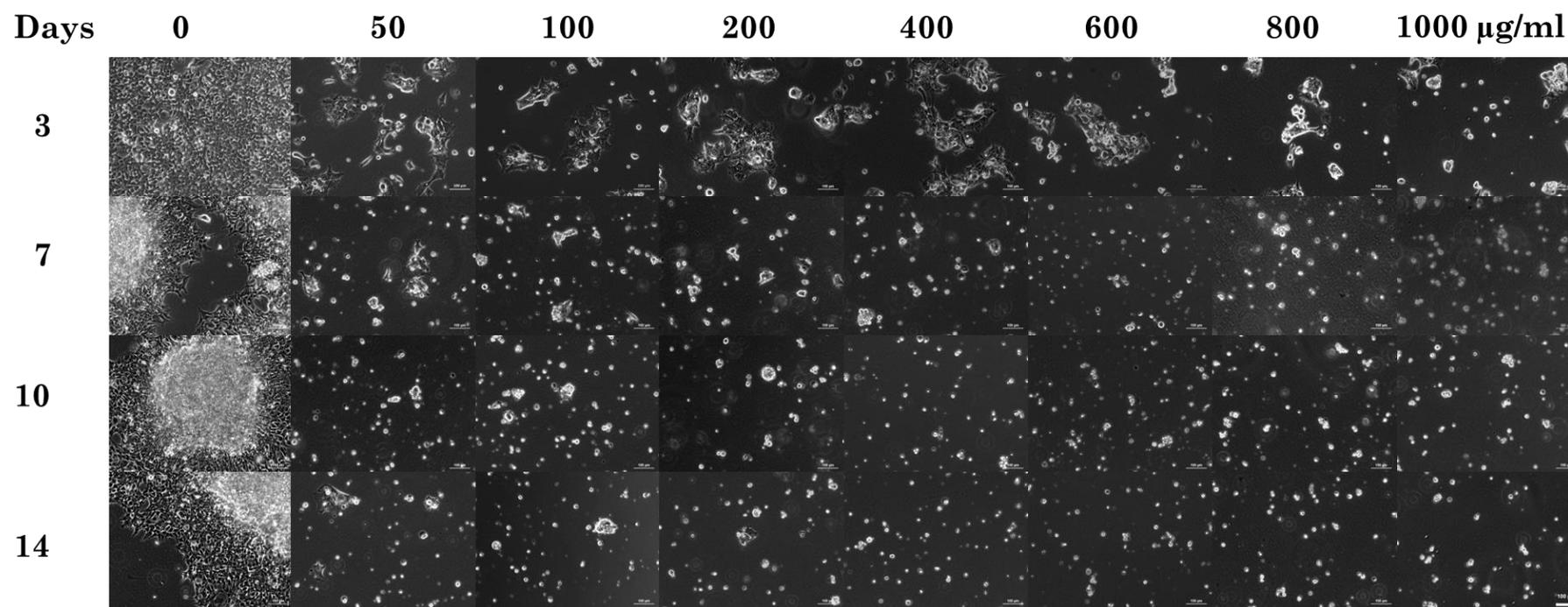


Figure 6-1 Determination of non-transfected HEK293T cells viability when cultured in varying concentrations of Zeocin antibiotic:

Following passage, HEK293T cells were diluted in standard 10% DMEM growth media to provide a cell count of 1.7×10^5 per 2ml, providing the approximate 25% confluency 24 hours post seeding in a six well plate recommended by the zeocin selection protocol. On seeding, the 2ml of cell suspension was supplemented with varying volumes of zeocin (stock concentration $100\mu\text{g}/\mu\text{l}$), to provide the defined antibiotic concentrations recommended in production of a kill curve. The concentration of zeocin was replenished every 72hours, by replacing 1ml of the existing 10% DMEM growth media with a fresh 1ml aliquot, supplemented with the necessary volume of zeocin. Abnormal cell morphology and cell death was then captured by phase contrast microscopy imaging at day 3, 7, 10 and 14 at 100x magnifications using the Zeiss Microscope. The zeocin was demonstrated to be effective, with estimates of cell death in excess of 90% observed for all doses following 3 days of exposure.

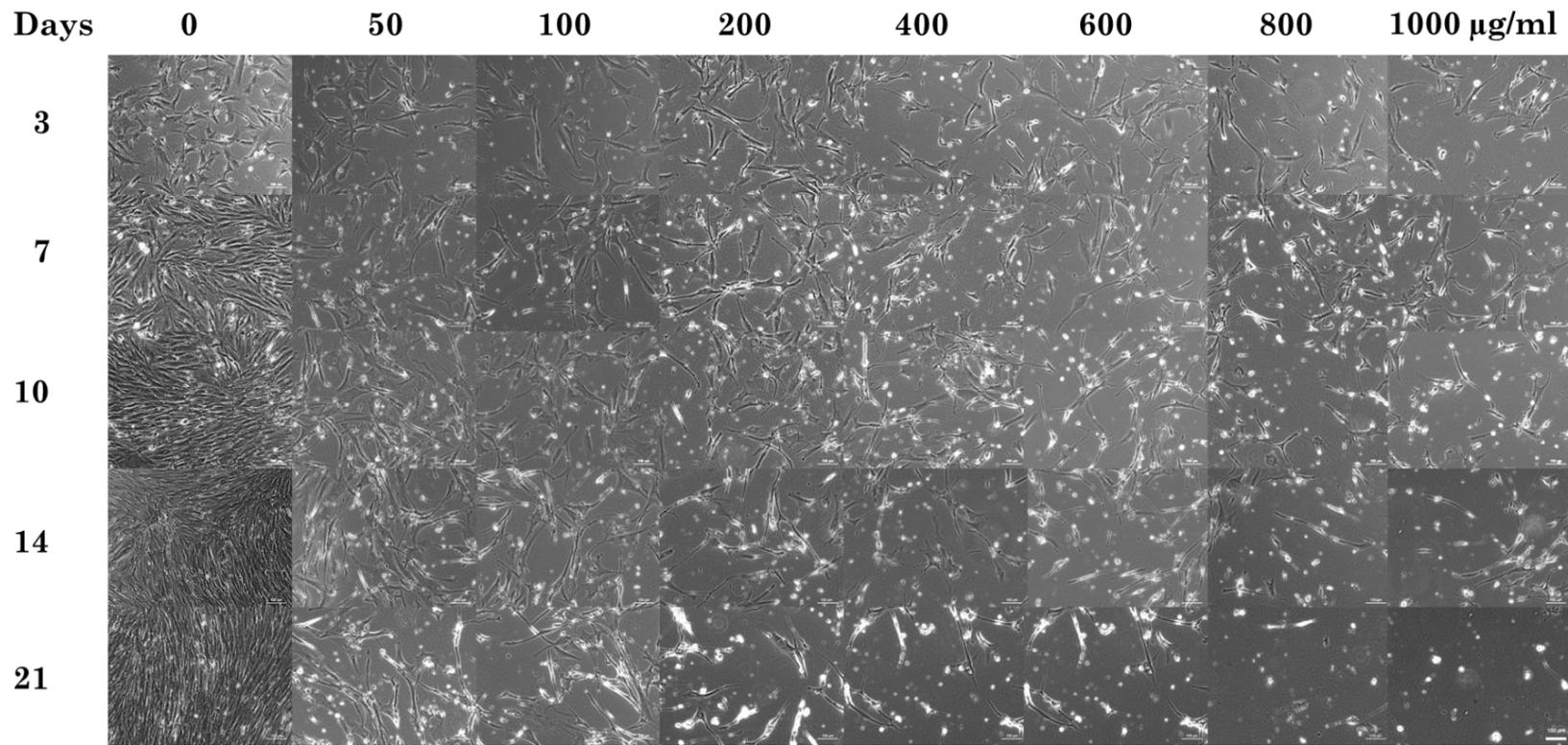


Figure 6-2 Determination of immortalised $\Delta 45-52$ DMD myoblast cells viability when cultured in varying concentrations of Zeocin antibiotic:

Following passage immortalised $\Delta 45-52$ patient myoblasts were diluted in standard myoblast growth media to provide a cell count of 8.4×10^4 per 2ml, providing the approximate 35% confluency 24 hours post seeding in a six well plate, a little in excess of that recommended by the zeocin selection protocol. On seeding, the 2ml of cell suspension was supplemented with varying volumes of zeocin (stock concentration $100\mu\text{g}/\mu\text{l}$, to provide the defined antibiotic concentrations recommended in production of a kill curve. The concentration of zeocin was replenished every 72hours, by replacing 1ml of the existing myoblast growth media with a fresh 1ml aliquot, supplemented with the necessary volume of zeocin. Abnormal cell morphology and cell death was then captured by phase contrast microscopy imaging at day 3, 7, 10 and 14 at 100x magnifications using the Zeiss Microscope. The zeocin was demonstrated to be effective at the higher doses, with estimates of cell death in excess of 90% observed at 21 days.

6.3.2 Transient Transfection of HEK293T Cells with Cas9, Guide RNA and Repair Template Plasmids and Polyclonal Enrichment using Zeocin Selection:

6.3.3 Dual Transfection Experimental Design

The demonstration that the CBh promoter within the all-in-one Px458 plasmid could drive both Cas9 and GFP expression prompted a dual transfection experimental design to be explored in the attainment of HDR-mediated integration of MD1 Δ Ex1. Utilising Px458-Guide 1 (Px458-G1, in figures) to introduce a double stranded break (DSB) and pUC57- hINT1- MD1 repair template constructed to encode MD1 Δ Ex1 cDNA and a downstream zeocin selection cassette, hitherto referred to as MD1 repair template (MD1 RT), to enable HDR repair. Notably, the use of the all-in-one Px458 plasmid had two inherent advantages: it enabled the use of GFP fluorescence as a proxy of transient transfection efficiency providing a visual indication of the delivery of gene-editing reagents and the assurance that both constituents of the Cas9 system, the guide and Cas9 nuclease encoding components, were co-delivered to each fluorescent cell.

To maximise the chance of attaining the successful integration of MD1 Δ Ex1 into DMD intron 1, two co- transfection conditions were employed, using a 1:1 and 2:1 ratio of Px458-Guide 1: MD1 RT plasmids, equating to 4 μ g and 6 μ g of total DNA per transfection respectively. These co-transfections were performed in triplicate alongside controls, including: a Viafect transfection reagent only mock, Px458-Guide 1 alone and MD1 RT alone to assert that any downstream observations were attributed to the co-delivery of the gene-editing and repair template plasmids. Notably, when delivered alone Px458-Guide 1 was delivered at either 2 μ g in the 1:1 or 4 μ g in the 2:1 conditions respectively. Whereas the MD1 repair template alone was routinely delivered at 2 μ g in both conditions.

6.3.3.1 Visual Assessment of Transient Transfection Efficiency:

Phase contrast and fluorescent microscopy images in the FITC channel were taken of HEK293T 48 hours post transfection. As anticipated Px458-Guide 1 and MD1 RT co-transfected, and Px458-Guide 1 alone samples, at both doses showed high levels of GFP expression (Figure 6-3). This is consistent with previous findings that transfection of 2 μ g and 4 μ g of Px458-Guide 1 resulted in 71.4% \pm 4.0 (N=3) and 92.7% \pm 3.0 (N=3) of GFP positive HEK293T cells respectively. Whilst promising, GFP fluorescence only provides transfection efficiency for the gene-editing Px458 plasmid in isolation; it does not provide information pertaining to the delivery of the exogenous repair template or indeed the co-localisation of the two constructs. However, given the similar sizes of the Px458 and MD1 repair plasmid of approximately 9.3kb and 9.9kb, the inference is transfection efficiency of the two plasmids would have been comparable and co-localisation would likely occur in a high proportion of cells.

Following this visual indication of successful transient transfection at 48 hours the treated HEK293T cells from the two transfection conditions were passaged and seeded at cell number of 1.7×10^5 cells per well into two six well plates. The media was supplemented on seeding with a volume of zeocin antibiotic in line with either a 600 $\mu\text{g}/\text{ml}$ or 800 $\mu\text{g}/\text{ml}$ dose; consistent with the conditions outlined by the zeocin sensitivity assay. This gave rise to a total of 4 conditions, in which both the 4 μg transfection using a 1:1 ratio of Px458-Guide 1: MD1 RT and the 6 μg using a 2:1 ratio of Px458-Guide 1: MD1 RT were exposed to 600 $\mu\text{g}/\text{ml}$ and 800 $\mu\text{g}/\text{ml}$ of zeocin. Conditions will be referred to throughout this chapter as 4 μg 600 μg , 4 μg 800 μg , 6 μg 600 μg and 6 μg 800 μg respectively. This coding system was selected to highlight the amount of plasmid (μg) transfected and the amount of zeocin per ml used for each condition.

6.3.3.2 DNA Extraction and PCR Amplification to Identify Successful Exogenous Repair Template Integration at DMD Intron 1 Site:

HEK293T cells in excess of the 1.7×10^5 cell count required for zeocin selection, attained from Mock, Px458-Guide1 and MD1 RT controls and co-transfected Px458-Guide 1 and MD1 RT samples were subject to Qiagen DNA extraction. This was undertaken for both 4 μg 1:1 and 6 μg 2:1 transfection conditions.

The resultant genomic DNA harvested was quantified and 250ng was amplified to assess for integration of the exogenous repair template. This was performed using a single-round 35-cycle PCR using either 'Outer LAH' or 'Outer RAH' primers, or a nested PCR in which 20 cycles of amplification were performed using outer primers and 1 μl of resultant product was applied to a second PCR using 'Inner LAH' or 'Inner RAH' primers. The resultant products were resolved on 1% agarose by gel electrophoresis. Both amplification processes utilised forward and reverse primers that spanned the genome-repair template junction, amplifying across the arms of homology included in the exogenous repair template respectively. This was performed for the left arm of homology, in which the forward primer was in the HEK293T genome and the reverse primer was within the repair template and the right arm of homology, in which the forward primer was in the repair template and the reverse was situated in the HEK293T genome; hitherto referred to as LAH and RAH PCRs respectively.

Following single round amplification of 250ng of DNA harvested from both 4 μg and 6 μg Px458-Guide 1 and MD1 RT co-transfected samples; a low intensity amplicon of 1291bps was resolved for the LAH PCR. This was consistent with the size anticipated. However unexpectedly, no amplicons were resolved for the RAH PCR in these samples. Finally, as anticipated, no amplicons were resolved for LAH or RAH PCRs in the Mock, Px458-Guide 1 or MD1 RT control conditions (Figure 6-4A and Figure 6-5A). Taken in isolation these results may

indicate that the LAH and RAH PCRs occur with varying efficiencies or that partial integration of the exogenous repair template has occurred.

To further assess the samples for low levels of exogenous repair template integration, nested LAH and RAH PCRs were undertaken with the outlook to increase the yield of resulting amplicons. In contrast to the single round amplification, nested LAH and RAH PCRs showed evidence for integration in Px458-Guide 1 and MD1 RT co-transfected samples, with 1139bp and 1171bp amplicons being resolved from inner PCRs respectively. All of these bands occurred with high intensity. An additional amplicon was also observed on resolution of nested LAH amplicons, estimated to be 1500bps in size; this is speculated to be due to mis-priming of the primers. As anticipated, no amplicons were resolved for LAH or RAH PCRs in Mock and Px458-Guide 1 alone samples. However, counter-intuitively a faint amplicon corresponding to the LAH PCR 1139bp amplicon was present in MD1 RT alone sample on the 6 μ g gel; furthermore faint amplicons corresponding to 1171bp RAH PCR amplicon were present in MD1 RT alone samples, in both 4 μ g and 6 μ g gels (Figure 6-4B and Figure 6-5B). Taken in combination these results indicate that 48 hours post transient transfection, low levels of exogenous repair template integration has occurred in all triplicate repeats of Px458-Guide 1 and MD1 RT co-transfections with both 4 μ g and 6 μ g conditions. Additionally, unexpectedly low levels of random integration may also have occurred following transient transfection of 2 μ g MD1 RT in isolation. This form of nuclease independent integration has previously been reported for the albumin safe-harbour locus, thus it was speculated that intron 1 of the *DMD* gene may have an open chromatin structure permitting this form of integration (Barzel et al. 2014).

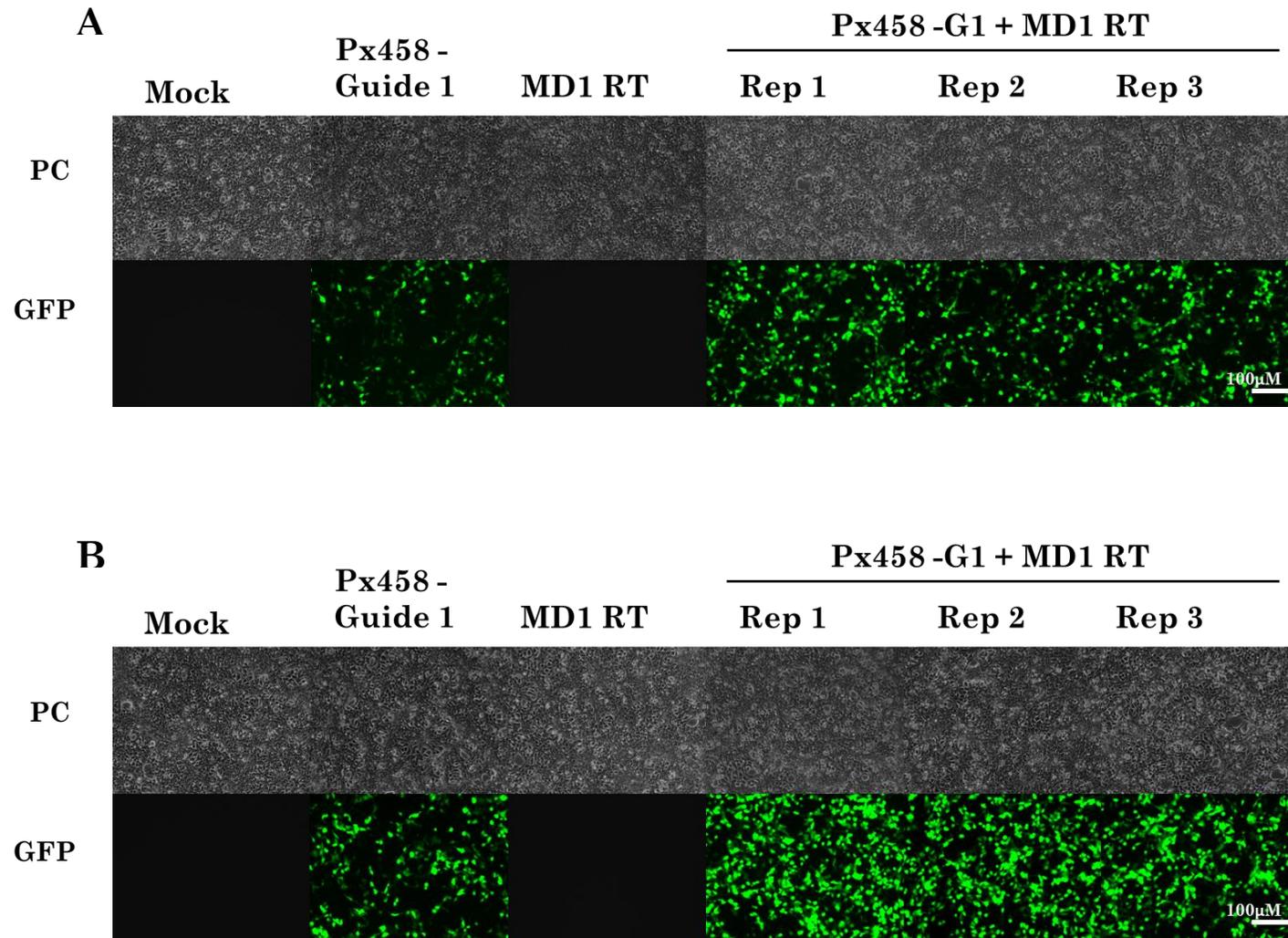


Figure 6-3 Microscopy Images of HEK293T Cells 48 hours post Co-transfection of Px458-Guide 1 at a 2µg or 4µg dose, with 2µg MD1 repair template, alongside Untreated, Px458-Guide 1 or MD1 RT alone.

(A) Phase contrast (PC) and fluorescent images in the FITC channel (GFP) of HEK293T transfected with 2µg of gene editing GFP-plasmid. Mock (M) = 25µl of Viasfect transfection reagent alone; Px458-Guide 1 = 2µg of Px458 alone; MD1-RT = 2µg of MD1-RT alone; Px458-Guide 1 + MD1 RT = co-delivery of 2µg Px458-Guide 1 and MD1 RT in a 1:1 ratio across three technical replicates (Rep 1-3). **(B)** Phase contrast (PC) and fluorescent images in the FITC channel (GFP) of HEK293T transfected with 4µg of gene editing GFP-plasmid., same as **(A)** except 4µg of Px458-Guide 1 was used in delivery alone and co-delivery alongside 2µg MD1 RT constituting a 2:1 ratio.

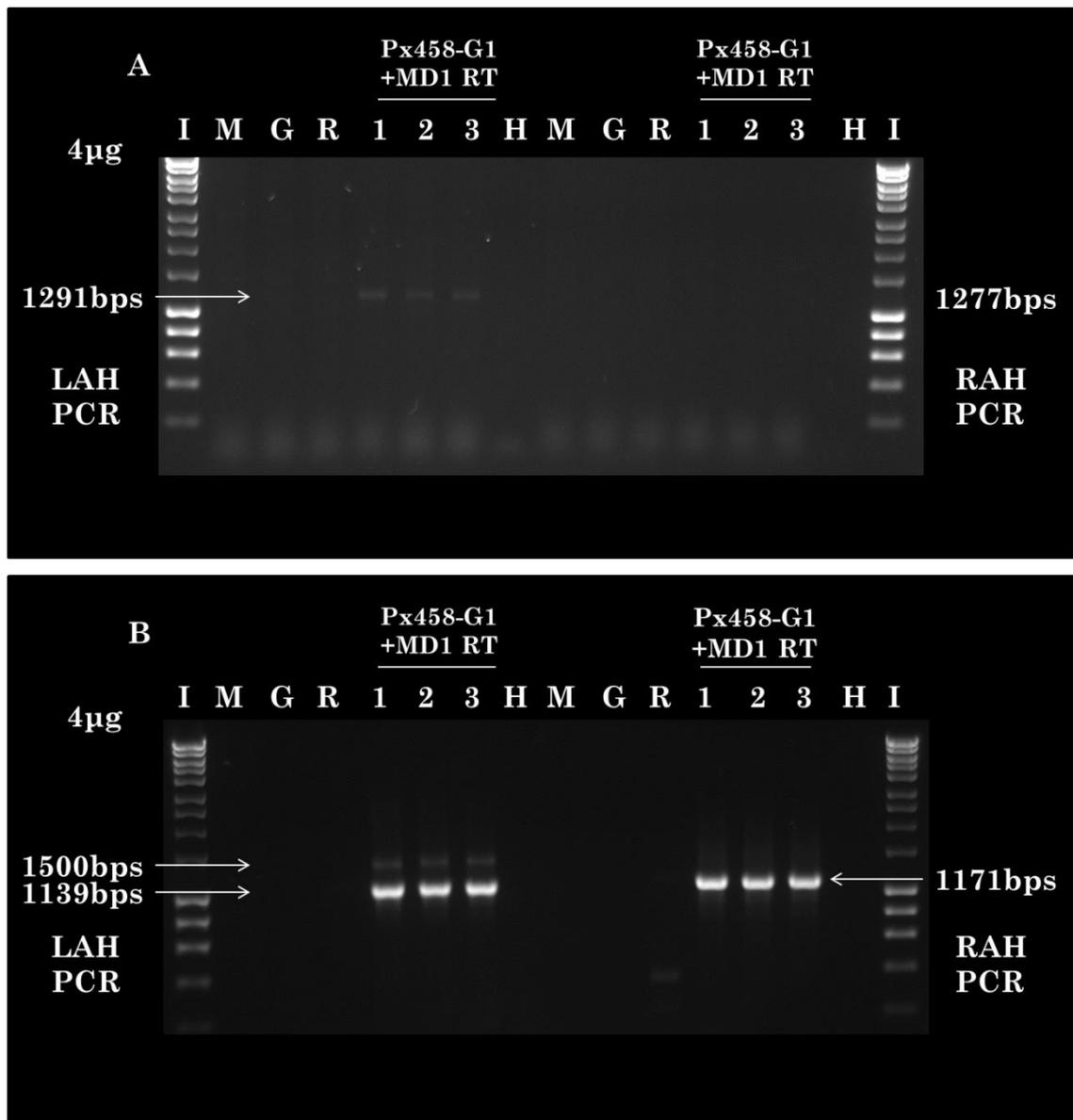


Figure 6-4 An examination of repair template integration into DMD intron 1, using PCR amplification; spanning from the endogenous genomic sequence, across the left or right arms of homology into the repair template. Performed for the 4 μ g (Px458-Guide 1 1:1 MD1-RT) condition.

A) Gel electrophoresis to resolve PCR products resulting from a single round of amplification performed on genomic DNA harvested from the 4 μ g (Px458-Guide 1 1:1 MD1-RT) condition, 48 hours post transient transfection. PCRs were performed and resolved spanning either the left arm of homology (LAH) or the right arm of homology (RAH). Samples included: Mock (M) – cells incubated with Viafect transfection reagent alone, Px458-Guide 1 (G) - 2 μ g of Px458-Guide 1 transfected alone, MD1 RT (R) - 2 μ g of MD1-RT transfected alone and triplicate co-transfection of 2 μ g Px458-Guide 1 and 2 μ g MD1-RT, denoted by 1-3 and a no-template water control (H). In all single round conditions, 250ng of genomic DNA was subjected to a 30 cycle GoTaq G2 Flexi PCR, prior to being resolved on a 1% agarose gel alongside hyperladder I. B) A Nested PCR performed on the same samples as A) except 1 μ l of a single 20 cycle outer PCR performed on 250ng of genomic DNA, was added to a second 30 cycle PCR to improve, the amplicon yield.

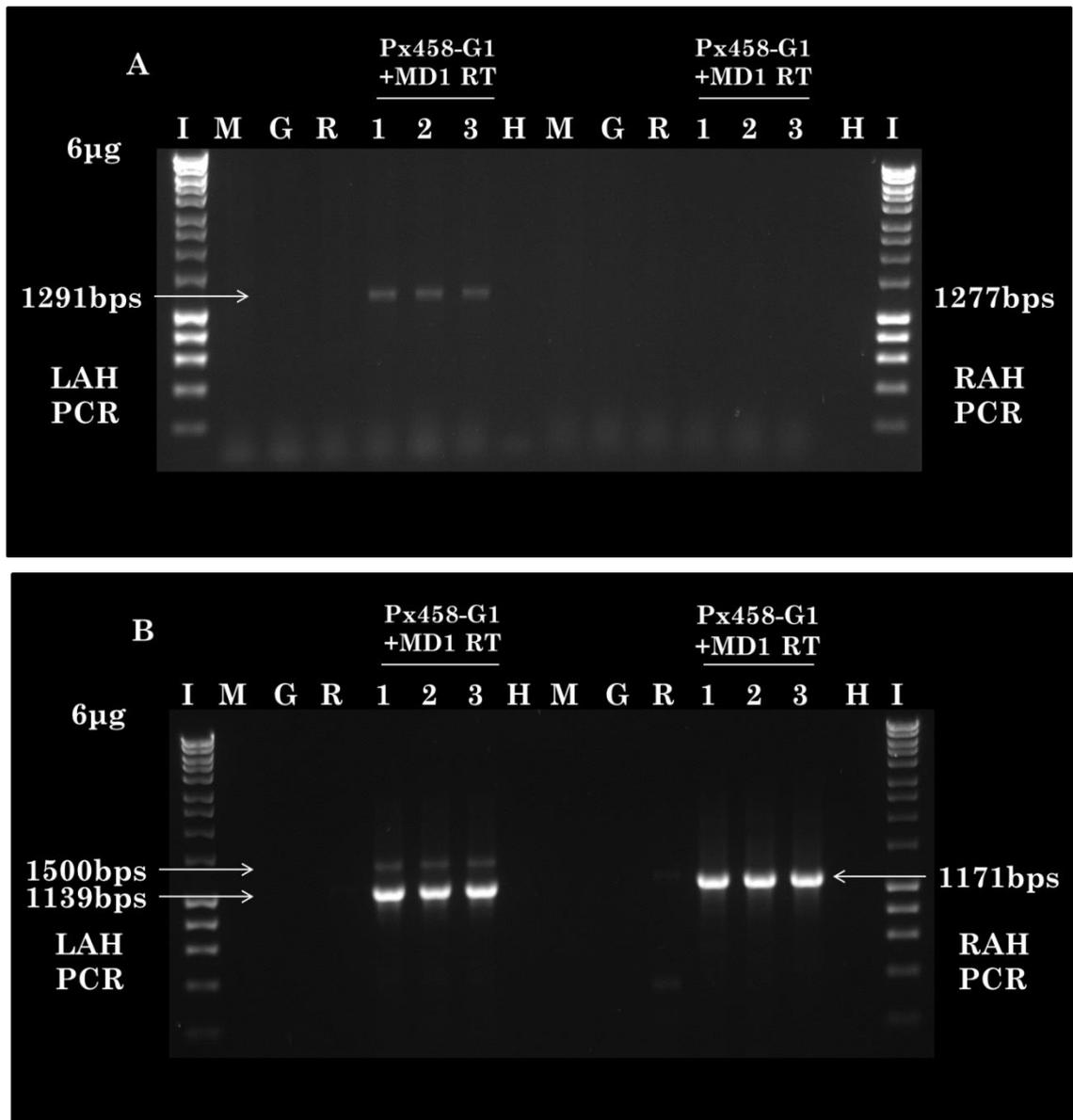


Figure 6-5 An examination of repair template integration into DMD intron 1, using PCR amplification; spanning from the endogenous genomic sequence, across the left or right arms of homology into the repair template. Performed for the 6 μ g (Px458-Guide 1 2:1 MD1-RT) condition.

A) Gel electrophoresis to resolve PCR products resulting from a single round of amplification performed on genomic DNA harvested from the 6 μ g (Px458-Guide 1 2:1 MD1-RT) condition, 48 hours post transient transfection. PCRs were performed and resolved spanning either the left arm of homology (LAH) or the right arm of homology (RAH). Samples included: Mock (M) – cells incubated with Viafect transfection reagent alone, Px458-Guide 1 (G) - 4 μ g of Px458-Guide 1 transfected alone, MD1 RT (R) - 2 μ g of MD1-RT transfected alone and triplicate co-transfection of 4 μ g Px458-Guide 1 and 2 μ g MD1-RT, denoted by 1-3 and a no-template water control (H). In all single round conditions, 250ng of genomic DNA was subjected to a 35 cycle GoTaq G2 Flexi PCR, prior to being resolved on a 1% agarose gel alongside hyperladder I. B) A Nested PCR performed on the same samples as A) except 1 μ l of a single 20 cycle outer PCR performed on 250ng of genomic DNA, was added to a second 30 cycle PCR to improve, the amplicon yield.

6.3.4 Positive Selection: The Enrichment of Stably Transfected HEK293T-MD1 Cells

The indication that exogenous repair template integration may have occurred at low levels, producing a heterogeneous HEK293T culture, prompted continued zeocin selection of HEK293T treated cells. In an analogous fashion to the zeocin sensitivity assay, 1ml of media was removed every 72 hours and replenished with 1ml fresh media supplemented with either 600µg/ml or 800µg/ml of zeocin. Furthermore, phase contrast and fluorescent microscopy imaging undertaken in the FITC (GFP) Channel were taken at 3, 7, 14, 18, 24 and 27 day time points to examine the selection process and cell viability. This was undertaken for all 4 conditions previously outlined, including: The 4µg 1:1 Px458-Guide 1 and MD1 RT conditions: 4µg 600µg (Figure 6-6) and 4µg 800µg (Figure 6-7) and the 6µg 2:1 Px458 Guide 1 and MD1 RT conditions: 6µg 600µg (Figure 6-8) and 6µg 800µg (Figure 6-9) respectively. The general trends of this positive selection process, rather than individual conditions will be described for the purposes of clarity.

Initial phase contrast images captured at day 3, showed low cell number across all of the conditions with viable cells estimated to constitute 10-30% confluency for all treatment conditions: Mock, Px458-Guide 1, MD1 RT alone and co-transfected Px458-Guide 1 and MD1 RT triplicates in the 4µg conditions. Notably, cell viability was reduced in the 6µg conditions relative to the 4µg conditions at this stage, with 10% confluency estimated in 6µg 600µg/ml conditions and 5-10% for 6µg 800µg/ml generally observed across treatments. This could be attributable to cells being exposed to higher DNA and transfection reagent concentrations. This stated, clusters of cells were present in some Px458-Guide 1 and MD1 RT co-transfected replicates in the 6µg condition, as exemplified by 6µg 600µg/ml repetition 1 and 2 and 6µg 800µg/ml repetition 2 images. Furthermore, fluorescent images taken at this time point indicated a higher proportion of GFP positive cells on visual assessment in 6µg relative to 4µg conditions. This suggests higher transfection efficiency in 6µg conditions and potentially a greater chance for MD1 integration to occur.

By Day 7, Mock and Px458-Guide 1 alone treatments showed extensive cell death across all 4 conditions, comparable to the >95% death observed in the zeocin sensitivity assay. In contrast samples transfected with MD1 repair template alone or co-transfected with Px458-Guide 1 and MD1 repair template, showed an increase in cell number. This observation was more pronounced in 4µg and 6µg conditions under 600µg/ml zeocin selection, where cell number was visually estimated to double relative to day 3. Notably, 4µg and 6µg samples placed under 800µg/ml still showed evidence of surviving cell clusters but the numbers of cells surviving appeared more variable. This is apparent in Px458-Guide 1 and MD1 RT co-transfection

repetition 1 in the 4µg 800µg/ml condition whereby cell number had a marked increase relative to the MD1 RT alone and other co-transfected repetitions. Further evidence of variability also apparent in Px458-Guide 1 and MD1 RT co-transfected repetition 2 of the 6µg 800µg/ml condition, where the cell number was reduced relative to repetitions 1 and 3 and MD1 repair template alone treatments. In addition, the number of GFP-positive cells at this stage appeared markedly reduced across all conditions when visually compared to the Day 3 time point. This is likely attributable to the effects of clonal dilution, and the subsequent expression of the Px458 plasmid being dissipated.

At day 14 and 18 cells appeared rounded and detached in mock and Px458-Guide 1 alone samples across all conditions. This indicated complete cell death had occurred and these treatments were not imaged after these time points. In contrast, in Px458-Guide 1 and MD1 RT co-transfected samples, clusters of cells examined had continued to proliferate and were able to fill a field of capture as observed in 4µg 600µg/ml, 6µg 600µg/ml and 6µg 800µg/ml conditions. The 4µg 800µg/ml condition appeared as a notable exception to this observation, with few small cell clusters being observed at day 14 and only co-transfected repetitions 1 and 3 appearing to fill approximately 60-70% of the field of capture at day 18. This was attributed to a combinatorial effect of lower anticipated exogenous repair template integration and the higher dose of zeocin selection. Interestingly, the MD1 repair template alone treatments were not anticipated to survive zeocin selection. This stated, the number of cells in MD1 RT alone treatment in the 4µg conditions, were visually estimated to be comparable to the Px458-guide 1 and MD1 RT co-transfected samples at day 14 and 18 time points. Moreover, evidence of cell growth in MD1 RT alone samples was present among the 6µg conditions, although a reduction in cell number relative to co-transfected repetitions was observed. Finally, by day 14, GFP expression was almost completely absent and thus images in the FITC channel were no longer captured.

By day 20 visible localised areas of cell growth had occurred in MD1 RT alone and co-transfected Px458-Guide 1 and MD1 RT samples across all conditions (not shown). Thus the cells within these treatments were passaged and seeded back into a 6 well plate under continued zeocin selection to avoid detachment and promote healthy monolayer growth. Mock and Px458-guide 1 alone samples were discarded at this time.

Post passage, all cells were then imaged at two further time points, days 24 and 27. At the day 24, the number of surviving cells in 4µg conditions appeared reduced across MD1 RT alone and co-transfected treatments relative to the 6µg conditions. In addition, it should be noted that one sample, the MD1 RT alone treatment in the 4µg 800µg/ml condition did not survive the

passage process. Finally, by day 27, all MD1 RT alone and co-transfected Px458-Guide 1 and MD1 treatments, with the exception of the single condition noted above attained over 70% confluency in a 6 well plate and appeared to be proliferating in a manner comparable to that expected of untreated HEK293T.

In summary, in Mock and Px458-guide 1 alone treatments, cells exhibited a rounded morphology and detachment from the plate after 14 days of selection consistent with cell death. Unexpectedly, HEK293T cells subject to MD1 RT alone treatments, with the exception of one sample in the 4 μ g 800 μ g/ml condition, appeared to survive the zeocin selection process. Finally, all Px458-Guide 1 and MD1 RT co-transfected triplicates, survived the zeocin selection process across all conditions, achieving greater than 70% confluency by the latest 27 day time point.

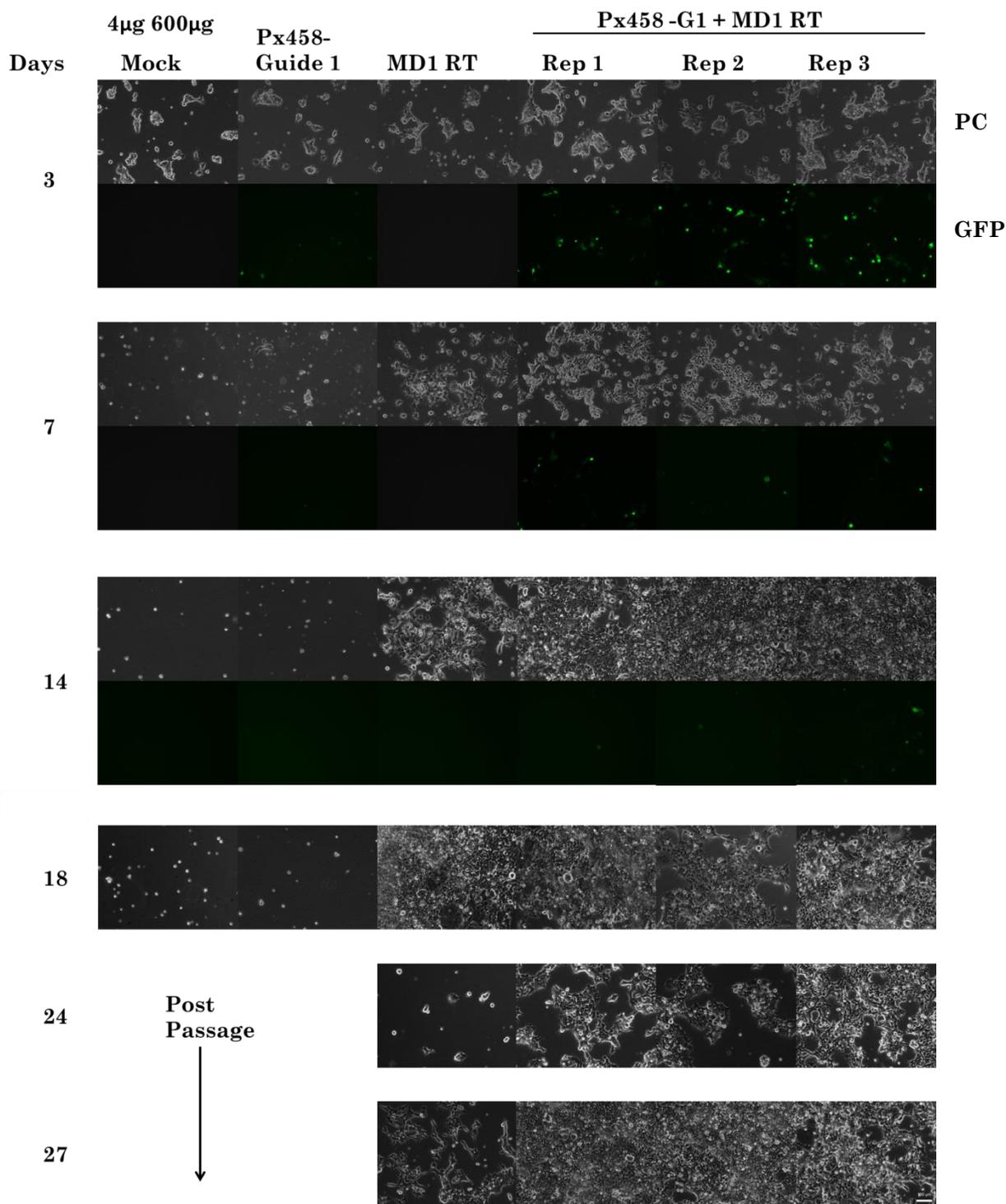


Figure 6-6 Microscopy images of HEK293T post 4µg (Px458-Guide 1 1:1 MD1-RT) co- transfection, alongside controls during the process of zeocin selection. At 48 hours post transfection, HEK293T cells were passaged and seeded at 1.7×10^5 cells per well and exposed to 600µg/ml of zeocin.

Samples included: Mock (M) – cells incubated with Viafect transfection reagent alone, Px458-Guide 1 (G) - 2µg of Px458-Guide 1 transfected alone, MD1 RT (R) - 2µg of MD1-RT transfected alone and triplicate co-transfection of 2µg Px458-Guide 1 and 2µg MD1-RT, denoted by1-3. A 1ml growth media aliquot was removed, and substituted with 1ml fresh growth media supplemented with 600µg/ml of Zeocin every 3 days to maintain selection. Cell death initially ensued, however by day 20, confluent HEK293T localised patches of cell growth were present and the cells were passaged and placed into a fresh six well plate. Phase Contrast and fluorescent images in the FITC channel (GFP) were taken at time points of 3, 7 and 14 days. Phase contrast microscopy was continued to 18, 24 and 27 days after the GFP expression had diminished.

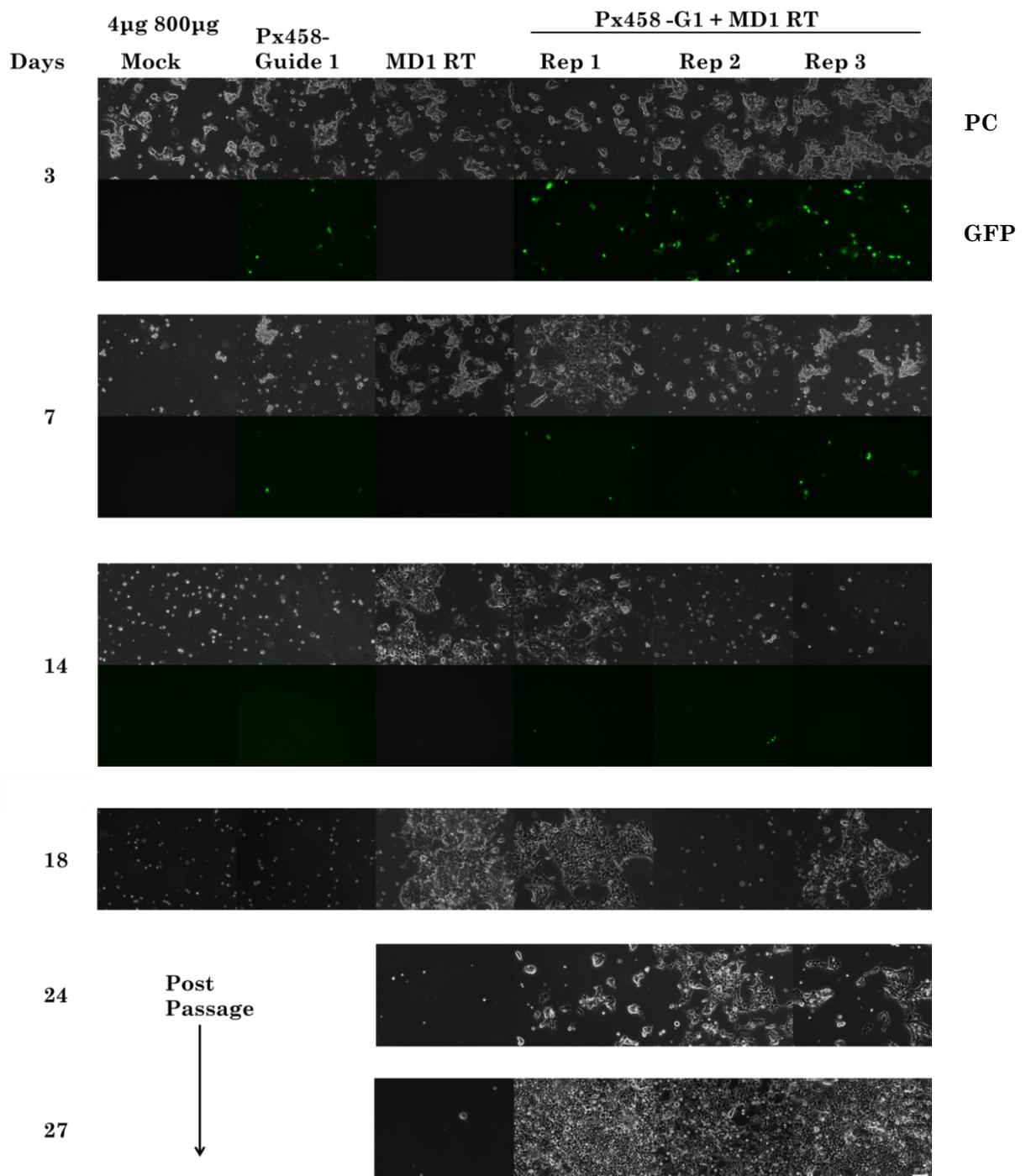


Figure 6-7 Microscopy images of HEK293T post 4μg (Px458-Guide 1 1:1 MD1-RT) co- transfection, alongside controls during the process of zeocin selection. At 48 hours post transfection, HEK293T cells were passaged and seeded at 1.7×10^5 cells per well and exposed to 800μg/ml of zeocin.

Samples included: Mock (M) – cells incubated with Viafect transfection reagent alone, Px458-Guide 1 (G) - 2μg of Px458-Guide 1 transfected alone, MD1 RT (R) - 2μg of MD1-RT transfected alone and triplicate co-transfection of 2μg Px458-Guide 1 and 2μg MD1-RT, denoted by 1-3. A 1ml growth media aliquot was removed, and substituted with 1ml fresh growth media supplemented with 800μg/ml of Zeocin every 3 days to maintain selection. Cell death initially ensued, however by day 20, confluent HEK293T localised patches of cell growth were present and the cells were passaged and placed into a fresh six well plate. Phase Contrast and fluorescent images in the FITC channel (GFP) were taken at time points of: 3, 7 and 14 days; phase contrast was continued to 18, 24 and 27 days after the GFP expression had diminished.

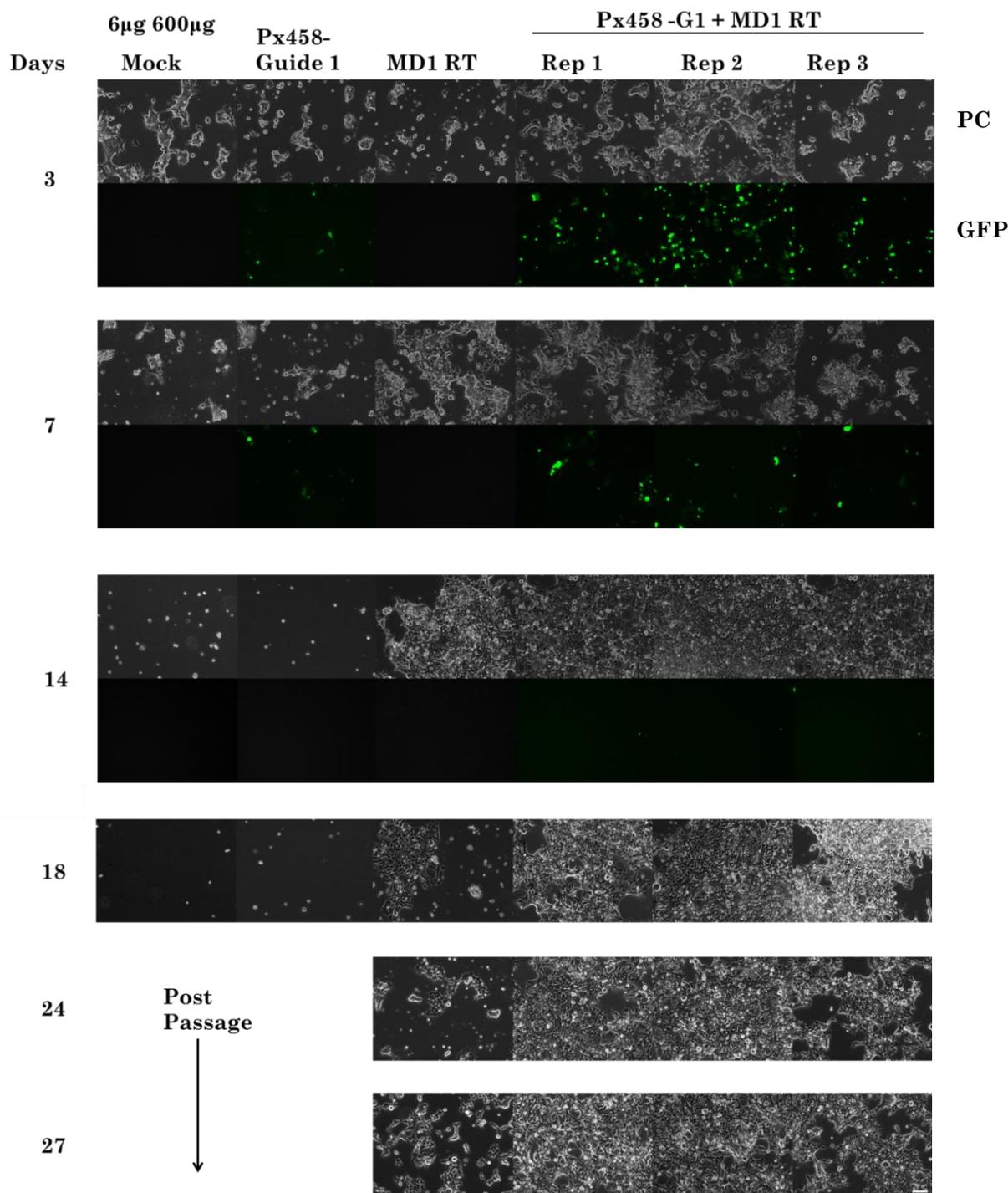


Figure 6-8 Microscopy images of HEK293T post 6µg (Px458-Guide 1 2:1 MD1-RT) co- transfection, alongside controls during the process of zeocin selection. At 48 hours post transfection, HEK293T cells were passaged and seeded at 1.7×10^5 cells and exposed to 600µg/ml of zeocin.

Samples included: Mock (M) – cells incubated with Viafect transfection reagent alone, Px458-Guide 1 (G) - 4µg of Px458-Guide 1 transfected alone, MD1 RT (R) - 2µg of MD1-RT transfected alone and triplicate co-transfection of 4µg Px458-Guide 1 and 2µg MD1-RT, denoted by 1-3. A 1ml growth media aliquot was removed, and substituted with 1ml fresh growth media supplemented with 600µg/ml of Zeocin every 3 days to maintain selection. Cell death initially ensued, however by day 20, confluent HEK293T localised patches of cell growth were present and the cells were passaged and placed into a fresh six well plate. Phase Contrast and fluorescent images in the FITC channel (GFP) were taken at time points of: 3, 7 and 14 days; phase contrast was continued to 18, 24 and 27 days after the GFP expression had diminished.

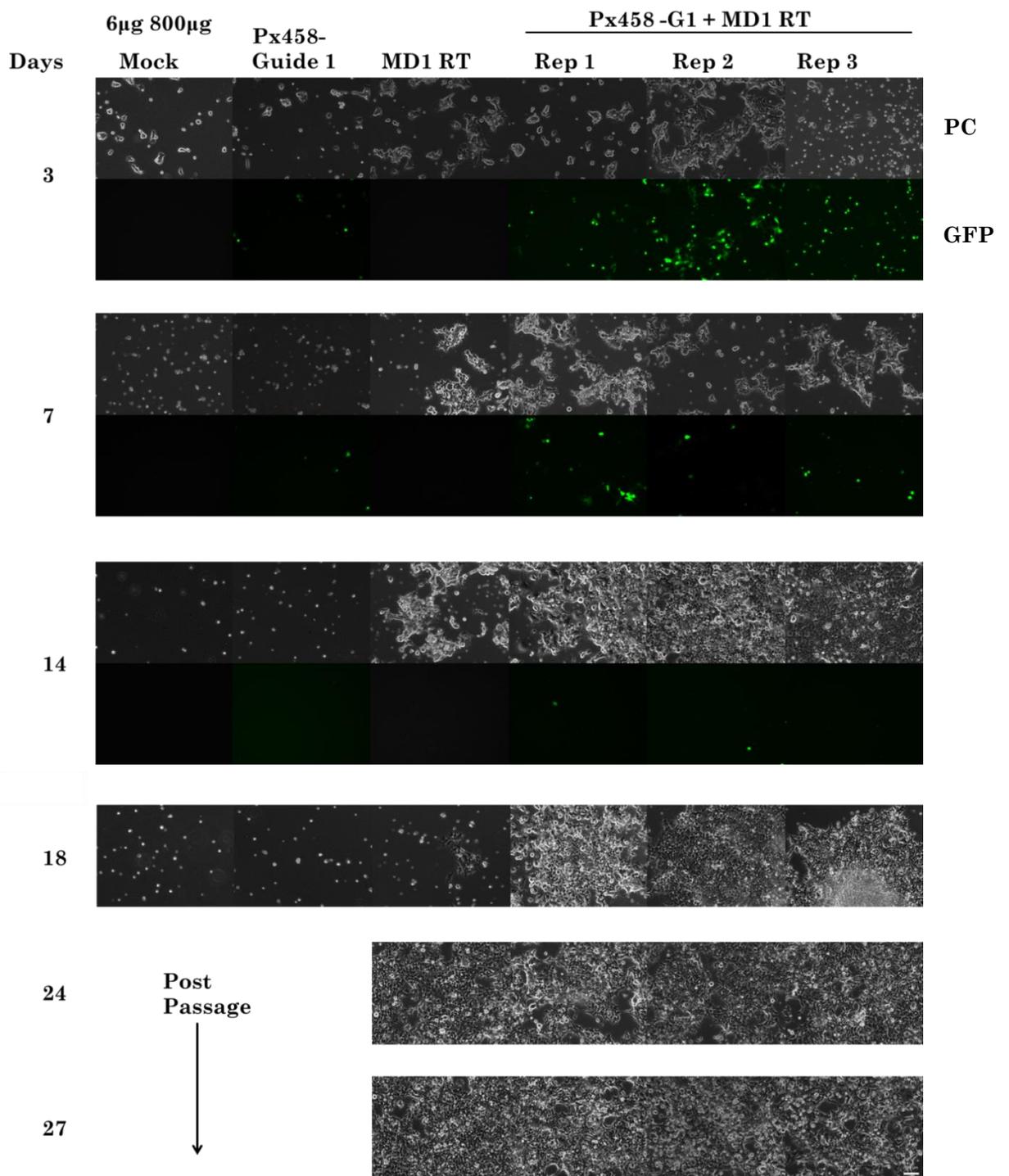


Figure 6-9 Microscopy images of HEK293T post 6µg (Px458-Guide 1 2:1 MD1-RT) co- transfection, alongside controls during the process of zeocin selection. At 48 hours post transfection, HEK293T cells were passaged and seeded at 1.7×10^5 cells and exposed to 800µg/ml of zeocin.

Samples included: Mock (M) – cells incubated with Viafect transfection reagent alone, Px458-Guide 1 (G) - 4µg of Px458-Guide 1 transfected alone, MD1 RT (R) - 2µg of MD1-RT transfected alone and triplicate co-transfection of 4µg Px458-Guide 1 and 2µg MD1-RT, denoted by 1-3. A 1ml growth media aliquot was removed, and substituted with 1ml fresh growth media supplemented with 800µg/ml of Zeocin every 3 days to maintain selection. Cell death initially ensued, however by day 20, confluent HEK293T localised patches of cell growth were present and the cells were passaged and placed into a fresh six well plate. Phase Contrast and fluorescent images in the FITC channel (GFP) were taken at time points of: 3, 7 and 14 days; phase contrast was continued to 18, 24 and 27 days after the GFP expression had diminished.

6.3.5 Molecular Characterisation of Zeocin-Enriched Polyclonal Cultures and Establishing Single Colonies Allowing for Clonal Expansion:

Once polyclonal cultures of zeocin selected HEK393T cells were established, they were expanded; where the contents of a single well were split into two wells, through two sequential passages, for each individual treatment. This culminated in four 6 well plates: comprising of MD1 RT alone and triplicate Px458-Guide 1 and MD1 RT co-transfections for each of the four conditions (4µg 600µg, 4µg 800µg, 6µg 600µg and 6µg 800µg).

For each condition, this expansion provided 3 plates for DNA, RNA and protein harvests respectively. This was to permit molecular characterisation of the polyclonal cultures, assessing endogenous repair template integration and MD1 transgene expression from the full length Dp427m dystrophin promoter. Additionally the cells seeded on the final plate were passaged and serially diluted to attain individual colonies for each condition.

6.3.5.1 Examination of Exogenous Repair Template Integration at DMD intron 1:

It was speculated that zeocin selection would enrich HEK293T populations where integration of the exogenous repair template had occurred. On this basis DNA was harvested from the expanded polyclonal culture and 250ng was amplified to assess for integration of the exogenous repair. This was performed using a single round PCR across the genome-repair template junction for both the LAH and RAH and products were resolved by gel electrophoresis, as described in section 6.3.3.2.

This was performed for all MD1 RT alone and co-transfected Px458-Guide 1 and MD1 treatments across the 4 conditions. Strikingly, for all Px458-Guide 1 and MD1 RT co-transfected samples, amplicons of 1291bp and 1277bps were resolved by gel electrophoresis for the LAH and RAH PCRs respectively. These amplicons are in line with the sizes anticipated, and consistent with the results obtained from the nested PCR undertaken at the 48 hour time point, indicate that integration of the exogenous repair template has occurred (Figure 6-10). Notably, as previously observed a single larger amplicon estimated to be 1500bps in size was apparent following resolution of LAH PCRs, likely attributable to mis-priming of the primers. Interestingly, no amplicons were resolved for MD1 RT alone treatments, subject to single round LAH or RAH PCRs. This finding in combination with the observation that MD1 RT alone transfected HEK293T cells survived zeocin selection, indicates either low level targeted integration at the *DMD* intron 1 locus or random integration of the MD1 RT has occurred.

Importantly, the resolution of LAH and RAH amplicons following a single round of amplification as opposed to the nested PCR required to enhance the amplicon signal at the transient 48 hour time point, suggests that the zeocin positive selection was successful. This indicates that the

HEK293T culture has been enriched for cells in which the specific integration of the exogenous repair template is present.

6.3.5.2 Examination of MD1 Transcription from the Endogenous Dp427m DMD

Promoter:

Once evidence was attained that indicated the exogenous repair template had been successfully integrated into DMD intron 1, focus was then shifted to examine whether expression of the MD1 Δ Ex1 transgene was driven by the endogenous Dp427m full-length dystrophin promoter. To assess this, expanded polyclonal cultures were harvested for total RNA and a reverse transcription PCR was undertaken on 500ng of purified RNA using random primers to produce cDNA.

The cDNA was then assessed for Dp427m- MD1 exon 2 splicing using a forward primer designed to span the Dp427m- MD1 exon-exon 2 junction and a reverse primer within exon 5. In this design an ~250bp amplicon would only be visualised following amplification and resolution by gel electrophoresis if correct splicing had occurred. Furthermore, due to the sequence optimisation of the MD1, primers design would detect Dp427m –MD1 exon1-exon 2 splicing, distinctly from any native dystrophin splicing that may occur, in the culture.

Interestingly, in all MD1 RT alone and co-transfected Px458-Guide 1 and MD1 RT treatments, a 250bp amplicon was resolved by gel electrophoresis of amplified products across all conditions (Figure 6-11). Indicating Dp427m-MD1 exon 2 splicing is occurring across all conditions. When considered in combination with the survival of MD1 RT alone transfected HEK293T cultures, following zeocin selection, it could suggest low levels of targeted integration of the repair template alone occurred in this treatment.

6.3.5.3 Examination of MD1 Protein Expression:

Provisional indication of mRNA expression prompted an examination of MD1 protein expression. In this instance 50 μ g total protein lysate, was subject to western blotting with the ManEx10-11c antibody. Crucially this would highlight that the MD1 mRNA transcribed by the endogenous promoter is functional, enabling MD1 protein synthesis.

The MD1 construct, encodes a 137kDa protein that can be visualised by western blotting (see Chapter 5). However in both MD1 RT alone and co-transfected Px458-Guide 1 and MD1 RT treatments, a truncated protein product, estimated to be 117kDa in size, was identified across all conditions (Figure 6-12A and Figure 6-13A). Notably, this band was not present in untreated HEK293T cell lysate, indicating this protein expression is likely a result of the treatment performed (Figure 6-12B and Figure 6-13B).

The resolution of truncated MD1 protein product was unexpected. It could be speculated to arise due to aberrant splicing being favoured over the Dp427m splicing observed, or an internal deletion within the integrated MD1 cDNA.



Figure 6-10 An examination of repair template integration into DMD intron 1, following polyclonal enrichment. This was performed by PCR amplification that spanned from the endogenous genomic sequence, across the left or right arms of homology into the repair template.

Gel electrophoresis to resolve PCR products resulting from a single round of amplification, performed on genomic DNA harvested from both 4µg (Px458-Guide 1 1:1 MD1-RT) Top and 6µg (Px458-Guide 1 2:1 MD1-RT) Bottom conditions, exposed to either 600µg/ml or 800µg/ml of zeocin for 28 days. PCRs were performed and resolved spanning either the left arm of homology (LAH) or the right arm of homology (RAH). Samples included: MD1 RT (R) - 2µg of MD1-RT transfected alone and triplicate co-transfection of either 2µg Px458-Guide 1 and 2µg of MD1 RT or 4µg Px458-Guide 1 and 2µg MD1-RT, denoted by 1-3 and a no-template water control (H). In all single round conditions, 250ng of genomic DNA was subjected to a 35 cycle GoTaq G2 Flexi PCR, prior to being resolved on a 1% agarose gel alongside hyperladder I. The findings indicate evidence for a positive integration of the MD1 exogenous repair template in DMD intron 1 in co-transfected Px458-Guide1 and MD1 RT at this stage.

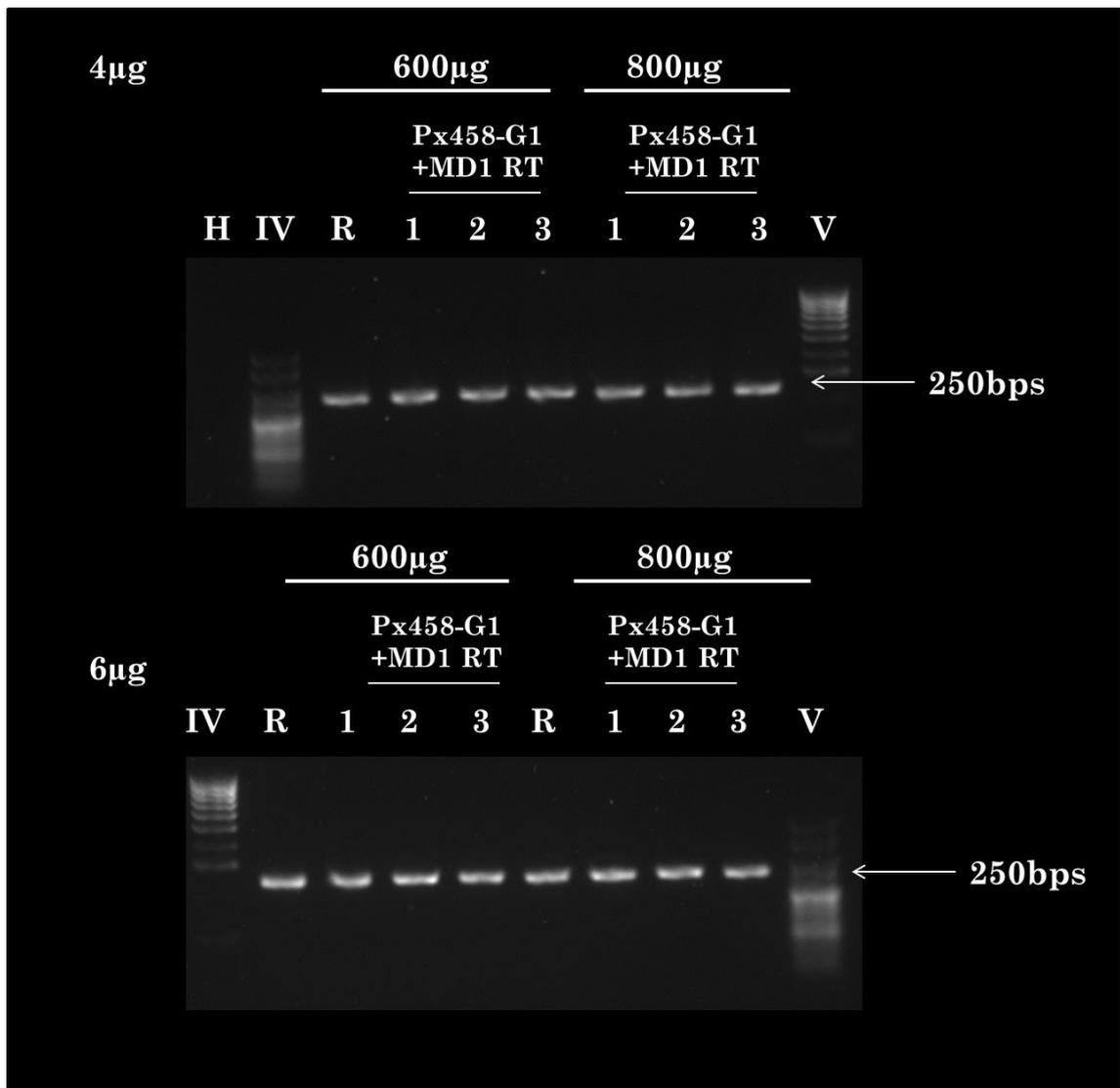


Figure 6-11 An examination of endogenous Dp427m promoter and exon 1 element splicing to exon 2 of MD1, undertaken on RNA harvested from polyclonal HEK293T cultures.

Gel electrophoresis to resolve PCR products resulting from a single round of amplification of cDNA. Importantly, cDNA was produced from 500ng of total mRNA lysate, and amplified with isoform specific primers, in which the forward primer targeted the Dp427m Exon 1 and MD1 Exon 2 splice junction. This was performed for both 4µg (Px458-Guide 1 1:1 MD1-RT) top and 6µg (Px458-Guide 1 2:1 MD1-RT) bottom conditions, for the MD1 RT (R)-2µg of MD1-RT transfected alone and triplicate co-transfection of either 2µg Px458-Guide 1 and 2µg of MD1 RT or 4µg Px458-Guide 1 and 2µg MD1-RT, denoted by 1-3 and a no-template water control (H). The presence of a band indicates splicing between the endogenous full length Dp427m and the modified MD1 transgene integrated into the genome.

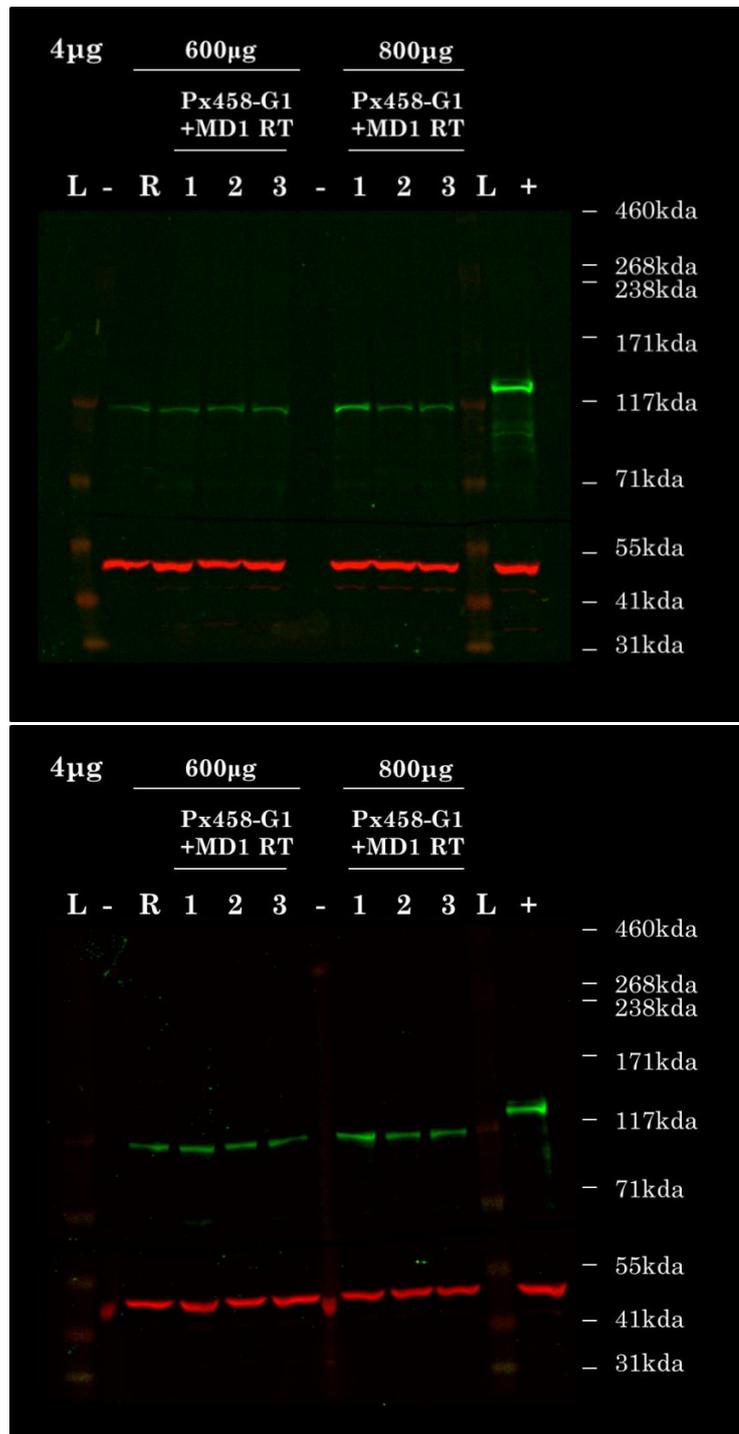


Figure 6-12 Western blotting for MD1 expression in HEK293T 4µg (Px458-Guide 1 1:1 MD1-RT) condition after polyclonal zeocin selection.

The four samples surviving either 600µg/ml or 800µg/ml zeocin selection: MD1 RT (R)-2µg of MD1-RT transfected alone and triplicate co-transfected 2µg Px458-Guide 1 and 2µg of MD1 RT samples, denoted as 1-3; were harvested for total protein lysate. A 50µg sample was subject to western blotting using the MannEx10-11c antibody to examine, for MD1 protein expression and Ab70774 (AbCam) for the 44kDa protein alpha tubulin as a loading control. This was performed alongside a positive control of MD1 transfected HEK293T protein lysate (Top) and untreated HEK293T protein lysate negative control (Bottom). Noticeably, the wild type MD1 protein is 137kDa and a truncated protein estimated to be 117kDa was expressed across all samples.

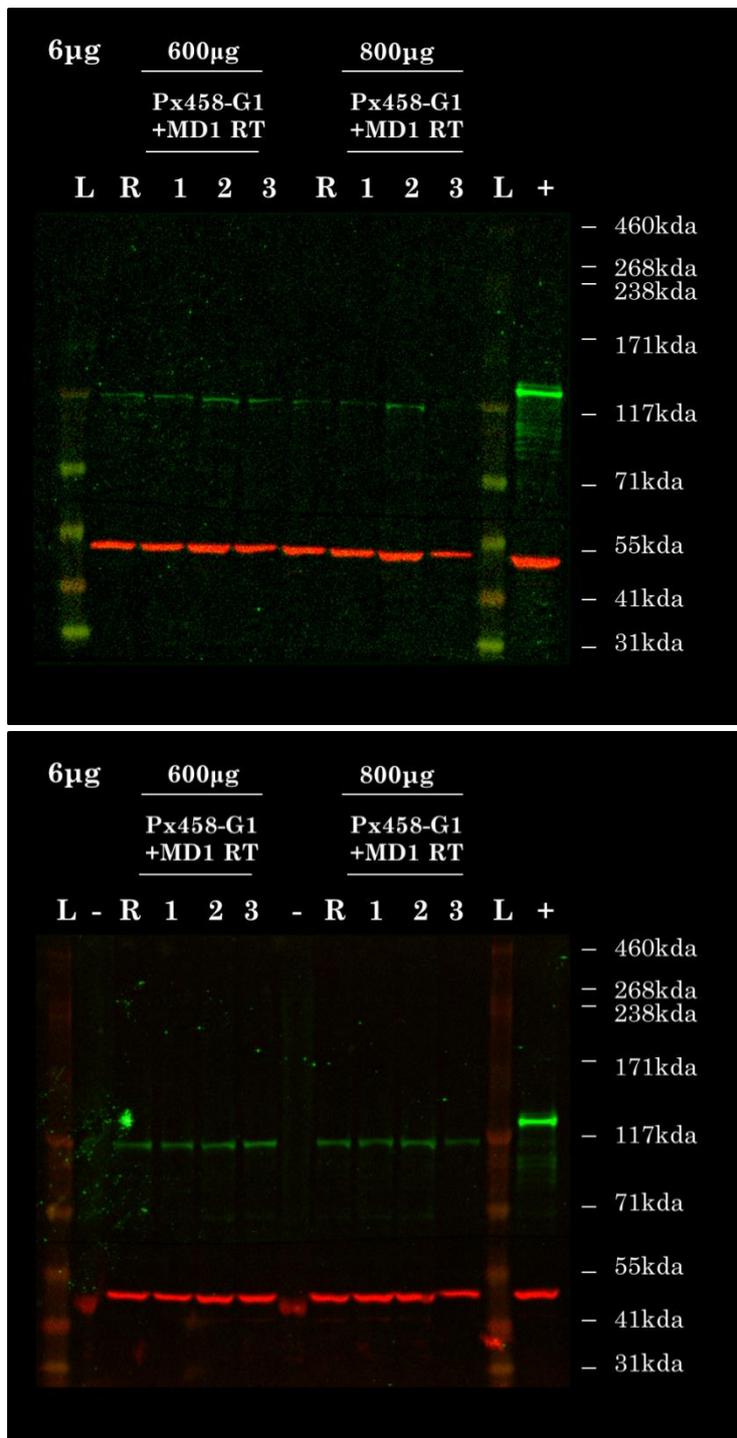


Figure 6-13 Western blotting for MD1 expression in HEK293T 6µg (Px458-Guide 1 2:1 MD1-RT) condition after polyclonal zeocin selection.

The four samples surviving either 600µg/ml or 800µg/ml zeocin selection: MD1 RT (R)-2µg of MD1-RT transfected alone and triplicate co-transfected 4µg Px458-Guide 1 and 2µg of MD1 RT samples, denoted as 1-3; were harvested for total protein lysate. A 50µg sample was subject to western blotting using the MannEx10-11c antibody to examine, for MD1 protein expression and Ab70774 (AbCam) for the 44kDa protein alpha tubulin as a loading control. This was performed alongside a positive control of MD1 transfected HEK293T protein lysate (Top) and untreated HEK293T protein lysate negative control (Bottom). Noticeably, the MD1 protein is 137kDa and a truncated protein estimated to be 117kDa was expressed across all samples.

6.3.6 Establishing Single Colonies: Further Enrichment and Molecular Characterisation:

The final plate of polyclonal HEK293T cells was used to establish single colonies for MD1 RT alone and triplicate co-transfected Px458-Guide1 and MD1 RT treatments for each of the four conditions. In pursuit of this, cells were passaged counted and seeded at serial dilutions of 10^5 or 10^4 cells into 10 cm cell culture dishes, in 10mls of media supplemented with zeocin. Notably, the zeocin concentration applied to the 10cm cell culture dish was kept consistent with the condition from which the treatment was derived at the polyclonal level. Treatments derived from the $4\mu\text{g}$ $600\mu\text{g/ml}$ and $6\mu\text{g}$ $600\mu\text{g/ml}$ conditions were placed in media supplemented with $600\mu\text{g/ml}$ of zeocin and those from $4\mu\text{g}$ $800\mu\text{g/ml}$ and $6\mu\text{g}$ $800\mu\text{g/ml}$ were placed in media supplemented with $800\mu\text{g/ml}$ of zeocin respectively. The premise for this being that this would allow for individual cells to attach to the dish, with a suitable radius of space to enable continued proliferation until a single colony achieving a diameter of 2 -3mm across is formed, without coming into contact with other cells.

6.3.6.1 Single Colony Visualisation and Expansion:

After culturing, colonies between 2-3mm in diameter forming an opaque circular region of cells could be clearly visually identified in all 10cm dishes. This stated, 10 cm dishes seeded at the lower 10^4 cell count yielded fewer colonies with greater intervening distances, which was deemed beneficial for ensuring clonality and thus these dishes were used to pick colonies. To identify colonies expanded from a single cell with a suitable radius to enable picking, plates were examined by phase contrast microscopy (Figure 6-14). This highlighted that some colonies were not suitable for selection as the colonies had intersected and may therefore be considered polyclonal in nature. However, in many instances colonies in isolation were identified, and circled for identification.

Once identified, colonies were removed from the 10cm dish using filter discs dipped in 1x trypsin and transferred to a 24 well plate. The cells attached to these discs were then cultured until 60-70% confluency was attained, and then the cells were passaged and transferred to a 6-well plate. These cultures were then subjected to continued expansion in a manner analogous to the polyclonal population.

Subsequently, four 6 well plates were produced that enabled harvesting of DNA, RNA and protein as before, and a final plate for stock cells to be frozen. Importantly, all the cells that are expanded in this instance were derived from a single cell and thus are monoclonal in nature for the genetic modification that has occurred.

6.3.6.2 Crystal Violet Staining of Individual Colonies:

Once picked and transferred to a 24 well plate, the colonies remaining attached to the 10 cm dishes were stained with 0.001% crystal violet solution. This served to highlight a general trend that the number of colonies present on the MD1 RT alone dish were reduced relative the replicates of Px458-Guide 1 and MD1 RT co-transfection, as exemplified by colonies obtained from the 6µg 600µg/ml condition (Figure 6-15). Unfortunately, this could not be directly quantified as the application of trypsin coated filter discs to the 10cm plate resulted in leaching of the trypsin into the nearby proximity in some instances. Thus whilst the filter disc was removed taking the colony of interest in isolation, nearby colonies were disrupted upon the application of crystal violet stain and subsequently detached.

6.3.6.3 Examination of Monoclonal Cultures for Exogenous Repair Template Integration:

Once monoclonal cultures were expanded, genotyping was undertaken using a single round PCR as previously described. For all monoclonal cultures expanded from Px458-Guide 1 and MD1 RT co-transfected clones, amplicons of 1291bps and 1277bps were resolved for LAH and RAH PCRs respectively. This is consistent with the polyclonal analysis and indicates that integration of the exogenous repair template has occurred at the *DMD* intron 1 site. One distinction from the polyclonal analysis identified, was that 1291bp and 1277bp amplicons corresponding to LAH and RAH PCRs were also present in monoclonal cultures expanded from MD1 RT alone clones (Figure 6-16 and Figure 6-17). This provides an indication that some targeted integration, has occurred in HEK293T exposed to the MD1 RT in isolation.

6.3.6.4 Examination of RNA harvested from Clonal Cultures:

The same examination pertaining to transcription of the MD1ΔEx1 transgene performed on the polyclonal populations was repeated at the monoclonal stage of enrichment. As anticipated, this yielded results consistent with the polyclonal analysis. The anticipated RT-PCR 250bp amplicon spanning from the Dp427m promoter to exon 5 of the MD1ΔEx1 transgene was present across both MD1 RT alone and Co-transfected Px458-Guide 1 and MD1 treatments (Figure 6-18). Sequencing of the PCR products at both the polyclonal and monoclonal stages revealed the sequence anticipated for correct Dp427m – MD1 Exon 2 splicing (not shown).

6.3.6.5 Examination of Protein Harvested from Clonal Cultures:

Western blotting performed on 50µg of total protein at the monoclonal stage with MannEx10-11c as described previously was also congruent with results obtained from polyclonal populations. In all MD1 RT alone and co-transfected Px458-Guide 1 and MD1 RT treatments, a 117kDa MD1 truncated protein was present (Figure 6-19 and Figure 6-20). Interestingly, however, when western blots were over-exposed at this stage a low level of full length 137kDa

MD1 protein could be visualised in addition to the 117kDa truncated MD1 protein (Figure 6-21). This indicates some full-length MD1 can be transcribed and successfully synthesised as a protein. The presence of truncated protein is therefore likely due to the occurrence of aberrant splicing. This is speculated as if the truncated protein appeared as a result of internal deletion it would not be possible to synthesise the full length MD1 protein.

6.3.6.6 Re-examination of Computational Splicing in the Context of the MD1 Repair Template:

The presence of low level full length 137kDa and high levels of truncated 117 kDa MD1 protein, prompted a re-examination by bioinformatics of potential cryptic splice sites within the MD1 exogenous repair template (Figure 6-22). Interestingly, whilst the artificial beta globin splice acceptor appeared strongest at position 181 nucleotides in accordance with Human Splice Finder bioinformatics analysis, there were low levels of upstream splicing enhancers. Furthermore, downstream of this splice acceptor was both a cluster of relatively strong splice acceptors and enhancers around 201nts. It is thus plausible that this cluster could work antagonistically, competing for Dp427m promoter and exon 1 element splicing.

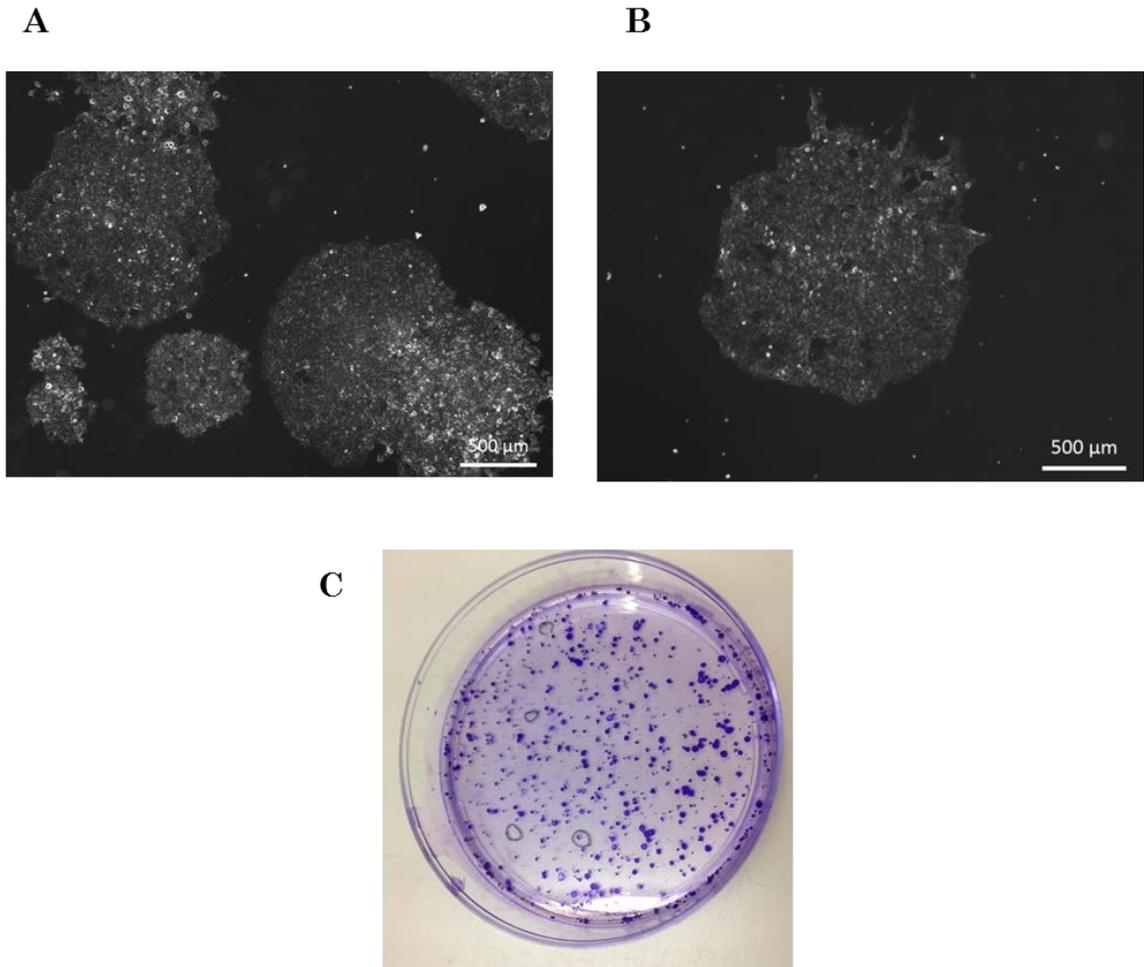


Figure 6-14 Phase contrast microscopy images and visual analysis of single colonies on a 10 cm dish using crystal violet.

(A) Phase contrast microscopy image of multiple colonies in close proximity with one another, these are unsuitable for passaging due to likelihood they have intercepted and may be polyclonal in nature. (B) Phase Contrast microscopy image of a single colony with ample diameter 2-3 mm, with suitable distance from other colonies to assert that this population arose from a single cell. Both microscopy images were taken at 200x magnification. (C) Crystal violet stains of the 10 cm dish these colonies were imaged from.

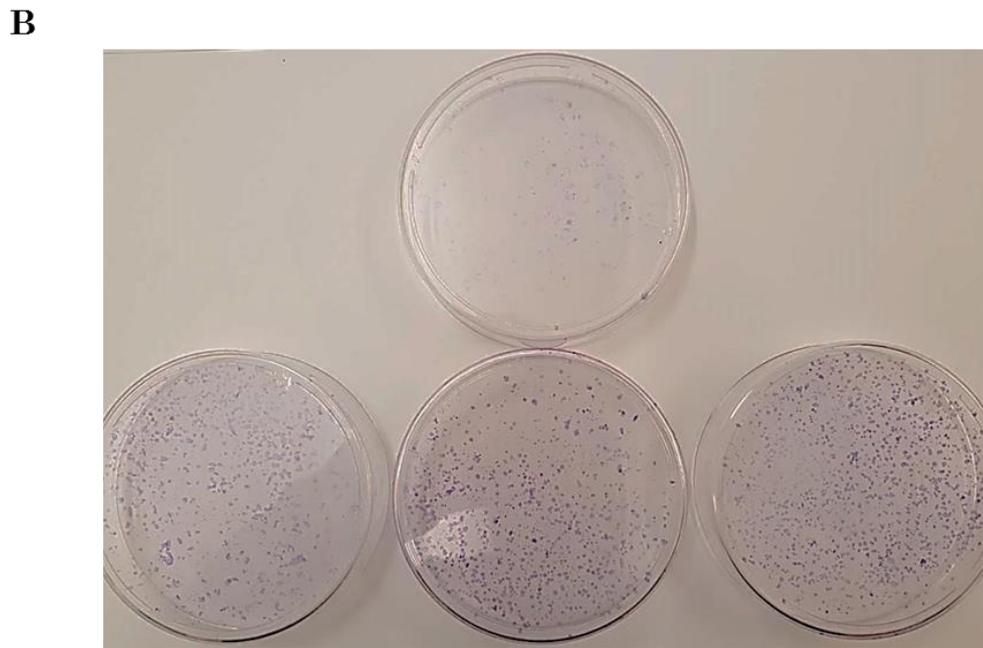
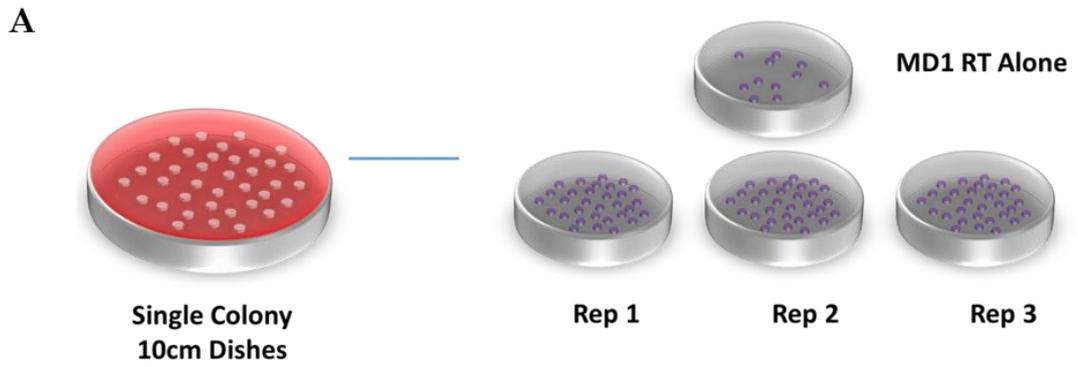


Figure 6-15 Crystal Violet Stains performed on the single colonies from the MD1 repair template alone and triplicate 4µg Px458-Guide 1 and 2µg of MD1 RT co-transfected samples, following 600µg/ml zeocin selection.

A schematic showing the relative frequencies of colonies observed post crystal violet staining, for MD1 RT Alone condition and individual repetition of 4µg Px458-Guide 1 and 2µg of MD1 RT co-transfected samples. B) Crystal violet stains of the MD1 RT alone (Top) and three repetitions of Px458-Guide 1 and MD1 co-transfection. Notice the trend that colonies appear to be increase in the latter relative to the former.



Figure 6-16 An examination of repair template integration into the *DMD* intron 1, undertaken on DNA harvested from three individual clones from the 4µg 600µg and 800µg conditions. This was performed by a single round PCR amplification that spanned from the endogenous genomic sequence, across the left or right arms of homology into the repair template.

Gel electrophoresis to resolve PCR products resulting from a single round of amplification. Performed on genomic DNA harvested from both 4µg (Px458-Guide 1 1:1 MD1-RT) exposed to either 600µg/ml or 800µg/ml of zeocin. PCRs were performed and resolved spanning either the left arm of homology (LAH) or the right arm of homology (RAH). Samples included: MD1 RT (R) - 2µg of MD1-RT transfected alone and triplicate co-transfection of either 2µg Px458-Guide 1 and 2µg of MD1 RT denoted by 1-3 and a no-template water control (H). A single round of amplification was performed in which 250ng of genomic DNA was subjected to a 35 cycle GoTaq G2 Flexi PCR, prior to being resolved on a 1% agarose gel alongside hyperladder I. The findings indicate evidence for a positive integration of the MD1 exogenous repair template in both repair template alone samples and co-transfected samples.

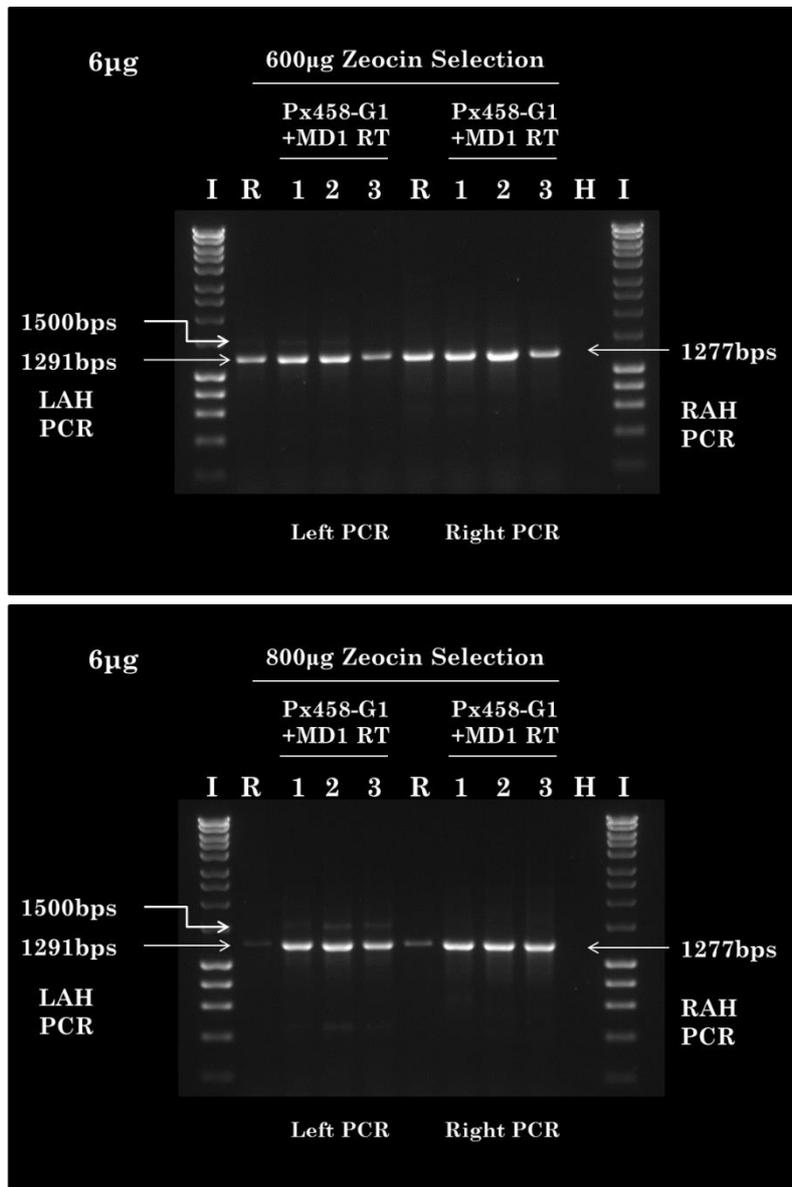


Figure 6-17 An examination of repair template integration into the *DMD* intron 1, undertaken on DNA harvested from three individual clones from the 6µg 600µg and 800µg conditions of HEK293T cells. This was performed by a single round PCR amplification that spanned from the endogenous genomic sequence, across the left or right arms of homology into the repair template.

Gel electrophoresis to resolve PCR products resulting from a single round of amplification. Performed on genomic DNA harvested from both 6µg (Px458-Guide 1 2:1 MD1-RT) exposed to either 600µg/ml or 800µg/ml of zeocin. PCRs were performed and resolved spanning either the left arm of homology (LAH) or the right arm of homology (RAH). Samples included: MD1 RT (R) - 2µg of MD1-RT transfected alone and triplicate co-transfection of either 2µg Px458-Guide 1 and 2µg of MD1 RT denoted by 1-3 and a no-template water control (H). A single round of amplification was performed in which 250ng of genomic DNA was subjected to a 35 cycle GoTaq G2 Flexi PCR, prior to being resolved on a 1% agarose gel alongside hyperladder I. The findings indicate evidence for a positive integration of the MD1 exogenous repair template in both repair template alone samples and co-transfected samples.

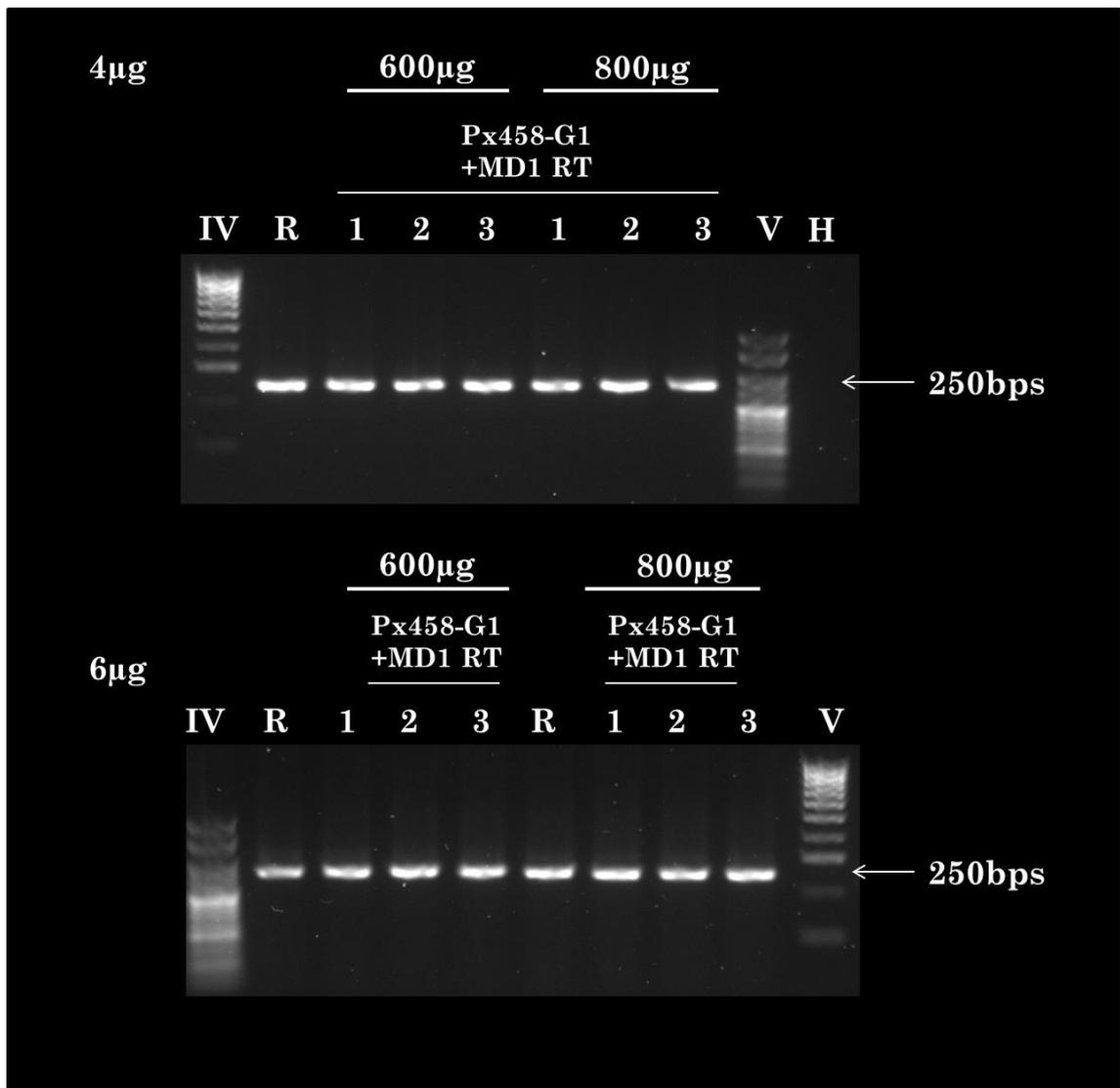


Figure 6-18 An examination of the endogenous Dp427m promoter and exon 1 splicing to MD1 exon 2, using an RT-PCR across the exon 1-exon 2 junction, undertaken on RNA harvested from monoclonal HEK293T cultures.

Gel electrophoresis to resolve PCR products resulting from a single round of amplification of cDNA. Importantly, cDNA was produced from 500ng of total mRNA lysate, and amplified with isoform specific primers, in which the forward primer targeted the Dp427m Exon 1 and MD1 Exon 2 splice junction. This was performed for both 4µg (Px458-Guide 1 1:1 MD1-RT) top and 6µg (Px458-Guide 1 2:1 MD1-RT) bottom conditions, for the MD1 RT (R)-2µg of MD1-RT transfected alone and triplicate co-transfection of either 2µg Px458-Guide 1 and 2µg of MD1 RT or 4µg Px458-Guide 1 and 2µg MD1-RT, denoted by 1-3 and a no-template water control (H). The presence of a band indicates splicing between the endogenous full length Dp427m and the modified MD1 transgene integrated into the genome.

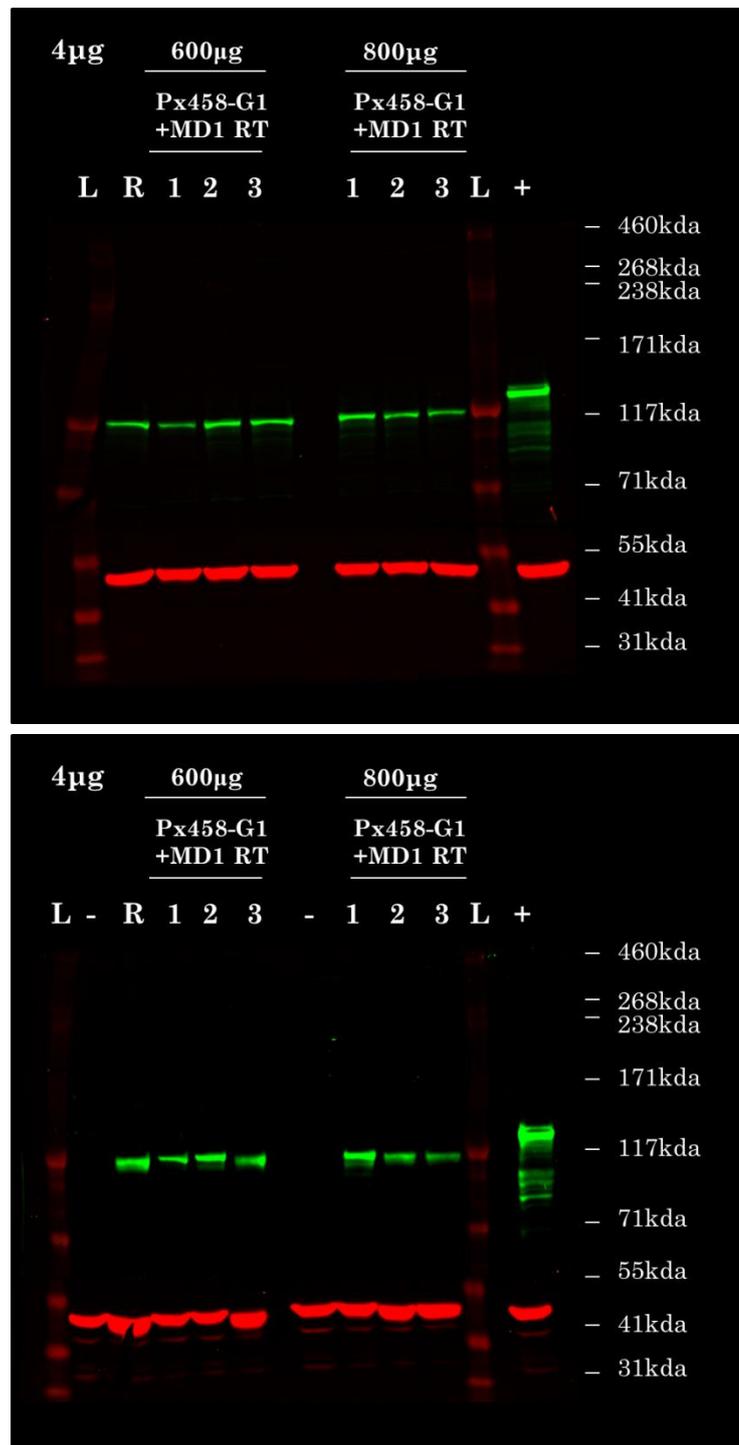


Figure 6-19 Western blotting for MD1 expression in monoclonal samples from the HEK293T 4µg (Px458-Guide 1 1:1 MD1-RT) condition.

Individual colonies expanded for the four conditions: MD1 RT (R)-2µg of MD1-RT transfected alone and triplicate 2µg Px458-Guide 1 and 2µg of MD1 RT co-transfected samples, denoted as 1-3 subject to 600µg/ml or 800µg/ml zeocin selection; were harvested for total protein lysate. A 50µg sample was subject to western blotting using the MannEx10-11c antibody to examine MD1 protein expression and Ab70774 (AbCam) for the 44kDa protein alpha tubulin, used as a loading control. This was performed alongside a positive control of MD1 transfected HEK293T protein lysate (Top) and untreated HEK293T protein lysate negative control (Bottom). Noticeably, the MD1 protein is 137kDa and a truncated protein estimated to be 117kDa was expressed across all samples.

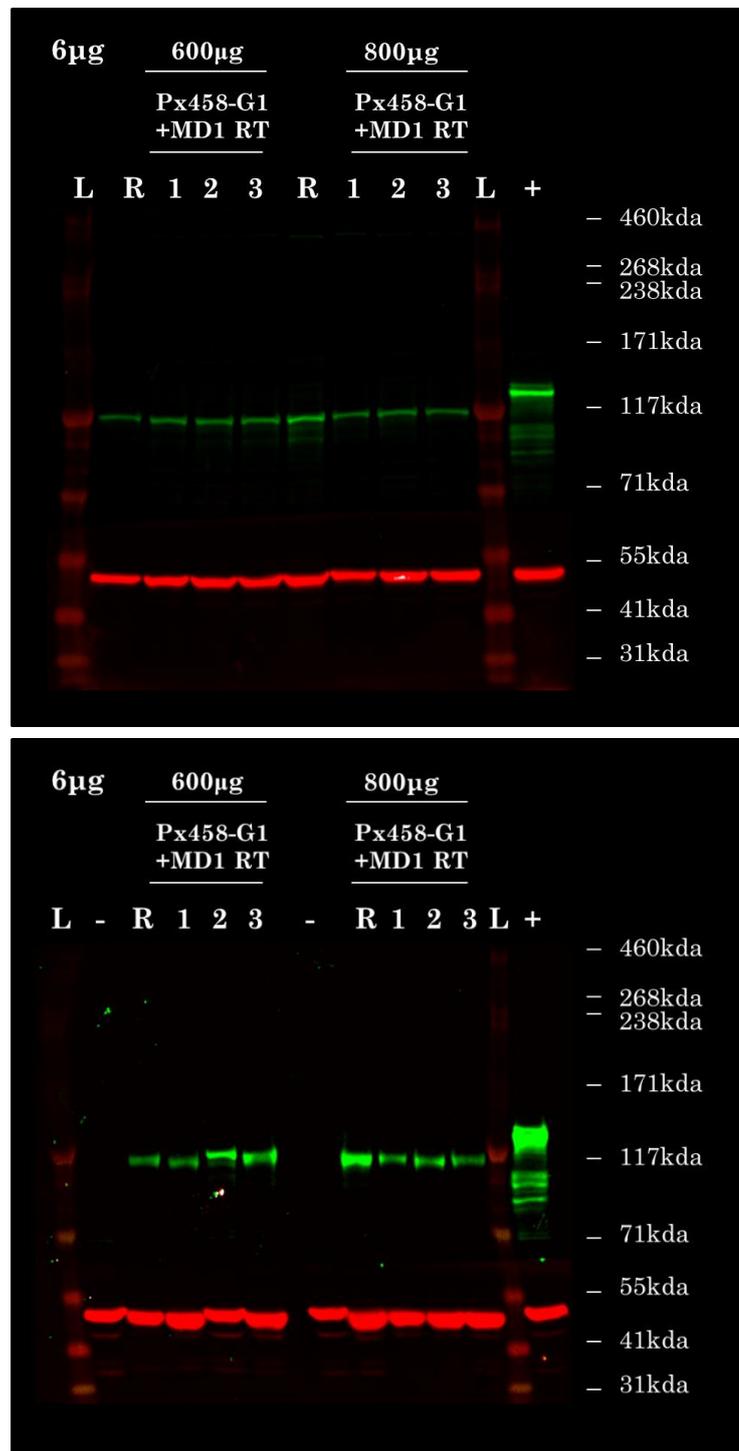


Figure 6-20 Western blotting for MD1 expression in monoclonal samples from the HEK293T 6 μ g (Px458-Guide 1 2:1 MD1-RT) condition.

Individual colonies expanded for the four conditions: MD1 RT (R)-2 μ g of MD1-RT transfected alone and triplicate 4 μ g Px458-Guide 1 and 2 μ g of MD1 RT co-transfected samples, denoted as 1-3 subject to 600 μ g/ml or 800 μ g/ml zeocin selection; were harvested for total protein lysate. A 50 μ g sample was subject to western blotting using the MannEx10-11c antibody to examine MD1 protein expression and Ab70774 (AbCam) for the 44kDa protein alpha tubulin, used as a loading control. This was performed alongside a positive control of MD1 transfected HEK293T protein lysate (Top) and untreated HEK293T protein lysate negative control (Bottom). Noticeably, the MD1 protein is 137kDa and a truncated protein estimated to be 117kDa was expressed across all samples.

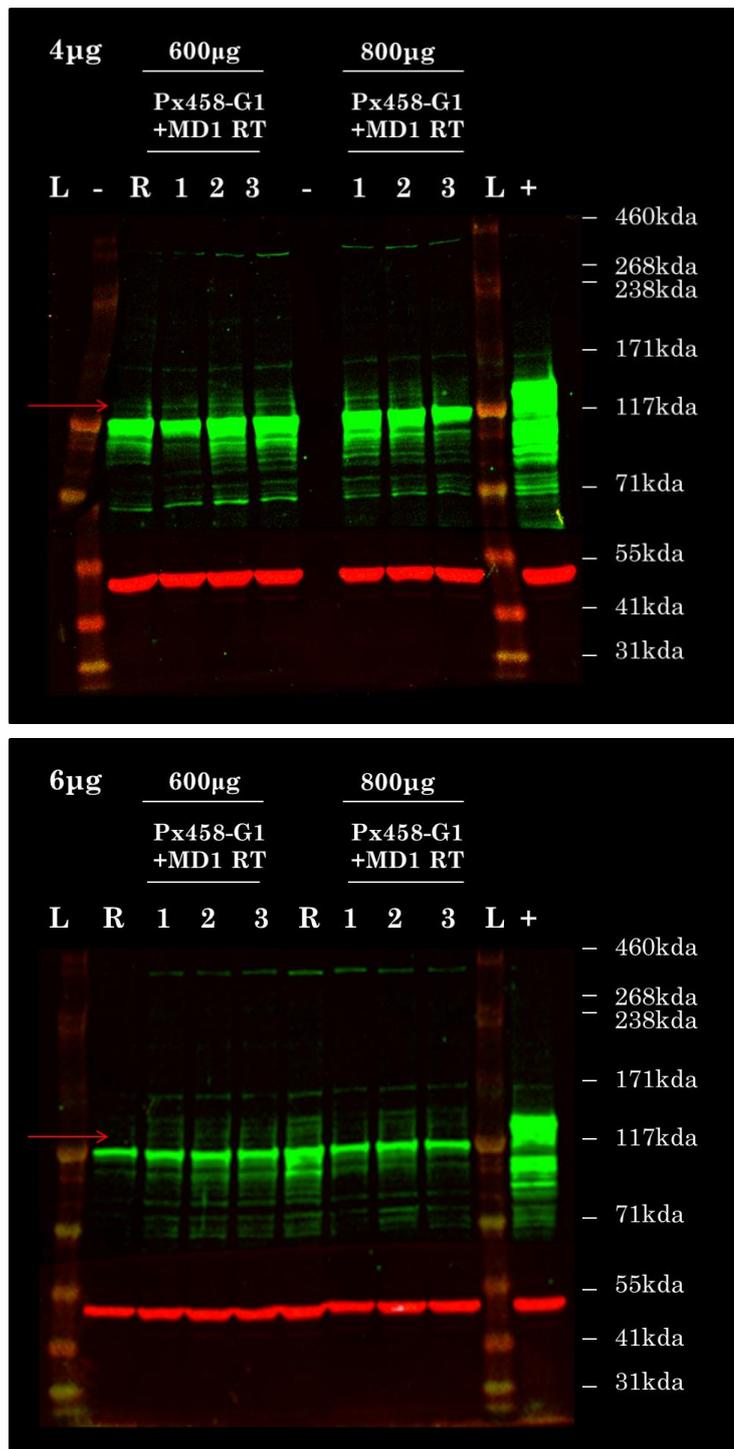


Figure 6-21 Overexposure of Western blots of monoclonal samples from 4µg (Px458-Guide 1 1:1 MD1-RT) and 6µg (Px458-Guide 1 2:1 MD1-RT) conditions; examination for full length MD1 expression.

Individual colonies expanded for the four conditions: MD1 RT (R)-2µg of MD1-RT transfected alone and triplicate 2µg or 4µg Px458-Guide 1 and 2µg of MD1 RT co-transfected samples, denoted as 1-3 subject to 600µg/ml or 800µg/ml zeocin selection; were harvested for total protein lysate. A 50µg sample was subject to western blotting using the MannEx10-11c antibody to examine MD1 protein expression and Ab70774 (AbCam) for the 44kDa protein alpha tubulin, used as a loading control. This was performed alongside a positive control of MD1 transfected HEK293T protein. Noticeably, the MD1 protein is 137kDa and a truncated protein estimated to be 117kDa was expressed across all samples, however on overexposure a faint band indicating full length MD1 is apparent as indicated by the red arrow.

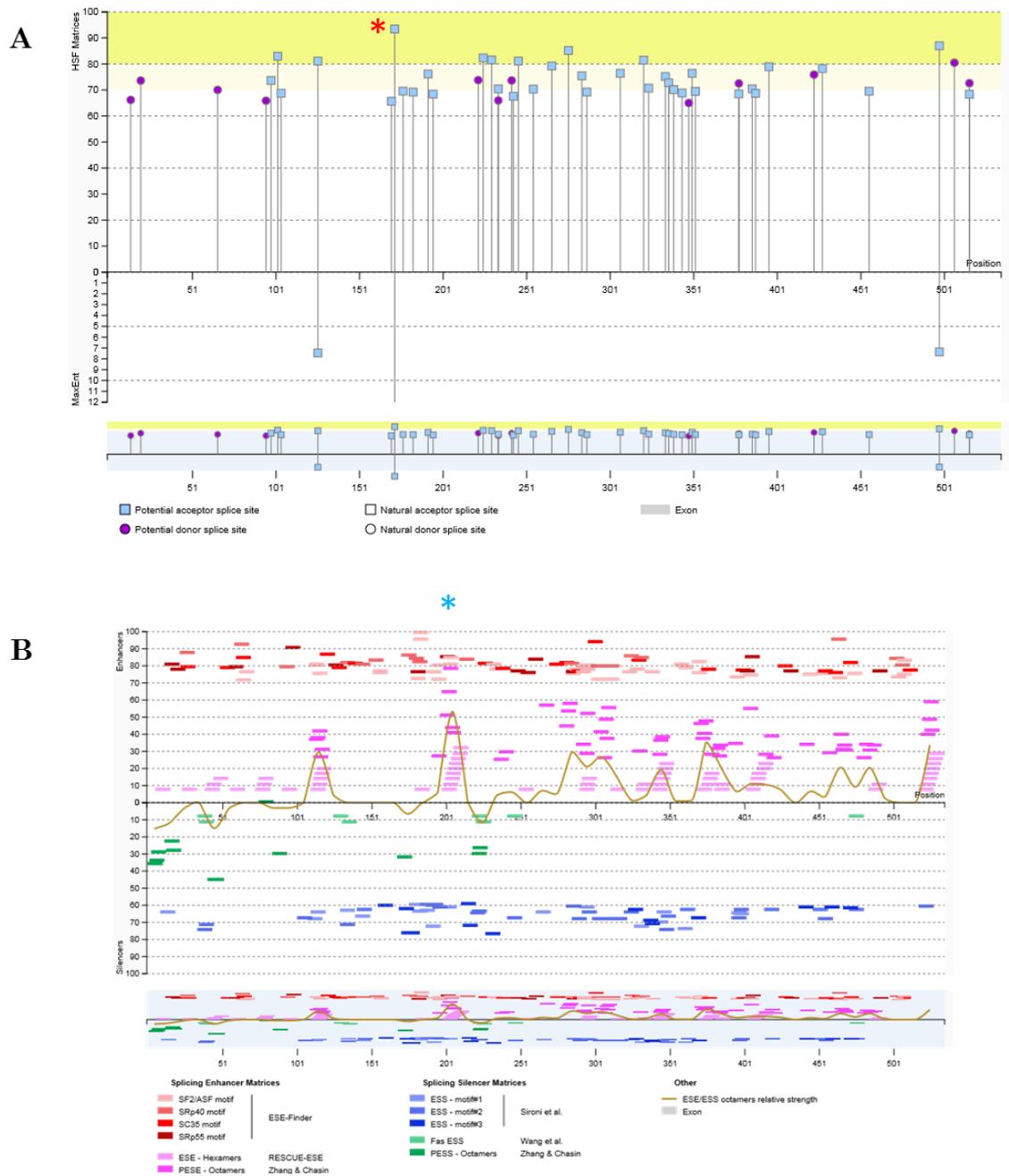


Figure 6-22 Post Hoc bioinformatics analysis of the 5' region of the MD1 exogenous repair template, using Human Splice Finder, to predict the presence of splicing acceptors and enhancers.

(A) Identification of splice acceptors (blue squares) and splice donors (purple circles) across the MD1 exogenous repair template with the artificial beta globin splice acceptor indicated by a blue star. Notice this is the strongest splice acceptor sequence present; however there is a downstream cluster around the 201 nucleotide region downstream. (B) Identification of splicing enhancers (pink/red rectangles) and splicing silencers (blue/green) rectangles, across the sequence. Interestingly few enhancers are upstream of the splice acceptor however; there is a cluster downstream around 201 nucleotides; this is indicated by the red star.

6.3.7 Patient Myoblasts: A Parallel CRISPR Cas9- HDR Mediated Knock-in of the MD1 Transgene:

The gene-editing approach was then expanded to immortalised DMD patient myoblasts carrying a deletion of exons 45 to 52. As this muscle cell line harbours an out-of-frame dystrophin mutation, it was deemed a more appropriate cell lineage to test MD1 Δ Ex1 knock-in downstream of the exogenous Dp427m full length dystrophin promoter and exon 1 sequence.

6.3.8 Assessment of Transient Transfection Efficiency:

The same 1:1 and 2:1 co-transfections of Px458-Guide 1: MD1 RT was undertaken alongside controls as previously described for the HEK293T integration study. Phase contrast and fluorescence microscopy images in the FITC channel, were taken on immortalised Δ 45-55 patient myoblasts 48 hours post transfection (Figure 6-23). Notably, the transient transfection was successful but the GFP expression level visually appeared slightly lower than levels observed upon 2 μ g and 4 μ g transient transfections of Px458 to this cell line previously, being $38.92\% \pm 1.55$ and $42.89\% \pm 0.67$ respectively.

Following this visual indication of successful transient transfection at 48 hours, the treated immortalised patient myoblasts were passaged and seeded at cell count of 8.3×10^4 into two six well plates. The media was supplemented on seeding with a volume of zeocin antibiotic in line with either a 800 μ g/ml or 1000 μ g/ml dose, consistent with the conditions outlined by the zeocin sensitivity assay.

6.3.8.1 DNA Extraction and PCR Amplification Processes to Identify Successful Integration:

Immortalised Δ 45-52 patient myoblasts in excess of the 8.3×10^4 cell count required for zeocin selection, attained from Mock, Px458-Guide1 and MD1 RT controls and co-transfected Px458-Guide 1 + MD1 RT samples were subject to Qiagen DNA extraction. This was undertaken for both the 4 μ g 1:1 and 6 μ g 2:1 conditions.

Due to the lower transient transfection, 250ng of harvested DNA was amplified only by nested PCR to assess for the integration of the exogenous repair template. Notably, a low intensity amplicon of 1139bps was present in the 6 μ g, 2:1 Px458-Guide 1 and MD1 RT co-transfected samples. Although far from a stringent and robust result, this was indicative that a low level of transgene integration could have occurred, a prospect that warranted further investigation (Figure 6-24).

6.3.8.2 Zeocin Selection of Immortalised Patient Myoblast Polyclonal Populations:

In consideration of the positive PCR across the left genome-repair template junction, it was deemed pertinent to continue the positive selection process, for the 6 μ g transfection. Phase contrast and fluorescent microscopy imaging undertaken in the FITC (GFP) channel were taken

at 3, 7, 14, 18, 21 and 24 day time points to examine the selection process and cell viability for both 6µg 800µg/ml (Figure 6-25) and 6µg 1000µg/ml (Figure 6-26) zeocin streams.

At day 3, phase contrast images showed some cell survival was apparent in both of the conditions. Notably, cells were estimated to be at approximately 10-15% confluent in the 6µg 800µg/ml condition across all treatments, whereas slightly higher cell numbers were observed in the 6µg 1000µg/ml condition. With Px458-Guide 1 alone and MD1 RT alone showing approximately 10% confluency, but Mock and co-transfected Px458-Guide 1 and MD1 RT samples showing estimated confluency of 50%. This could be attributed to a slightly higher cell number being seeded at 48 hours. In addition, few GFP expressing cells were observed at this time and a high level of auto-fluorescence of the patient myoblast culture made it difficult to delineate cells undergoing cell death from those positively transfected. As a consequence fluorescent imaging was halted at this stage.

By day 7, the cell number appeared to increase an estimated 2 fold relative to day 3, in the 6µg 800µg/ml condition. This stated cell number appears to remain comparable in the 6µg 1000µg/ml condition. This was observed across all treatments.

At day 14, 6µg 800µg/ml showed a reduction in cell number, becoming comparable once more with the cell number observed at day 3. In contrast in the 6µg 1000µg/ml, a marked reduction in cell viability was observed in mock, Px458- Guide 1 alone and MD-1 alone treatments. Interestingly, however, cell number only exhibited subtle decrease in Px458-Guide 1 and MD1 RT transfected samples.

The final time-points of day 18-24 showed continued cell decline and death in both conditions, across all treatments. Interestingly, a few cells were still present in the 6µg 1000µg/ml condition. This stated no proliferation or expansion of this cell population appeared to occur. This resulted in the discontinuation of this experiment.

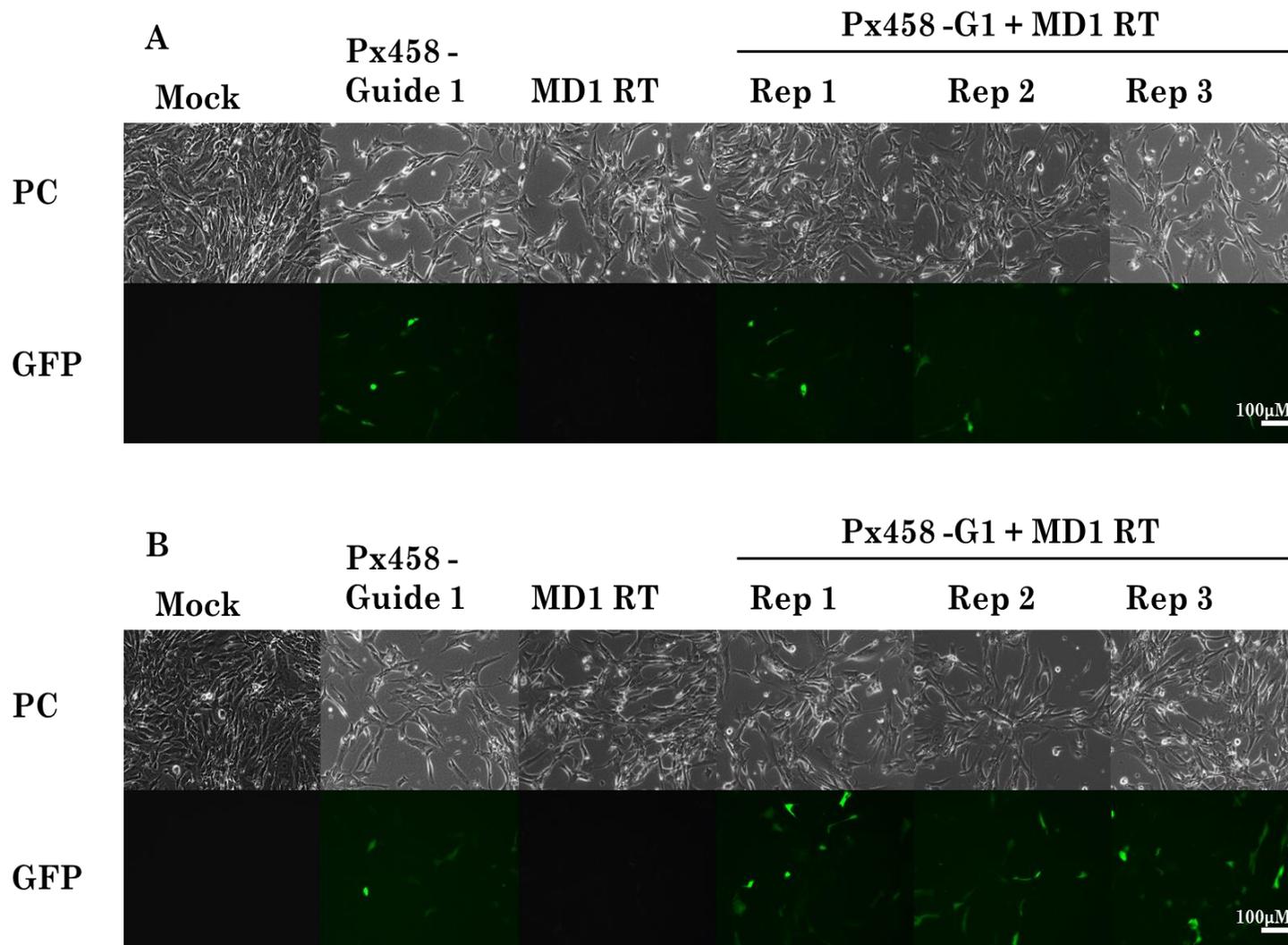


Figure 6-23 Images of Immortalised Δ 45-52 Patient Myoblasts 48 hours post Co-transfection of Px458-Guide 1 at a 2 μ g or 4 μ g dose, with 2 μ g MD1 repair template, alongside Untreated, Px458-Guide 1 or MD1 RT alone.

(A) Phase Contrast (PC) and fluorescent images in the FITC channel (GFP) of Immortalised Δ 45-52 patient myoblasts transfected with 2 μ g of gene editing GFP-plasmid. Mock (M) = 25 μ l of Viafect transfection reagent alone; Px458-Guide 1 in which 2 μ g Px458 alone; MD1-RT= 2 μ g of MD1 RT alone; Px458-Guide 1 + MD1 RT represents co-delivery of 2 μ g Px458-Guide 1 and MD1 RT in a 1:1 ratio across three technical replicates (Rep 1-3). **(B)** Phase Contrast (PC) and fluorescent images in the FITC channel (GFP) of Immortalised Patient Myoblasts transfected with 4 μ g of gene editing GFP-plasmid., same as **(A)** except 4 μ g of Px458-Guide 1 was used in delivery alone and co-delivery alongside 2 μ g MD1-RT constituting a 2:1 ratio.

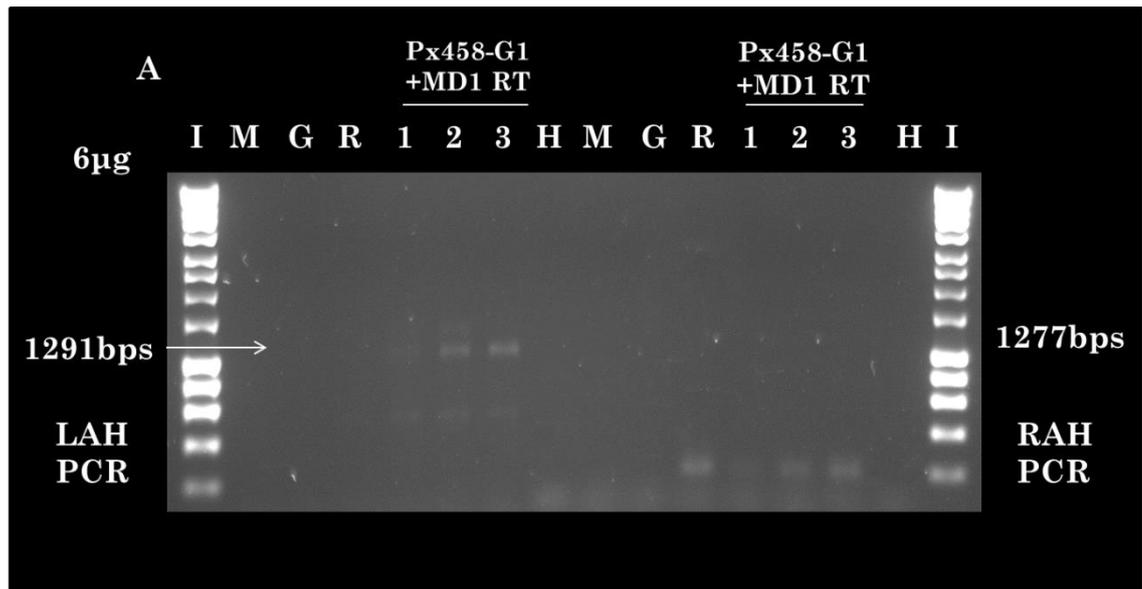


Figure 6-24 An examination of repair template integration into DMD intron 1, using PCR amplification; spanning from the endogenous genomic sequence, across the left or right arms of homology into the repair template. Performed for the 6 μ g (Px458-Guide 1 2:1 MD1-RT) condition.

A) Gel electrophoresis to resolve PCR products resulting from a Nested PCR amplification performed on genomic DNA harvested from the 6 μ g (Px458-Guide 1 2:1 MD1-RT) condition, 48 hours post transient transfection. In which 1 μ l of a single outer PCR comprised of 20 cycles performed on 250ng of genomic DNA, was added to a second 30 cycle PCR to improve the amplicon yield. PCRs were performed and resolved spanning either the left arm of homology (LAH) or the right arm of homology (RAH). Samples included: Mock (M) – cells incubated with Viafect transfection reagent alone, Px458-Guide 1 (G) - 2 μ g of Px458-Guide 1 transfected alone, MD1 RT (R) - 2 μ g of MD1-RT transfected alone and triplicate co-transfection of 2 μ g Px458-Guide 1 and 2 μ g MD1-RT, denoted by 1-3 and a no-template water control (H).

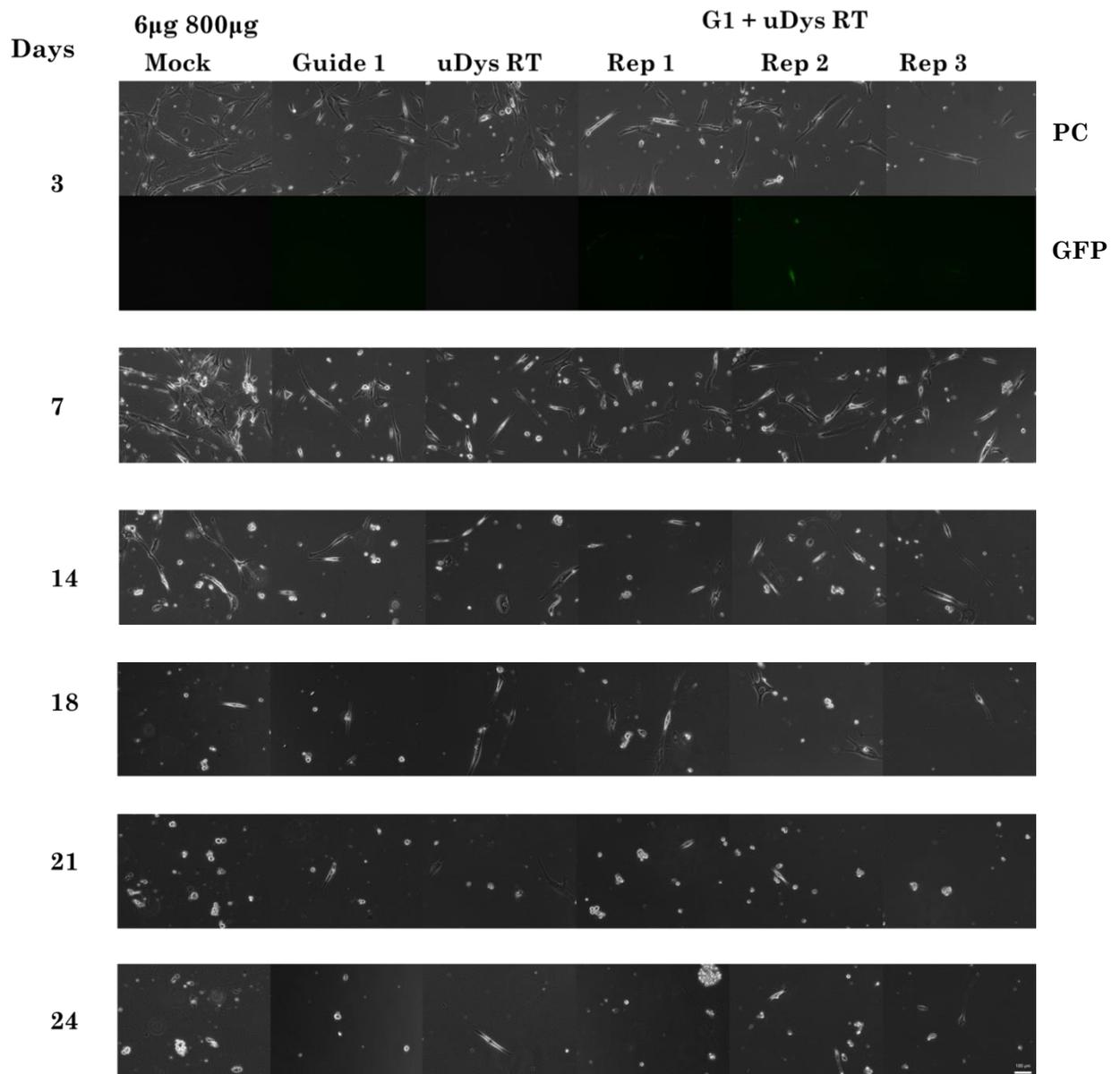


Figure 6-25 Microscopy images of Immortalised $\Delta 45-55$ patient myoblasts post 6µg (Px458-Guide 1 2:1 MD1-RT) co- transfection, alongside controls during the process of zeocin selection. At 48 hours post transfection, Patient myoblasts were passaged and seeded at 8.3×10^4 cells per well and exposed to 800µg/ml of zeocin.

Samples included: Mock (M) – cells incubated with Viafect transfection reagent alone, Px458-Guide 1 (G) - 4µg of Px458-Guide 1 transfected alone, MD1 RT (R) - 2µg of MD1-RT transfected alone and triplicate co-transfection of 4µg Px458-Guide 1 and 2µg MD1-RT, denoted by1-3. A 1ml growth media aliquot was removed, and substituted with 1ml fresh growth media supplemented with 800µg/ml of Zeocin every 3 days to maintain selection. Phase Contrast and fluorescent images in the FITC channel (GFP) were taken at time points of: 3 days; phase contrast was continued to 7, 14, 18, 21 and, 24 days after the GFP expression had diminished.

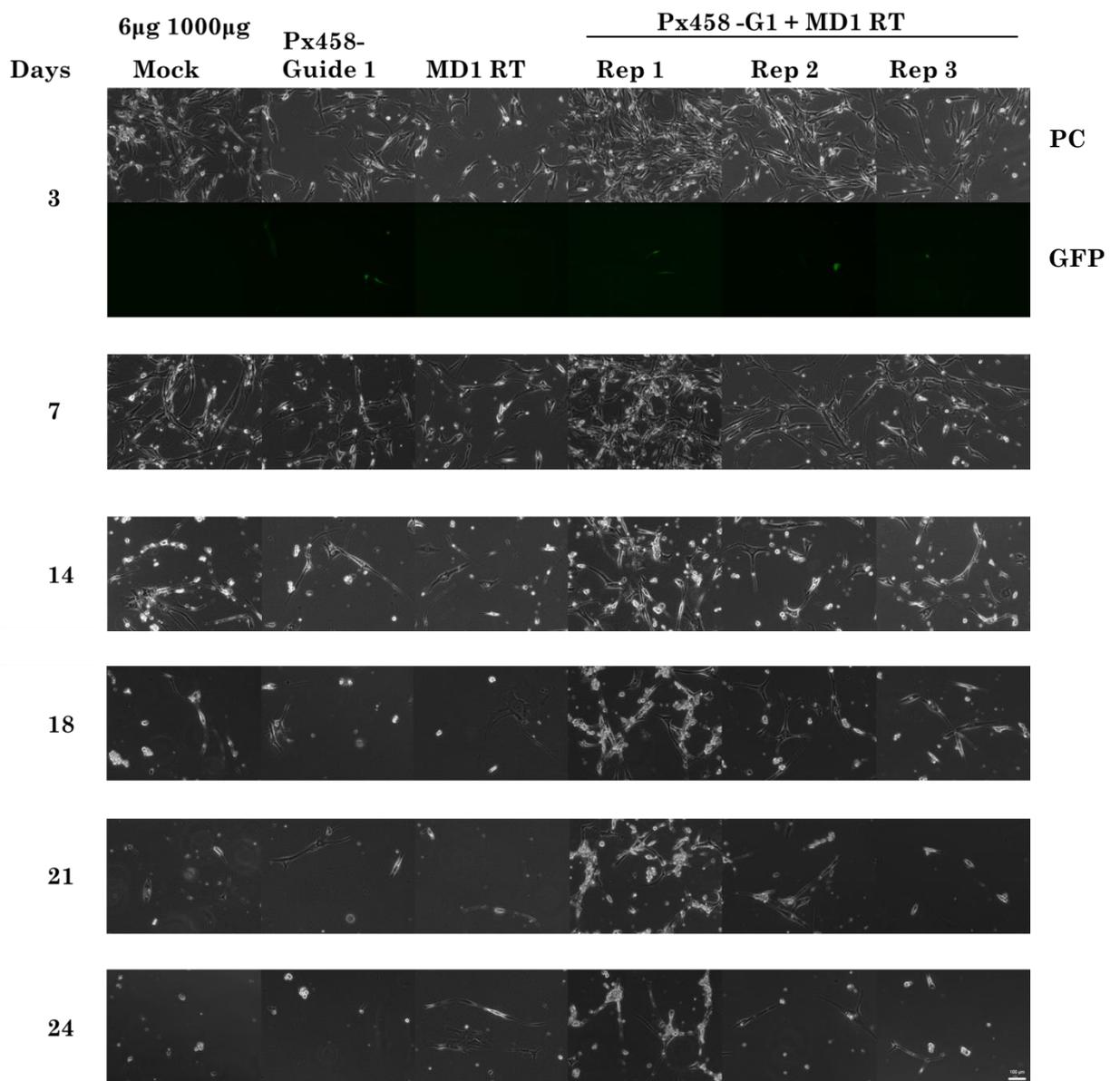


Figure 6-26 Microscopy images of Immortalised $\Delta 45-55$ patient myoblasts post 6µg (Px458-Guide 1 2:1 MD1-RT) co- transfection, alongside controls during the process of zeocin selection. At 48 hours post transfection, Patient myoblasts were passaged and seeded at 8.3×10^4 cells per well and exposed to 1000µg/ml of zeocin.

Samples included: Mock (M) – cells incubated with Viafect transfection reagent alone, Px458-Guide 1 (G) - 4µg of Px458-Guide 1 transfected alone, MD1 RT (R) - 2µg of MD1-RT transfected alone and triplicate co-transfection of 4µg Px458-Guide 1 and 2µg MD1-RT, denoted by 1-3. A 1ml growth media aliquot was removed, and substituted with 1ml fresh growth media supplemented with 1000µg/ml of Zeocin every 3 days to maintain selection. Phase Contrast and fluorescent images in the FITC channel (GFP) were taken at time points of: 3 days; phase contrast was continued to 7, 14, 18, 21 and, 24 days after the GFP expression had diminished

6.4 Discussion

6.4.1 HEK293T, HDR-mediated MD1 Transgene Integration:

6.4.1.1 *Transient Transfection of HEK293T:*

The transfection efficiency of HEK293T was striking when using the all-in-one gene-editing plasmid Px458 (U6-Guide-CBh-Cas9-2A-GFP) in which green fluorescence was used as a proxy for plasmid uptake. In both the 4 μ g and 6 μ g transient transfection, 80-90% of cells were estimated to exhibit green fluorescence. This served to highlight that the transfection efficiency of gene-editing reagents was not the limiting factor in this investigation.

Interestingly, the same successful integration of the MD1 transgene was not extended to the three component system trialled comprising of: Px165 (CBh-Cas9), U6-Guide, and the MD1-RT (data not shown). A likely explanation of this can be afforded in the first instance by a lower efficiency of delivery. Transient transfection is fundamentally reliant upon electrostatic interaction between cationic lipids and anionic nucleotides to facilitate the introduction of DNA into cells (Zuris et al. 2014). The general rule of thumb being the larger the piece of DNA or the greater the number of components sought to be introduced into the cell, the lower the overall transfection efficiency. This arguably was the main justification for the development of a chimeric guide comprising crRNA-Hairpin loop- tracrRNA coined single guide RNA (sgRNA), the conventional scaffold used throughout this investigation. This single scaffold, enables the delivery of crRNA and tracrRNA together as one component able to initiate gene editing in mammalian cell culture as opposed to the two separate constituents observed in the endogenous system (Jinek et al. 2012). In addition to the introduction of the three components, their co-expression may also be a limiting factor. However, it should be noted that the three-component gene editing system described had no fluorescent marker to indicate successful DNA introduction to the HEK293T cells. Thus a methodological error pertaining to the transfection mixture constitution cannot be excluded. Future investigations should aim to fluorescently tag all components required for gene-modification to enable suitable tracking throughout in-vitro development.

6.4.1.2 *Integration of the MD1 Transgene at the Genomic and Transcriptomic levels:*

Importantly the PCR amplification extending from the genome-repair template and repair template – genome on the left across the left and right arms of homology respectively yielded an amplicon indicative of integration across all stages examined. The stages examined included 48 hours post transfection, polyclonal HEK293T-MD1 and monoclonal HEK293T-MD1. In the former stage a nested PCR was required to detect the desired amplicon, whereas in the latter

two stages a single round of amplification was sufficient; this is consistent with a zeocin-mediated enrichment of genetically-corrected cells.

The survival of HEK293T cells treated with the MD1 RT alone yielded no amplicon associated with integration at the 48 hour post transfection stage. Initially, the absence of an amplicon led to speculation that random or non-targeted integration of the transgene and zeocin cassette had occurred. Notably, there was an overall trend that this treatment yielded a lower proportion of viable cells, relative to conditions in which the repair template was co-transfected with gene-editing reagents. The combinations of these two observations added weight to this hypothesis. However, at the single colony stage the PCR amplicon was positive for site-specific integration of the repair template when used alone, a finding that was substantiated by the later determination of the occurrence of Dp427m promoter and exon 1 element and MD1 exon 2 splicing in the MD1 RT alone treatment. In light of these findings, a new postulation of low level targeted integration became favoured. This being founded upon the basis that low levels of HDR can occur in the absence of a DSB, albeit with much less efficiency (Jasin & Berg 1988; Jasin & Rothstein 2013); this coupled to the finding that some genomic loci are more susceptible to DNA integration events, two well-documented examples being the AAVSI and Albumin Loci (Kotin et al. 1990; Barzel et al. 2014). Admittedly, these studies have a viral based integration focus; however, the genomic locus being predisposed to integration may be postulated to facilitate plasmid based integration as well. The target site within intron 1 for this investigation appears highly conserved; the possibility of a regulatory function and the associated open chromatin structure cannot be overlooked.

Both of these theories demand further investigation, although all of the colonies yielded a positive PCR, indicating MD1 integration at monoclonal stage and there was no evidence of random integration. The process of Southern Blotting may elucidate whether copies of the MD1 transgene were integrated at a site distinct from that targeted by nucleases. If present this undesirable effect could be reduced with the use of negative selection utilising the Herpes Simplex Virus thymine kinase (HSVtk), which when integrated into a region outside of the arms of homology sensitises cells to ganciclovir (Schwartz et al. 1991). This could also be achieved by utilising the analogous strategy using Diphtheria Toxin A fragment (Araki et al. 2006). In contrast, the potential of the region of homology identified having a favourable chromatin structure or indeed an associated regulatory function could be investigated by its subsequent destruction or use with an up-regulatory Cas9 system.

It should be noted that additional bands appear to occur upon the left integration PCR at around 1500bps in addition to the 1291bps, both upon nested conditions and in the final

polyclonal and monoclonal amplifications. Initially, this was anticipated to be outer PCR product at the 48 post transfection condition. However, its continued presence in single round amplification processes renders this argument inadequate. Thus this is largely thought to be considered to be a mis-priming artefact of the PCR, potentially highlighting that this PCR could be further optimised. This stated, the artefact arguably only appears in integration positive samples as seen by its absence in Mock, Px458-Guide 1 and MD1 RT alone samples 48 hours post transfection stage. As a consequence, this amplicon could likely be removed by the optimisation of the PCR but is largely considered inconsequential particularly when presented in combination with the right arm of homology PCR.

Once DNA integration had been asserted, the next parameter that warranted investigation was transcription which was undertaken at the polyclonal and monoclonal stages of HEK293T-MD1. Initially, it was anticipated that splicing would only occur in instances where gene-editing reagents were co-delivered with exogenous MD1 repair template, particularly at the polyclonal stage. However, the reverse transcription and subsequent PCR, yielded amplicon sizes congruent with Dp427m promoter and exon 2 splicing across all conditions at this stage including repair template alone and instances of co-delivery of gene-editing reagents. This adds additional weight to the theory of nuclease independent integration at the target locus. This likely indicates that nested amplification of the repair template alone DNA at the polyclonal stage, would likely yield an amplicon consistent with integration. The positive transcription result of the exogenous MD1 repair template alone is further validated by the MD1 protein expression at both the polyclonal and monoclonal HEK293T-MD1.

A major implication of this result is that the endogenous Dp427m promoter can be spliced to exon 2, potentially yielding MD1 transgene expression with the same temporal and spatial expression of dystrophin. This finding could be particularly pertinent to the field of DMD research if the ability to splice to the β -globin splice acceptor is retained across the other full-length dystrophin isoforms Dp427c, Dp427p1 and p2. This is especially important as permanent integration could enable prolonged dystrophin expression, as opposed to gene augmentation approaches.

6.4.1.3 MD1 Protein Expression within Enriched Polyclonal Populations and Monoclonal Populations:

An interesting albeit unanticipated finding was the presence of truncated microdystrophin at the polyclonal and monoclonal stages of this investigation (Full length MD1 137kDa, detected MD1 approximately 117kDa). It should be stated this truncated MD1 was the prominent protein expressed alongside low levels of full-length MD1 at the monoclonal stage of the investigation. The truncated MD1 was shown to result from the treatment of HEK293T cells

with the promoter-less MD1 endogenous repair template, as no protein of corresponding size was detected in HEK293T only cell lysate. Moreover, the presence of this protein product was not only limited to conditions where gene-editing reagents were co-expressed. Initially, this was thought to be attributable to some aberrant initiation of transcription from a downstream methionine constitutively expressed within the MD1-RT. However, the RT PCR data highlights that splicing is occurring between the full length Dp427m promoter and the MD1 Exon2 within the repair template in all conditions that survived to the clonal stage. Equally, all conditions showed evidence of integration at the target locus whether co-expressed with the Cas9 nuclease or expressed in isolation with absence of Cas9.

The evidence accrued indicates that this protein product is a positive dystrophin variant, due to the use of the MannEx10-11c antibody, which recognises an epitope between the exon 10-11 boundaries respectively. However, the use of MannEx10-11c in detection means that currently the exact nature and position of the truncation cannot be empirically asserted. This stated, the MD1 cloned from Exon 6 onwards into the Puc57 backbone, contains only the terminating stop codon downstream from Exon10 and 11c, making it likely that any truncation would likely be a resulting from the 5' end of the *DMD* gene. Due to time constraints, it was not possible to investigate this further. However, such investigations would include the use of a panel of antibodies screening the C-terminal and N-terminal of MD1 protein. Based upon this assumption a post-hoc bioinformatics analysis was performed in an attempt to ascertain the putative identity of this protein. Using the correct in-frame microdystrophin, protein predictions were produced using the downstream in-frame methionine residues. This pertained to 3 initiating ATGs, including 2 ATGs within exon 3, which would yield a 129kDa protein and a final ATG within Exon 6, which would yield a 117kDa protein. Interestingly, this may be consistent with the Internal Ribosome Entry Site (IRES) being used within Exon 5 to enable translational initiation from Exon 6 (Wein et al. 2014). The preferential utilisation of the IRES could be attributable to either the absence of a regulatory region perhaps destroyed as a process of the gene-editing undertaken that would modulate expression from exon 1, or indeed the potential for a sequence alteration that disrupt the reading frame. The latter would be particularly consistent as this mechanism was used to describe N-terminal truncations that were exceptions to the Monaco rule, yielding milder BMD phenotypes with an out-of-frame deletion (Monaco 1989).

In this study describing IRES functionality, the IRES was assessed by sequential deletions of dystrophin exons 1-6, which were fused to a downstream fluorescent reporter. Although the IRES activity was not strong enough to induce fluorescence expression in the HEK293T cell lineage, in the study undertaken (Wein et al. 2014); many parameters are different in this

research including: use of optimised MD1, introduced into the intron 1 site. The former is of particular importance as the sequence optimisation enhancing the level of subsequent transcription, could be postulated to promote the IRES to activity (Raab et al. 2010). For a further assertion of this IRES theory, prospects of aberrant splicing and the potential of cryptic splice sites would also have to be excluded via thorough investigation. A second post-hoc bioinformatics analysis was undertaken using Human Splice Finder to identify other putative splice sites (Desmet et al. 2009). This served to highlight that whilst the artificial beta-globin acceptor previously utilised successfully by our lab was the strongest splice acceptor, there are a multitude of splice enhancers present downstream around 201nts, which could influence aberrant splicing in the MD1 (Popplewell et al. 2013). This would enable splicing of the third exon which in turn could be speculated to bias the IRES translation. Thus it is plausible that more enhancers need to be appended to the 5' end of the MD1-RT, to facilitate the expression of the full length MD1 product.

In summary, evidence for Dp427m promoter exon 1 to exon 2 successful splicing as determined by PCR and sequencing and the presence of a full length MD1 transgene appear consistent with in-frame reading of the MD1 transgene. However, the presence of truncated MD1 necessitates further examination as to how this truncated protein form came into production in the context of IRES, Aberrant / Cryptic Splice sites and / or subtle sequence alterations. These investigations would need to be extended to both the repair template itself and the subsequent integration to find the source of any such mutation that may have occurred. In addition current examinations indicate that the integrated MD1 is being expressed from the endogenous Dp427m promoter. Further investigation examining the function of the resultant transgene is necessitated although this would likely require the use of an animal model such as the mdx mouse and an alternative delivery system.

6.4.2 Stability of HDR-mediated MD1 Transgene Integration in Δ 45-52 Patient Myoblasts,

Stable integration of the MD1 transgene could not be established in the Δ 45-52 patient myoblast culture. The numerous parameters of this investigation namely transfection efficiency, transgene Integration and cell survival on exposure to zeocin, could all be improved upon to enhance the chance of successful MD1 knock-in and subsequent expression.

6.4.3 Transfection Efficiency of Immortalised Δ 45-55 Patient Myoblasts:

The immortalised Δ 45-52 patient myoblasts utilised in this study appear relatively refractory to transient transfection, like many other primary-like cell lines. Introduction of tethered Cas9-2A-GFP constructs indicated notable successful transfection efficiencies in the range of 20 and 40%. However, considering the decreasing proportion of cells associated with each stage of

genome-engineering i.e. number of cells transfected, transfected cells that undergo gene-modification and finally, of interest genetically modified cells that has undergone HDR, the parameter of transfection efficiency could still be drastically improved in attainment of the final HDR-outcome, particularly as the PCR amplification across the left arm of homology, could indicate a low level of integration. Interestingly, the integration PCR is only detectable across a single junction. This is likely attributable to PCR efficiency with the right junction PCR occurring with less efficiency than the left. Delivery of gene-editing reagents remains a fundamental limiting step in the development of therapeutic approaches to numerous cell lines of physiological interest. However, much advancement has been made in the field at large that could be utilised to improve upon the current transfection efficiency.

In the context of a non-viral delivery strategy, multiple transfection reagents have been trialled in this investigation and a series of optimisations were undertaken to try and improve plasmid delivery to patient myoblasts (Chapter 4). Further improvements in transfection efficiency with plasmids would likely necessitate the examination of further transfection reagents and culture conditions. An alternative strategy for the delivery of plasmids too hard to transfect cell lines is electroporation whereby an electrical current is applied to cells to facilitate the uptake of plasmids. In the context of guide and Cas9 plasmids, this has shown comparable or improved efficiency across a range of mammalian cell lines relative to cationic lipid transient transfection based delivery systems (X. Liang et al. 2015). The reason such an approach remained unutilised in this investigation is that the optimisation process is relatively costly and time consuming. Furthermore, if not optimised correctly, cell viability is greatly diminished.

In addition to alternative delivery strategies consideration should be given to alternative delivery platforms for the Cas9 namely, mRNA and protein. Plasmid delivery platforms have a number of inherent limitations such as risk of random integration, higher associated toxicities and/ or off-target effects and lower gene modification across a number of cell lines, which are largely overcome by mRNA and protein gene-editing reagents (X. Liang et al. 2015; Yu et al. 2016; S. Kim et al. 2014; Zuris et al. 2014). It should be noted that transfection, is a multifactorial process related to the dose of reagents administered, the cell number seeded and finally any toxicity (Yu et al. 2016). In this instance, some toxicity was observed in the microscopy images of transfected patient myoblasts relative to the mock samples. If toxicity can be circumvented by the delivery of Cas9 protein as opposed to plasmid based platforms in a manner analogous to that observed in multiple cell lines, then the extent of gene editing can be increased. Moreover, an instance where there is less toxicity lower cell densities could be exploited, thereby enhancing the amount of reagent per cell available for gene editing (Yu et al. 2016). By extension of this, lower doses of reagents would be required and the proportion

of gene modification occurring would be heightened. This could represent a beneficial prospect with patient myoblast where the cell seeding of 8.3×10^4 is relatively low. In conventional transfections this is left for 48 hours prior to delivery, but the optimisation of a protein based system may allow protein transfection to be attempted at 24 hours post seeding. These types of approaches have been provisionally developed (Chapter 4) with the outlook to apply them in the context of patient myoblasts.

Finally, viral vectors should be considered as a potential delivery mechanism of gene-editing reagents. Notable examples of viral vectors established in the field of CRISPR/Cas9 include Adenoviral, Lentiviral and Adeno-Associated Viral vectors. Adeno and Lentiviral vectors, arguably are more suitable vectors owing to their large capacities and ability to transduce proliferating and non-proliferating cells (Maggio et al. 2014; Maggio et al. 2016; Ousterout et al. 2015; Ran et al. 2015). The former has been used in the context of multiplexing and the latter has shown promise orchestrating NHEJ-mediated deletions, across the *DMD* gene in myoblasts. Lentiviral vectors were under development in this thesis for utilisation in such a manner but time restrictions precluded exploitation. Perhaps the most promising candidate is however the adenoviral vector as a dual transfection would enable the delivery of larger transgene cDNA repair template, thereby enabling the introduction of the full range of dystrophin variants highlighted in Chapter 5.

6.4.3.1 Enhancing the HDR DNA Repair Pathway:

The presence of a PCR amplicon across the left genome-repair template junction provides an indication that low levels of HDR had occurred within the heterogeneous $\Delta 45-52$ patient myoblast culture. In addition to enhancing gene-editing reagents delivery, numerous strategies could be employed that have been shown to successfully bias the DNA repair pathway utilised in genome repair. In reference to the DSB resolution, NHEJ and HDR pathways occur concurrently and competitively. The NHEJ pathway is the predominant repair mechanism used throughout the mammalian cell cycle, thereby limiting the extent scientists can exploit the high fidelity repair afforded by HDR. However, it is suppressed by HDR during late G2 and S phases where a sister chromatid becomes available as a template for HDR (Maruyama et al. 2016; Velic et al. 2015; Heyer et al. 2010). Thus to enhance precise genome repair and subsequent transgene integration, two strategies may be employed: upregulation of the HDR repair pathway or suppression of the NHEJ repair pathway

In consideration of HDR being constrained to the late G2 / S phase, this could be addressed by the modulation of cell cycle. By transiently locking the cells into certain phases of the cell cycle, this can result in an increase in HDR efficiency, from the conventional values cited in

literature of 10-17% to a level reported in HEK293T of 36% (Lin et al. 2014; S. Kim et al. 2014; Zuris et al. 2014). However, manipulation of the cell cycle has largely been explored with ssODN exogenous donors and further work utilising larger transgene donors is warranted in the case of CRISPR-mediated gene knock-in. Interestingly, an alternative approach has been trialled in studies, in which the Cas9 protein is synchronised to G2 and S phases. This provides an alternative complementary approach to cell synchronisation and was achieved by tethering the N-terminal 110 amino acids of geminin to Cas9 protein. This post-translational modification rendered the endonuclease susceptible to proteolysis via endogenous ubiquitination, a process that geminin would ordinarily partake. The expression of geminin oscillates becoming lowered during the M and G1 phases of the cell cycle and higher at S and G2 phases. This process was used to successfully limit an engineered Cas9 protein expression to phases in the cell cycle where HDR is elevated (Gutschner et al. 2016). The combination of both approaches can be utilised to achieve desirable HDR outcomes.

Further promising research relates to fusion of yeast Rad52 protein to Cas9 protein, and such an approach has been utilised in chicken cell backgrounds (Wang et al. 2017). The premise here is that the expression of this Rad52 fusion intermediate will improve Rad51 filament formation and enhance the rate of strand invasion and subsequent HDR by extension. In isolation the overexpression of Rad52 protein increased HDR by up to 37 fold in HeLa cells, a striking result which was improved upon to 50 fold when combined with nucleus localising HIV Tat11 protein (Di Primio et al. 2005; Kalvala et al. 2010). Whilst the Cas9-Rad52 fusion protein has only to date been trialled in a chicken background, further work aiming to develop this fusion construct to enhance HDR correction in mammalian cells, could prove promising to the field at large.

The second strategy employs use of NHEJ-inhibitors which aims to suppress the NHEJ pathway in favour of HDR, enabling HDR to be exploited in the context of genome engineering. The transient downregulation of the NHEJ pathway could enhance the MD1 transgene integration observed within $\Delta 45-52$ patient myoblasts. In consideration that a low level of transgene integration was found by PCR, this should be explored. Fortunately, the plethora of mediators used in the NHEJ DNA repair pathway to resolve a DSB yields a number of candidates that could be modulated. As such, four classifications of biomolecules have become apparent: Ku70/80 inhibitors, Nucleotide DNA-Phosphokinase (DNA-PK) inhibitors, small molecule DNA-PK inhibitors and finally DNA IV Ligase inhibitors (Velic et al. 2015). This list of drug targets is extensive and non-exhaustive, thus for simplicity, discussion will only pertain to DNA-PK inhibitors and DNA IV ligase inhibitors.

In brief, during the c-NHEJ, a DNA DSB is stabilised by Ku70/Ku80 dyad molecules; these in turn recruit DNA-PKs which undergo auto-phosphorylation potentiating a phosphorylation cascade, ultimately promoting DNA end processing and ligation by DNA IV ligase. A range of morpholino-containing molecules with the propensity to inhibit DNA-PKs have been identified: including vanillin, NU7026 and NU7441 (Durant & Karran 2003; Leahy et al. 2004; Calabrese et al. 2004). These all represent potential candidates to suppress the NHEJ pathways and enhance the HDR, improving the MD1 integration occurring in patient myoblasts.

This could prove particularly promising when utilising an ex-vivo gene correction strategy, during which cells could be transiently exposed. However, some inherent concerns to the therapeutic use of these reagents relate to drug target specificity since DNA-PK is a member of and thereby similar to Phosphatidylinositol-3-kinases, potential toxicity and ability to sensitise the cells to DNA damage (Velic et al. 2015). These parameters require further examination prior to being used to supplement a translational approach.

Interestingly, a recent investigation utilised Sc7, a ligase IV inhibitor alongside Cas9 in both mammalian cell culture and murine zygotes. This was shown to potentiate HDR mediated insertion, increasing HDR to a striking 19 fold in the case of mammalian cell culture (Maruyama et al. 2016). However, this study highlighted some major considerations that would need to be met prior to employing such a strategy for MD1 integration in patient myoblasts. Initially, Sc7 exhibits a cell dependent response with one mammalian epithelial cell used in the study A549, showing only a moderate 3 fold increase at a low dose of 0.01 μ M and attenuation of HDR at higher doses, attributed to toxicity. This is in contrast to the 19 fold increase of HDR showcased in at a high dose 1 μ M with MuJullSO, a melanoma cell line. Moreover, the transgenes inserted were considerably shorter than MD1 with inserts of 200-800bps being inserted flanked by arms of homology. These are considerable shorter than the MD1 transgene and zeocin selection cassette, which is in the range of 5.1kb. In addition, only in the murine zygotes were gene-editing and HDR reagents delivered alongside the Sc7; all mammalian cell lines were engineered to have doxycycline inducible sgRNA and Cas9 expression. In short, whilst none of the conventional contraindications were shown with this study e.g. DNA IV ligase-associated apoptosis in culture conditions or failure of lymphocyte maturation and a shift towards alt-NHEJ at the murine zygote level, the applicability of such an approach may prove challenging in the context of patient myoblasts.

Another more broad study with the use of a Traffic Light Reporter (TLR), this being a system that indicates whether a DSB has been resolved by NHEJ or HDR by alternative fluorescent reporter, cited a more modest 4-5 fold increase of HDR with Sc7 (Chu et al. 2015). The

disparity in fold increase in HDR reported is likely attributable to the fact that reagents were delivered by transient transfection as opposed to being constitutively expressed. Thus this is more likely representative of the HDR increases likely to be observed with transient transfection into patient myoblast if employed. In addition to small molecule therapy with Sc7, this investigation also highlighted a 5 and 8 fold increase in HDR upon gene silencing of Ku70 and DNA ligase IV co expression of Cas9 with Adenovirus 4 E1B55k and E4orf6 proteins (using a tethered protein) respectively. This coupled to another NHEJ inhibitor study undertaken in *Cryptococcus neoformans*, admittedly a fungal background, highlights the requirement to screen multiple drug candidates and formats in order to attain the most efficient NHEJ pathway suppression (Arras & Fraser 2016).

To conclude, there are a variety of potential drug targets that could interrupt the NHEJ or indeed enhance the HDR pathway improving the successful integration of the MD1 transgene. However, such an approach could prove challenging to implement and could prove detrimental to cell cultures. Nonetheless, this represents a potential manner to improve upon the current outcomes.

6.4.3.2 Further Refinement of Zeocin Doses

A final revision to this investigation may be to reduce the zeocin concentration applied during the selection process. Whilst higher doses of zeocin were required to cause cell death to untreated patient myoblasts, this represented a somewhat artificial situation that does not account for the effects of transfection related toxicity and sparse seeding of muscle cells. Moreover, from the microscopy imaging undertaken, in some instances an increased myoblast number is present in some of the treated conditions relative to mock samples. This may be attributable to inaccurate seeding of cells in the presence of zeocin, particularly as high cell number has been indicated to reduce the efficacy of the drug (Invitrogen 2017). The alternative explanation being the cells have retained the resistance cassette but the dose of zeocin surmounts the protective effect of the cassette.

6.5 Conclusion

In conclusion this chapter has demonstrated that the human *DMD* intron 1 locus, a region common to transcription of all full-length dystrophin RNA isoforms, can undergo Cas9-mediated HDR integration of the MD1 Δ Ex1 transgene. In HEK293T cells, integration of the MD1 Δ Ex1 transgene was not only detected upon PCR at 48 hours post transfection and also at two sequential stages of DNA harvest following long-term zeocin enrichment. Interestingly, the Dp427m exon 1 was demonstrated by RT PCR and sequencing to splice to the inserted MD1 Δ Ex1 transgene. However, a smaller than predicted MD1 protein was detected and was the predominate product, alongside low levels of full length MD1 by western blot of zeocin-

enriched cells. It is theorised that this may be as a result of either aberrant splicing or increased utilisation of the IRES site on exon 5. None-the-less this provides valuable evidence relating to a novel near universal strategy of splicing a dystrophin transgene to the Dp427m promoter.

Further refinement is required to achieve successful integration in patient myoblasts, although initial evidence of a positive single junction PCR appears promising. Furthermore, additional research is warranted to assess the functionality of the resultant truncated MD1 protein. This stated, with some refinement, the near-universal therapeutic strategy presented could achieve the endogenous spatial and temporal expression of dystrophin transgenes and hold promise for nearly all patient cohorts.

Chapter 7 General Discussion

This investigation aimed to and provided the first proof of principle, *in vitro* demonstration of MD1 Δ Ex1 integration downstream of full length endogenous dystrophin promoters and exon 1 elements. This near universal permanent gene-correction approach could provide a clinical strategy distinctive from gene augmentation. It is distinctive as it would not require repeated administration, serving to reduce the risk of adverse immunogenic events in patients. Secondly, being integrated downstream of and driven by full-length endogenous dystrophin promoters, it speculated that this approach is unique relative to other gene-editing strategies on the basis of its high patient applicability. In addition, by targeting downstream of endogenous promoters and exon 1 elements this approach is likely to provide spatial expression profiles of MD1 consist with dystrophin *in vivo*. Furthermore, the improved transcription associated with sequence optimisation predicts that microdystrophin protein expression levels would likely be equal to or higher than those observed for dystrophin physiologically. The latter sequence optimisation, examined for MD1, Δ 45-55 and full-length dystrophin will undoubtedly make achieving a therapeutic threshold of dystrophin in patients easier.

These findings could hold therapeutic promise for the field of DMD, particularly if a successful demonstration of this strategy's ability to ameliorate disease phenotype could be achieved in an animal model. As a consequence the discussion will be divided into considerations of the current research, enhancement of transgene integration approaches and translation of research into therapeutic setting; the outlook for each of the sections will be to examine the work undertaken in the context of the current gene-therapy field.

7.1 Current Research Considerations:

7.1.1 Identify and Resolve the Source of the MD1 Truncation

Molecular characterisation of polyclonal and monoclonal HEK293T-MD1 cells indicated that splicing between the Dp427m exon 1 and exon 2 of the integrated MD1 transgene was successful, as confirmed by an RT-PCR across the exon junctions performed on mRNA lysate. However, when total protein was subject to analysis by western blotting with MannEx10-11c, only a faint band of full-length MD1 was observed in monoclonal cultures, being absent or below detection in preceding polyclonal cultures. Whilst, this served to validate that splicing between Dp427m Exon 1 and MD1 Exon 2 was occurring, the full-length MD1 protein was largely overshadowed by a high intensity band corresponding to a truncated MD1 protein estimated to be 117kDa in size.

It is speculated that the source of this truncation is attributable to aberrant splicing between the Dp427m exon 1 and MD1 transgene or an internal deletion of MD1 cDNA. In light of this, *In silico* predictions of splice sites strength using human splice finder, were re-examined (Desmet et al. 2009). Notably, whilst the artificial β -globin splice acceptor used for the repair template was ranked as the strongest splice acceptor, relative the native exon 2 splice acceptor, overall; a cluster of splicing enhancers were observed downstream of this, surrounding moderately ranked splice acceptors. Consequently, it is predicted that this could provide a favourable environment for splicing and a source for competition for the β -globin splice acceptor. Empirical evidence of this could be attained by undertaking RT-PCRs spanning the sequence between Dp427m exon 1 and predicted downstream splice acceptor candidates. If aberrant splicing was found to be the cause of truncation, the inclusion of native dystrophin exon 61 or 64 splice acceptors in place of the β -globin, may overcome this problem particularly as these were observed to be the strongest in triple-transplicing experiments (Lai et al. 2005). To eliminate the alternative possibility of internal deletions within the integrated MD1 cDNA, PCRs tiling the whole MD1 transgene should also be undertaken. In addition, RT PCRs could also be undertaken to assess the mRNA and findings could be further validated with the use of antibodies across the MD1, if a specific deletion is suspected.

Identifying the source of the truncation is of paramount importance to the investigation, as the nature of the 5' deletion may / may not influence the functionality of the MD1 construct (Banks et al. 2007; Wein et al. 2014). If the functionality, is in anyway impaired it could undermine the strategy proposed and thus requires resolution.

7.1.2 Assess the Region of Homology Identified between Human and Mouse Intron 1 for Potential Functionality:

The genomic region targeted at the 5' end of the *DMD* gene in this study showed high levels of sequence similarity between human and mouse. This, in combination with the unanticipated truncation of the integrated MD1 transgene, prompted further sequence analysis. It was speculated that the target site could represent a non-coding sequence with function. Interestingly, studies have indicated that only 1 -2% of the human genome codes for proteins, yet studies estimate that approximately 10% of the genome conserved (Lander et al. 2001; Wright & Sanjana 2016; Canver et al. 2017). Thus a potential expanse of non-coding DNA retains functions beyond the conventional mRNA and protein synthesis.

To investigate further, the human 345 nucleotide region of homology with 100 nucleotides of adjacent flanking sequence was subject to pairwise sequence analysis against the genomes of the following species: *Canis Lupis Familiaris*, *Macaque Fascicularis*, *Rattus Norvegicus* and *Sus*

Scrofa (Figure 7-1, Figure 7-2). Intriguingly, all species showed 82% or higher sequence similarity relative to human, with alignments ranging from 311 -351 in length (see Table 7-1).

Further scrutiny of the sequence highlighted that the single nucleotide polymorphisms (SNPs) present within the nucleases designed appeared to be conserved. This is particularly exemplified for Guide 1 and 4, where respective 12T-C and 9A-G SNPs, were conserved across 3 of the 4 species examined; the one exception being the *Macaque Fascicularis*. Furthermore, the rule is partially retained for guide 3 where Mouse and Rat harbour the same 9T-C, 20C-A polymorphisms. All SNPs noted followed a PAM proximal to distal nomenclature.

In light of this evidence, the sequences were examined using the UCSC browser for DNase I Hypersensitivity and histone acetylation/methylation; the presence of which are considered to be biochemical hallmarks of functional non-coding regions (Kent et al. 2002; Wright & Sanjana 2016). Despite Encyclopedia of DNA Elements (ENCODE) providing a high level of genome-scale annotation (Dunham et al. 2012), which UCSC utilises, no sequence annotation was present for the site of interest (Figure 7-3). This stated, the tool confirmed the high conservation of the sequence among the species in question. In addition short stretches of sequence to the left and right of the region examined, were indicated to have patents pending, due to predicted regulatory function. Although, there was no genomic annotation, this does not necessary disprove the presence of a functioning non-coding element. A recent study that produced Nanog, Rpp25, TdGF1 and Zfp42 GFP fusions, in mouse embryonic stem cell lines, identified several regulatory elements that were not indicated by annotation of biochemical hallmarks (Rajagopal et al. 2016).

Collectively, these findings indicate experimental assessment of the sequence to delineate potential function and suitability of this site for gene-editing is warranted. Loss of function studies, either by the destruction of the local sequence or steric hindrance of transcriptional machinery, would likely aid in this process (Canver et al. 2017). Wild-type myoblasts and those treated for NHEJ-mediated removal of the region of homology could be differentiated in parallel. If a disparity in dystrophin expression is observed between the two groups it would suggest this sequence retains a regulatory function. The exact motif within this region could then be ascertained by tiling sgRNAs across the site and causing defined sequence modification with micro-insertions and deletions. Alternatively, tethering a fluorescent marker to the 3' end of the *DMD* gene would allow for an examination of transcriptional activity, with catalytically inactivated dead Cas9 (dCas9) tethered to transcriptional activators or repressors (Gilbert et al. 2013; Chen et al. 2013). The level of fluorescence and fluorescence intensity would serve as parameters to assess whether the site functions to modulate

transcription of the *DMD* gene. Finally, of note a novel isoform Dp412e was recently elucidated downstream from the region of homology; on this basis one further speculation could be that this site could function as a promoter to the *DMD* gene (Massouridès et al. 2015).

Species	Sequence Similarity to Human	Gaps	Length of Alignment	Common Nucleases (Guide and TALEN)
Mouse	84.9%	5.5%	311	2 and TALEN
Dog	82.0%	1.9%	322	2(X19) and TALEN
Macaque	98.0%	3.1%	351	1, 2, 3 and TALEN
Rat	83.9%	6.4%	311	2 and TALEN
Pig	82.0%	4.0%	350	2(X19), 3 and TALEN

Table 7-1 : A table showing the sequence similarity attained by rigorous local alignment of the human region of homology sequence, identified in *DMD* intron 1, against a multitude of species.


```

Human      40 AGTTTATCCATTTTCGAAACATCTAA-TATTACATG----AGAATTACTT      84
          |||||..|||||.|| ||||| |||...|| |||||.|||.|
Rat       11796 AGTTTACTCATTTTCCAA-CATCTAAATATTGTGTGTGAGAGAATCTACAT 11844

          85 TTCTCATTAGATTACATAAACATCATGTTAAAAATGCAATATGTACATGTG      134
          |||.|||||.|||||.|||.|||.|||||.|||||.|||||.|||||.
Rat       11845 TTCTTATTAGATTACAAAGAAATCCTGTAAAAATGCAATATGTACATGTG 11894

          135 ACAACTGATCTTTTGGGAAAAAAGTCCATGAATCACTCCTGTAAACAT      184
          |||.|||||.|||||.|||.|||.|||||.|||||.|||||.
Rat       11895 ACAAGTATCTTCTGGGAAAA-----GTTGATGAGTCACTTCTGTAAACAT 11939

          185 TTGGAATTAATTCATTAACCAAAATAATATCATGCTTCACTTTTATAT      234
          |||||.|||||.|||||.|||||.|||||.|||||.|||||.|||||.
Rat       11940 TTGGAATAAATTCATTAACCAAAATAATACCATGCTTCAATTTTATAT 11989

          235 TTTGCTTCCTTAGCATCTGGTCTTAGTCTGTCTAATTACATTATAATGG      284
          |||||.|||||.|||||.|||||.|||||.|||||.|||||.|||||.
Rat       11990 TTTGCTTCCTTAGCATCTGGTCTTAGTCTGTCTAATTGCATTATAATGG 12039

          285 CTCTGTGTACACAGTGTCTAGAAAGATTTTTTGGGGGGGGGGCAAGA      334
          .|||||.|||||.|||||.|||||.|||||.|||||.|||||.
Rat       12040 TTCTGTGTGCAACAGATGCTAGAAAGATTTCTTTTGG-----CAAGA 12080

          335 ATGATGATTTT      345
          |||.|||.|||||
Rat       12081 ATGGTAATTTT 12091

```

```

Human      1 CTATTGACGAATTATTCATCTACCCGTGGAAGTGTCTGCAGTTTATCCAT      50
          |.|||||.|||.|||||.|||||.|||||.|||||.|||||.|||||.
Pig       11754 CCATTGGCAAGTTATTCATAATCCTTTGAGACTGTCTACTGTTTAGTCAT 11803

          51 TTCGAAACATCTAATATTACATGAGA----ATTACTTTTCTCATTAGAT      96
          ||..|||.|||||.|||||.|||||.|||||.|||||.|||||.
Pig       11804 TTTAAATATCTAATATTCTGTAAGAGGCAATTTACTTTTCTCATTGTAT 11853

          97 TACATAAACATCATGTTAAAAATGCAATATGTACATGTGACAACATGATCTT      146
          |||.|||.|||||.|||||.|||||.|||||.|||||.|||||.
Pig       11854 TACGTAGGCATCATGTTAAAAATGCAATATGTACATGCAGAAACTGCTCTT 11903

          147 TTGGGAAAAA-----GTCCATGAATCACTCCTGTAACATTTGGAAATAAA      195
          |..|||||.|||||.|||||.|||||.|||||.|||||.|||||.
Pig       11904 TAAGAAAAAAGGCCATGAATCATTCTTGTAAACATTTGGAAATAAA 11953

          196 TTCATTAACCAAAATAATATCATGCTTTCATCTTTTATATTTTGCTTCCTT      245
          |||||.|||||.|||||.|||||.|||||.|||||.|||||.|||||.
Pig       11954 TTCATTAACCAAAATAATATCATGCTTTCATCTTTTAAATTTTGCTTCCTT 12003

          246 AGCATCTGGTCTTAGTCTGTCTAATTACATTATAATGGCTCTGTGTAC      295
          |||||.|||||.|||||.|||||.|||||.|||||.|||||.|||||.
Pig       12004 AGCATCTGGTCTTAGTCTGTCTAATTGCATTATAATGGCTCTGTGTGC 12053

          296 AACAGTTGCTAGAAGATTTTTTGGGGGGGGGGCAAGAATGATGATTTT      345
          |||||.|||||.|||||.|||||.|||||.|||||.|||||.
Pig       12054 AACAGATGCTAGAAGATATTTT-----CAAGAATGGTAATTTT 12093

```

Figure 7-2 The human region of homology identified within the DMD common intron 1 aligned against consensus DMD sequences of *Rattus Norvegicus* and *Sus Scrofa*, using a rigorous local alignment tool ‘Emboss Matcher’.

The top local alignment is between human and *Rattus Norvegicus*, the bottom is between human and *Sus Scrofa*. Notice the region deemed to be highly homologous between mouse and human is highlighted in blue for both of the sequences retain high similarity.

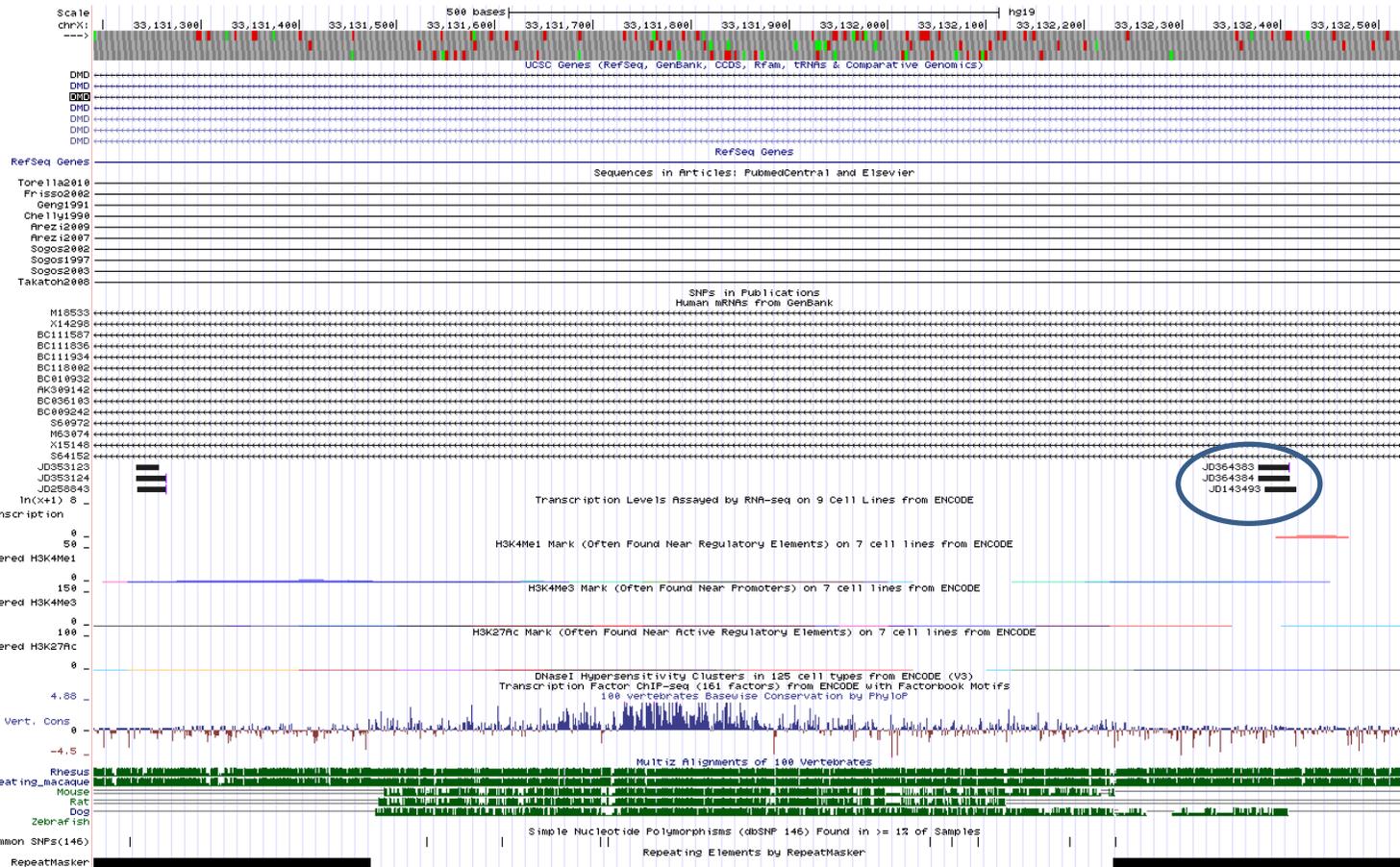


Figure 7-3 An examination of genomic annotation at the region of homology identified in DMD intron 1 using the UCSC genome browser.

Notably, this sequence has no histone acetylation tracts and DNaseI hypersensitivity annotation (All tracts labelled on the left hand-side), which are regarded as biochemical markers of accessibility for functional non-coding DNA sequences. In addition high levels of conservation are noted for rhesus, macaque, rat and dog sequences; as evidenced by the green multiz alignment and vertebrate conservation. Finally, sequences within the blue semi-circle have been patented as sequences with regulatory function.

7.2 Enhancement of Transgene Integration Approaches

7.2.1 Advances in CRISPR Guide Design: A Prospect for Improved CRISPR guides

At the initiation of this project in 2013 the prokaryotic adaptive immune system CRISPR/Cas9 had just been repurposed for mammalian gene editing (Jinek et al. 2013; Mali, Yang, et al. 2013; Cong et al. 2013). This simple gene-editing platform appeared attractive over predecessors being an RNA rather than protein guided customisable endonuclease (Mali, Esvelt, et al. 2013). One major limitation to the use of this system, at the start of the investigation, was the lack of specificity and efficacy-governing rules to guide design. The variable efficacy of guides and lack of understanding thereof, was certainly a hindrance to the computational design of gene-editing reagents. This was exemplified by Liu et al's study, which demonstrated that 41% of the 218 guides tested failed to show detectable on-target mutagenesis in Neuro-2A-cells, with mismatch detection assays (Liu et al. 2016a). Interestingly, studies indicated that guides targeted to the same genomic locus, showed variable efficacy. One example of this was Mali et al's paper, which demonstrated two guides targeting the AAVS1, separated by 19 nucleotides, showed an approximate two-fold difference in gene-modification frequencies (Mali, Yang, et al. 2013).

These findings lead to speculation that the nucleotide composition of a guide could influence the overall efficacy of the Cas9 system. This prompted a plethora of investigations to identify sequence determinants that govern guide efficacy; one milestone study screened 1841 guides targeted to 6 cell surface markers. The knock-out of markers was examined by Fluorescence Activated Cell Sorting (FACs), with the use of antibodies; the extent of knock-out was used as a proxy for guide efficacy. This study amongst many others confirmed that nucleotide composition had an influential role on on-target mutagenesis; this led to the derivation of sequence rule-sets, which served to inform computational prediction tools.

The sequence parameters deemed beneficial to a high efficacy guide included: i) The presence of guanine in position 1 of guide, ii) The presence of guanine in position 20, iii) Adenine residues positioned to the middle of the guide, iv) A cytosine in position 17 at the RNA-DNA cleavage junction and v) A poly GGGG runs in In-vitro transcribed guides. Those deemed detrimental included: i) A cytosine in position 3 of the guide and ii) A poly TTTT run.

Interestingly, numerous studies indicated that the presence of a guanine nucleotide in position 1 of guide designs was related to higher efficacy (Doench et al. 2014; Moreno-Mateos et al. 2015; Xu et al. 2015; Chari et al. 2015b; Liu et al. 2016a). This was attributed to two reasons,

guanine being a pre-requisite for correct transcription from U6 RNA III polymerase promoters and the 5' nucleotide enabling a stronger RNA: DNA hybridisation between guide and genome (T. Zhang et al. 2017; Moreno-Mateos et al. 2015). The requirement of the guide to have a certain 5' composition to enable efficient transcription was also expanded to in-vitro transcription studies in zebrafish using T7 and Spc6 promoters, the former was observed to have a bias to a 5' 'GG' and the latter showed comparable activity between guides with a 5' 'GG' or 'GA' sequence (Gagnon et al. 2014). The ability to transcribe an expanded repertoire of guides, served to indicate that the Sp6 was a superior promoter. As not all guide sequences identified computationally begin with a guanine residue, Cho et al examined the effect of appending guanine residues to guide sequences. The findings of the study suggest that the substitution of the first nucleotide of the guide to form GX19 was superior to simply appending 'GG' at the 5' end (Cho et al. 2014b).

In contrast to the guanine at position 1, indicated to be beneficial to transcription, the presence of a poly TTTT sequence is inhibitory to RNA III polymerase activity. It is speculated that this poly TTTT run promotes the premature termination of mRNA transcription. Interestingly, the first generation chimeric single guide RNA, used a TTTT motif to connect the crRNA-tracrRNA; this was later modified to a GAAA- tetraloop by Chen and Dang et al to serving to enhance on-target mutagenesis (Jinek et al. 2013; Dang et al. 2015; Chen et al. 2013). As a consequence guide sequences should avoid the inclusion of this motif as transcription of the guide may be variable or indeed unsuccessful.

Remaining sequence determinants are thought to influence secondary structures formed via intermolecular interaction upon hybridisation of the guide and genome/ RNA: DNA, or intramolecular interactions within the guide itself. In the case of intermolecular interaction, it is speculated that the guanine at position 20 of the CRISPR guide, facilitates the unwinding of genomic DNA and the formation of an R-loop structure between RNA and DNA. This R-loop formation is indicated to stabilise the RNA-DNA duplex, thereby facilitating more efficient genomic modification (Anders et al. 2014; Malina et al. 2015; Liu et al. 2016a). Further general trends indicated to improve RNA: DNA duplex interactions or enable greater DNA cleavage include: the avoidance of cytosine in position 3 (Doench et al. 2014; Moreno-Mateos et al. 2015), adenine nucleotides positioned to the middle of the guide (Doench et al. 2014; Moreno-Mateos et al. 2015; Xu et al. 2015) and the presence of cytosine in position 17 (Xu et al. 2015).

The relationship between the presence of a poly GGGG sequence and guide efficacy proved more difficult to determine. Wong et al indicated that the presence of a poly GGGG sequence, not only reduced the oligo yield upon synthesis, but also formed an unfavourable G-quad

secondary structure, hindering the formation of RNA: DNA duplexes and subsequent DNA cleavage (Wong et al. 2015). In contrast to this Moreno, indicated that poly GGGG sequences improved guide RNA stability and whilst Cas9 loading was slower, when the guide was associated with the genome the poly GGGG served to stabilise the DNA:RNA duplex (Moreno-Mateos et al. 2015). The increased stability of the guide itself and of the RNA: DNA duplex was predicted to improve guide efficacy.

The disparity of these findings was resolved upon the examination of the experimental designs; Wu based his findings upon the highest efficacy U6 transcribed guides from Doench's 2014 study, whereas Moreno-Mateos produced guides by *in vitro* transcription. It was speculated that the latter study, which injected a defined concentration of guide mRNA, placed more emphasis upon guide stability and mRNA degradation; whereas, guides delivered in a U6 plasmid format are continually transcribed. In light of this the presence of a poly GGGG sequence and a higher GC content, appear to be more favourable to the formation of secondary structures, enabling a greater persistence of mRNA guides post-delivery and thus a higher efficacy. These findings indicate that rule-sets of guide composition should be considered as dynamic and influenced by the experimental situation in which they are employed.

Intriguingly, Moreno's indications that *in-vitro* transcribed guides should have a higher GC content and/ or poly GGGG sequences were corroborated by studies, that injected mice oocytes with guide mRNA (Haeussler et al. 2016). As a consequence of this Doench's rule sets are argued to define U6 transcribed guides, whereas experiments aiming to produce guides by *in-vitro* transcription methods should adhere to Moreno's guide design parameters. A range of computational tools have since been designed adhering to parameters highlighted in the aforementioned studies, a few examples include: CRISPRScan (Moreno-Mateos et al. 2015), CRISPOR (Haeussler et al. 2016), Benchling (Benchling 2015) and SgRNA designer (sgRNA Designer 2017).

Importantly, sequence determinants are not the only parameter that influences CRISPR/Cas9 guide efficacy, the microenvironment of the genome to which the guide is targeted, is also an important consideration. Firstly, Protospacer Adjacent Motif (PAM) recognition site composition is important 'NGGX', PAMs in which N is a guanine and X a thymine nucleotide have been shown to produce higher level efficacy; this is likely attributable to favourable R-loop formation (Doench et al. 2014). Further studies have also indicated the importance of chromatin accessibility and PAM multiplicity. Biochemical hallmarks of high DNase I hypersensitivity, Histone 4-lysine-3, and low CpG methylation were all found to coincide with

the design of higher efficacy guides (Chari et al. 2015a; Wong et al. 2015; Xu et al. 2015; Liu et al. 2016a). Furthermore, a high frequency of PAM at the genomic target site is associated with poor gene-modification outcomes; this is speculated to be due to the Cas9 nuclease having to undergo a lot of genomic surveillance before target site recognition (Malina et al. 2015; Jinek et al. 2014).

In addition to all of the efficacy algorithms coded above, all of the computational *in silico* guide prediction tools stated, also use specificity governing rules. The first guide specificity rule sets were derived from four main studies (Hsu et al. 2013a; Pattanayak et al. 2013; Mali et al. 2013; Fu et al. 2013). These studies utilised a mixture of transcriptional studies, with the catalytically dead Cas9 (dCas9), and direct cleavage studies. Hsu et al's study represents a milestone to the understanding of CRISPR guide specificity. This study highlighted: i) The nature of the nucleotide mismatches e.g. A-G ii) The position of the mismatch being PAM proximal or PAM distal, ii) The number of mismatches occurring and whether they were concurrent or dispersed throughout the guide are all parameters that influence a guide's propensity to cleave an off-target site. These findings were recently used by Doench to produce a Cutting Frequency Determination (CFD) score, whereby the propensity for a guide to cut on-target relative to off target is presented as a ratio. The closer this ratio is to the value 1, the more optimal the predicted guide design (Doench et al. 2016).

The research presented here, highlights the extent to which the understanding of CRISPR guide design and targeting has advanced since the beginning of this investigation. This improved understanding of guide composition, genomic microenvironment and specificity, would have undoubtedly influenced the assessments of target sites and the guides selected for use within this study. Notably, 8 of the 9 total guides designed for this study, had neither a guanine at position 1 or 20, requiring the 5' guanine to be appended to enable transcription. Furthermore, the guide identified as common to human and mouse sequences, and thus identified as an ideal candidate for development, harboured a poly TTTT region. Consistent with the rule sets, the common guide /human guide two, was associated with relatively low activity when transcribed by the U6 RNA III polymerase promoter. Moreover, this guide produced low mRNA levels following *in vitro* transcription precluding its use in an experimental setting. In contrast human guide 1 which was shown to be the most efficacious guide throughout this investigation harboured some beneficial sequence determinants. These included: a cytosine in position 17, centrally located adenine nucleotides and an adenine nucleotide at position 1, which is most favoured after guanine.

Interestingly, when the region was analysed again with the use of CRISPOR, a more advanced *in silico* guide prediction software (See Chapter 3); all guides attained threshold score of 50 for specificity, but only guide 1 scored a suitable efficacy score. This guide enabled the integration of the MD1 transgene downstream of endogenous promoters, attaining the primary aims of this investigation. This stated when the method was transitioned to patient myoblasts, integration was identified at low levels. This could be speculated to be due to a number of parameters including reduced transient transfection efficiency, limited introduction of DSB and even an altered level of HDR repair occurring within this cell line.

Nonetheless, human guide 1 remains a candidate for continued investigation of the genetic-correction strategy proposed based on the predicted specificity profile and the reduced propensity for off-target effects. Future work would aim to scrutinise guide 1 more thoroughly and investigate supplementary methods to improve HDR outcomes. In addition, the common intron 1 sequence should continue to be analysed with more recent guide design tools. On identification, target sites within the genome should be assessed on the basis of biochemical hallmarks of DNAI hypersensitivity, acetylation and the presence of SNPs. If both profiles appear favourable, the site should be sequence and mismatch detection assays undertaken. The combination of these two strategies would ensure the most successful outcomes for this gene-editing strategy moving forward.

7.2.2 Improved Homology Directed Repair and the Prospect for Non-Homologous End Joining-mediated Integration Approaches

Numerous strategies have been presented throughout this investigation with regard to enhancing the integration of the MD1 cDNA at the DMD intron 1 site. Briefly, these include: the inhibition of the NHEJ pathway (Maruyama et al. 2015), the synchronisation of the cell cycle at G2 and S phases (Lin et al. 2014) and overexpression or enhancement of (HDR) mediatory proteins (Di Primio et al. 2005; Wang et al. 2017). These considerations were made due to the inherently low efficiency of a HDR-based strategy, due to its limitation to late G2/ S phases of the cell cycle (Heyer et al. 2010). Importantly the modification of the exogenous repair template, to include endonuclease target sites outside the arms of homology, has also been shown to enhance the occurrence of HDR. Although the exact mechanism of how this works has not been delineated, it is speculated that the cleavage of the exogenous repair template *in situ* serves to improve integration outcomes. Moreover, dispensing with extensive homology arms, introducing endonucleases at the termini of short nucleotide stretches of microhomology, or directly adjacent to the cDNA transgene, has been shown to enable NHEJ-mediated integration (Sakuma et al. 2015; Suzuki et al. 2016). As NHEJ is generally regarded as

a high efficiency process not constrained by the cell cycle, this finding could serve to redefine gene-editing approaches.

This investigation used a classical HDR approach, producing a targeted DSB in the genome and utilising an exogenous cDNA repair template to facilitate the integration of the MD1 transgene from Exon 2 onwards. Importantly, the cDNA repair template conventionally requires arms of homology ranging from 0.2-1kb upstream and downstream of the DSB, that are completely isogenic to the genome (Byrne et al. 2015). However, the length of homology arms increases to 2kb, when a HDR approach is used in the context of induced Pluripotent Stem Cells (iPSCs) (Byrne et al. 2015; Zhang et al. 2017). These extensive arms of homology increase the size of the exogenous repair template considerable and make delivery strategies more challenging.

In this instance 1kb arms of homology were included directly adjacent to the cDNA cassette, to maximise successful integration outcomes in HEK293T and immortalised patient myoblasts harbouring deletions of exons 45 - 52. Interestingly, during the production of this thesis, Zhang et al demonstrated that the inclusion of endonuclease genomic target sequence, directly adjacent to arms of homology increased cDNA integration efficiency. Reportedly, this approach produced 2 and 5 fold increases in integration efficiency, when using an mCherry transgene in HEK293T and iPSCs cultures respectively (Zhang et al. 2017). This study also indicated optimal levels of HDR were achieved using 600-900bp arm of homology flanked by cleavage sites, using this double cut system. The utilisation of this approach could streamline exogenous repair template designs and improve HDR outcomes in iPSC and other therapeutic cell candidates.

Zhang's study highlights that modification of the exogenous repair template used in this investigation could also improve HDR efficiency. Before considering the inclusion of restriction sites, the arms of homology within the exogenous repair template require refinement. Firstly, the arms of homology were constructed from consensus sequences to ensure timely de novo synthesis could occur. However, if the arms of homology are not completely isogenic to genomic sequences adjacent to the DSB, this can reduce integration efficiency (te Riele et al. 1992; Zhang et al. 2017). To improve the current exogenous repair template construction, PCR designs should be tiled across the 1kb either side of the cleavage site; the region should then be amplified and sequenced in $\Delta 45-52$ Immortalised Patient myoblasts. This would aid the identification and elimination of any sequence polymorphisms and improve HDR outcomes. Secondly, there is a 49 and 32 nucleotide gap, between the left and right arms of homology and the guide 1 cleavage site respectively. This is present as the highest efficacy guide at the region of homology had not yet been identified, thus the arms of homology were designed to flank the highly homologous 100 nucleotides. Numerous studies have shown that arms of

homology designed in closer proximity to the DNA cleavage site are associated with improved HDR outcomes (te Riele et al. 1992; Byrne et al. 2015; Zhang et al. 2017). Thus arms of homology should be extended to the Guide 1 cleavage site; this could be performed using extended self-annealed oligo cloning. In spite of these limitations the current exogenous repair template enabled successful integration of the MD1 exon 2 onwards transgene at high enough efficiency for HEK293T cells to survive zeocin selection. Although evidence indicated some MD1 integration occurred in immortalised patient myoblasts, the culture was unable to survive the selection process; these strategies alone or in combination could improve the HDR enabling the persistence of MD1+ve-Myoblasts.

7.2.2.1 A Non-Homologous End Joining Approach to Integration

In addition to the modification of arms of homology, improving the first generation repair template, which encoded a Cyan Fluorescent protein (CFP) with the initiating methionine removed between FseI and NotI restriction sites, could provide a model system to examine improved integration approaches. It was originally speculated that successful integration of this first generation exogenous repair template into human DMD intron 1, would yield a partial dystrophin-CFP fusion protein driven by the Dp427m promoter; thereby providing a visual indication of HDR. Provisional experiments yielded no such visual indication; this could be attributed to a variety of reasons: the fusion protein adopting an unfavourable conformation and being degraded, aberrant splicing hindering CFP fluorescence and in the absence of a promoter driven CFP, the prospect of a dysfunctional fluorescent protein cannot be excluded.

It would be beneficial to simplify the first generation repair template, to encode only a fluorescent marker driven by a CMV or Spc512 promoter, between left and right arms of homology. This construction would also enable a visual indication of integration and an indirect measurement by FACS analysis. Such a template could prove instrumental in assessing the optimal HDR enhancement strategies prior to translating such approaches to the MD1 exogenous repair template. Furthermore, due to the low efficiency of the HDR pathway, there is a growing trend in the field to exploit the NHEJ and Micro homology-Mediated End Joining (MMEJ) DNA repair pathways for transgene integration (Heyer et al. 2010). The main difference between these two methods being the proximity of genomic endonuclease restriction sites encoded on the exogenous repair template (Sakuma et al. 2015; Suzuki et al. 2016). These alternative approaches could be trialled using this CFP with the outlook to transition a successful strategy to the exogenous MD1 template, without prolonged antibiotic selection.

Two milestone studies that demonstrated the utility of the NHEJ DNA repair pathway in mediating cDNA integration was Obligate Ligation-Gated Recombination 'ObLiGaRe' and Homology Independent Targeted Integration 'HITI' (Maresca et al. 2013; Suzuki et al. 2016). These approaches are reliant upon the inclusion of an endonuclease recognition site in reverse orientation directly adjacent to the cDNA for integration. In this manner the exogenous repair template and the genomic DNA locus are simultaneously cleaved, and the cDNA 'ligated' into the genome. Notably, ObLiGaRe utilised Zinc Finger Nucleases (ZFNs) whereas HITI used CRISPR/Cas9 but the underlying strategy remains the same. Interestingly, ObLiGaRe demonstrated that following a modification of the AAVSI site to contain high efficacy ZFN target sites, integration of a 15kb inducible transgene was possible (Maresca et al. 2013). In contrast, HITI exploited the propensity for NHEJ to occur in both proliferating and post mitotic cells; using this approach to target and correct rat neurons of a genetic defect causing retinitis pigmentosa (Suzuki et al. 2016). During the latter method, constructs below 5kb were cloned into AAV vectors and injected into the subretinal space of rat models.

The research presented has some important implications to furthering this investigation; i) It is possible through this method to introduce a 15kb cDNA ii) Utilising an NHEJ approach this method is able to target proliferative and post-mitotic cells iii) This approach also circumvents the requirement of HDR for extensive arms of homology and occurs throughout the cell cycle. Utilising this NHEJ-mediated approach for this investigation, co-transduction of AAVs encoding *Staphylococcus aureus* or Split *Streptococcus pyogenes* with AAV-MD1 could be envisaged (Wright et al. 2015; Ran et al. 2015). Furthermore, this system could facilitate the integration of the larger transgenes such as Quasi $\Delta 45-55$ and full length dystrophin, which still remain the potential gold standard in therapeutic application.

Whilst this approach has many advantages over conventional HDR, this method is arguably less precise than the former. If a plasmid repair template is utilised, it would necessitate the removal of its bacterial backbone and production in a minicircle format (Kay et al. 2010). This is due to the integration of the plasmid backbone being as likely as the transgene of interest. A further guard against this would be to use negative selection, by encoding a toxin in the plasmid backbone such as Herpes Simplex Virus -Thymidine kinase (HSV-TK). If this is integrated into DMD intron 1 rather than the MD1 transgene, then it can process ganciclovir if is exposed to, producing a toxin that results in cell death (Elion 1982; Schwartz et al. 1991). In addition, there is an increased risk of multiple insertion events in a head-to-tail format at the genomic locus of interest (Suzuki et al. 2016). A final concern is related to the orientation of transgene integration particularly as spCas9 produced a blunt flush genomic cut. One strategy proposed to overcome this is to use bacterial orthologues of Cas9 that produce DNA cleavage

with cohesive overhangs, such as Cpf1 (Zetsche et al. 2015). Alternatively, 10-25 nucleotides of sequence isogenic to the DNA adjacent to the cleavage site could be included between the transgene and endonuclease cleavage site. This MMEJ utilises DNA 3' end resection during DNA repair to generate 5' overhangs; this could provide 'directional ligation' of the MD1 transgene into the genome (Sakuma et al. 2015). The approaches presented are promising and exciting alternative to HDR integration but warrant more extensive research and application.

7.3 In Vivo Translation:

7.3.1 Cell Engraftment Approaches in Pre-clinical Animal Models and DMD Patients.

To further refine the near universal gene-editing strategy presented, an in-vivo demonstration of safety and efficacy is required in pre-clinical DMD animal models. This could be achieved through an *ex vivo* genetic modification and subsequent engraftment strategy, in which genetically modified muscle precursor cells are administered to dystrophin deficient mice. Since the first demonstration of myoblast engraftment in the 1970s, cell based therapies have focused upon the treating muscular dystrophies through the administration of mono-nucleated precursor cells (Partridge 1978). The premise for this therapeutic approach being wild type or genetically corrected mono-nucleated precursors can either: fuse to each other producing mature unaffected myofibres or fuse to regenerating patient muscle fibres. In the case of the latter, the donor nucleus enters the syncytial myofibre, and enables the synthesis of donor derived proteins overcoming the genetic mutation; this was first exemplified through injection of normal muscle precursor cells into the *mdx* mouse. It was demonstrated by these studies that the injected normal muscle precursor could fuse with pre-existing or regenerating *mdx* muscle fibres, rendering many of them dystrophin positive and rescuing from the dystrophin deficiency (Partridge et al. 1989).

In the context of the current study, the genetically modified muscle precursor cells would contain the MD1 transgene from exon 2 onwards, integrated downstream of endogenous promoters. Importantly, the reconstituted MD1, is unlikely to restore all dystrophin protein associated functions; particularly as it lacks spectrin-like repeats 16 and 17 that neuronal Nitric Oxide Synthase (nNOS) is indicated to associate with (Lai et al. 2009). This stated, the MD1 transgene is speculated to contain the minimal necessary protein domains to restore structural associations between the F-actin and Dystrophin Associated Protein Complex (DAPC) (Harper et al. 2002). Additionally, its use in gene augmentation approaches in mouse and canine preclinical models have demonstrated amelioration of dystrophic pathology (Foster et al. 2008; Koo et al. 2011; Le Guiner et al. 2017). Moreover, muscle derived cells were, genetically

corrected via a lentiviral transduction using a lentivirus encoding minidystrophin cDNAs and engrafted into TA muscles of *mdx* nude mice. This showed some success but was unable to achieve high enough expression levels to affect physiological function (Meng et al. 2016). This study used lentiviral transduction to attain transgene integration, native cDNA sequences of mini-dystrophin constructs and a muscle specific promoter different from the endogenous dystrophin promoter. It is speculated that the advantages of the approach presented being: targeted integration, sequence optimisation and expression from the endogenous dystrophin promoter, all represent parameters that may permit more favourable outcomes.

Pertinent to the success of such an approach is the careful selection of muscle precursor cells to genetically correct and engraft. On a practical note GMP guidelines indicate cell candidates: should be readily available, easy to isolate /amplify and have a homing capacity to the site of muscle regeneration (Negrone et al. 2016). Furthermore an ideal cell candidate for the treatment of DMD, should also be: suitable for systemic delivery, as this would allow body wide correction rather than a single muscle post-intramuscular injection and be able to reconstitute the stem cell pool (Briggs & Morgan 2013). Currently, there are a plethora of cell candidates continually being identified and trialled, arguable the most examined are: myoblasts, satellite cells, pericytes/mesoangioblasts, CD133+ cells, Human Embryonic Stem cells and Induced Pluripotent Stem Cells.

In the context of preclinical studies, an immune tolerant, dystrophin deficient mouse would provide the best model to examine this strategy. The recently developed Rag2^{LL2rb}DMD^{-/-} mouse, that lacks T, B and NK cells and has dystrophic phenotype could prove to be one such suitable candidate (Vallese et al. 2013). Alternatively, the NSG-*mdx*^{4cv} mouse could also be utilised, which illustrates dystrophin deficiency with a reduction of NK cells (Arpke et al. 2013). Both of these models would prove suitable for allogeneic engraftment approaches, this being where the donor cellular material is distinct from the host (Briggs & Morgan 2013). As a consequence, if improved viability and persistence of immortalised Δ45-52 patient myoblasts could be attained following genetic modification, these cells could be expanded and injected intramuscularly. Whilst it is generally noted that myoblast engraftment is associated with low efficiency, owing to low survival and migration post injection (Lipton & Schultz 1979; Morgan et al. 1987; Beauchamp et al. 1999); such an approach would likely be sufficient to provide an initial indication of efficacy in vivo. Furthermore the engraftment efficiency could be improved in pre-clinical models if the muscle is 'primed' by inducing damage and subsequent muscle regeneration by pre-treatment with: irradiation (Morgan et al. 1993), myotoxins/ cardiotoxins (Harris 2003) or cryoinjury (Morgan et al. 1987).

The strategy presented is only likely to provide correction of a specific, defined and accessible muscle. While it is argued correction of single muscle could provide an improved quality of life to DMD patients, by for example maintaining upper limb motility if targeting the biceps brachii (Briggs & Morgan 2013; Skuk et al. 2007); skeletal muscle constitutes 50% of our body mass and dilated cardiomyopathy is currently the leading cause of premature death in DMD, hence, there is a requirement for systemic restoration of dystrophin (Bushby et al. 2010b; Long et al. 2016; Pini et al. 2017). Interestingly, satellite cells and myoblast progeny are not suitable for systemic delivery (Briggs & Morgan 2013). The most characterised cellular candidates with good tolerability following intra-arterial systemic delivery in preclinical models are the blood and bone marrow derived CD133+ (Torrente et al. 2004) and vessel derived pericytes/mesoangioblasts (Díaz-Manera et al. 2010; Sampaolesi et al. 2003; Sampaolesi et al. 2006; Tedesco et al. 2012). Thus for the true therapeutic potential of this gene-editing strategy to be achieved, gene modification needs to be trialled in one of these two cell candidates prior to intra-arterial injection. One elegant study by Tedesco et al in 2012 used fibroblasts and myoblasts from limb girdle muscular dystrophy (LGMD) patients, to generate iPSCs. Once an iPSC culture was obtained and mesoangioblasts-like cells were subsequently derived and genetically corrected to treat LMGD patients. Importantly iPSCs can theoretically proliferate indefinitely thereby providing a substantial and essential source of mesoangioblasts if these are used for cell-based therapies.

Importantly, while preclinical models of muscular dystrophy tend to promising levels of correction and/or functional improvements following cellular engraftment approaches, this is challenging to transition to human patients, as evidenced by the relative successes and challenges of three clinical trials (Skuk et al. 2007; Périé et al. 2014; Cossu et al. 2015).

The first Phase I clinical trial by Skuk et al utilised HLA-matched allogenic myoblast transplantation in three muscle groups of a 26 year old DMD patient on Tacrolimus, an immunosuppressive therapy, these being: the gastrocnemius, biceps brachii and thenar eminence. In spite of the low myoblast transplantation efficiency observed by previous clinical trials in the 1990s; this study asserted that more knowledge in relation to the numeracy and distribution of injections was required. Earlier studies by this group indicated that muscle precursor cells only fuse and have regenerative capacity for myofibres they come into contact with, being defined by the trajectory of the injection (Skuk et al. 2002; Skuk et al. 2006). As a result of these findings a high density injection protocol in which 25-100 injections are placed within 1 – 2mm of each other, was undertaken. Two biopsies taken 1 month and 18 months post transplantation indicated 27% and 34.5% dystrophin positive fibres in the gastrocnemius respectively. However, other muscle groups failed to show dystrophin positive fibres upon

biopsy and other than improved metacarpal flexion no functional improvement was observed. This trial was selected as a further Phase I/II variant of this approach, transplanting 30 million myoblast per cm³ per 3, 6 and 9cm³ of the extensor carpi radialis (Clinical Trial reference: NCT02196467) is currently underway.

The second arguably more successful Phase I/IIa clinical trial in Ocular Pharyngeal Muscular Dystrophy (OPMD) by Perie et al, combined Heller myotomy and an autologous transplantation (of healthy unaffected myoblasts, derived from other muscle groups) in 12 patients. Importantly, OPMD is characterised by dysphagia, a difficulty in swallowing and ptosis a drooping of eyelids; hence the trial was concerned with safety and the time taken to drink 80mls as end points (Périé et al. 2014). Notably, all 12 patients showed functional improvement as assessed by the mean time taken to drink reducing almost 2 fold from 23.7 pre-transplantation to 10.2 seconds 2 years post-transplantation. One critique of this study is that they was no Heller Myotomy alone patient control group was included due to low patient number; which would be beneficial particularly as myotomy in isolation is initially associated with an immediate functional improvement in swallowing. This stated a secondary progressive reoccurrence of dysphagia is normally observed 39 months after myotomy; interestingly of the 10 patients that agreed to long term follow up at 56 months following myotomy and myoblast transplantation, 8 were clinically stable.

Overall, these trials highlight some success in the field of myoblast transplantation, the latter an elegant example of the benefit of autologous engraftment, dispensing with the requirement for life-long immunosuppressive therapies. Whilst such approaches hold promise such studies are once again limited by the propensity to correct a single, defined and accessible muscle. Thus these studies highlight that application of myoblast transplantation in a therapeutic setting is limited and will not provide the body wild correction sought by DMD patients. Furthermore, a repeated high density injection of myoblasts does not appear to be a feasible long –term treatment of muscular dystrophies.

The final Phase I/IIa proof-of-principle clinical trial demonstrated large scale production of mesoangioblasts in line with good manufacturing practice was achievable (Cossu et al. 2015). The primary aim of the study was to assess the safety and tolerability of delivering HLA-matched mesoangioblasts by intra-arterial infusion, in 5 DMD patients. The results presented were deemed promising with functional stabilisation of 2 of the 3 ambulant patients included in the trial. Although no functional improvement was identified throughout the course of the trial; authors attribute this to the low cell number of mesoangioblasts used and indicate an increase in cell number may improve clinical efficacy. However, results of this trial should be

interpreted with caution, particularly as comparison of pre and 2 months post-transplantation biopsies highlighted quantifiable increases in dystrophin protein in 2/5 patients; the differences in dystrophin levels being 3-11% and 11-22% respectively. Furthermore 3/5 patients showed evidence of sub-cutaneous asymptomatic 'livedo reticularis' a purplish mottled colour of the skin, which can indicate impaired circulation. This in combination with one patient having a thalamic stroke following the fourth and final infusion of mesoangioblasts indicates that further investigation of how mesoangioblast infusion influences circulation is warranted. This stated the patiently cohort included in the trial is relatively small, hence if the protocol can be improved to reduce side effects, mesoangioblasts could still show efficacy if the cell number is escalated.

This trial illustrates the propensity for cells to be delivered systemically in the treatment of DMD, with some subtle dystrophin restoration observed at sub-optimal cell numbers. This is a notable advancement in the field of cell-based therapies, although the side effects observed, indicate further assessment of safety and tolerability is required. Notably, one further cell candidate is currently in clinical trial. The clinical trial aims to assess the safety and efficacy of autologous transplantation of patient bone marrow derived mesenchymal stem cells in the treatment of DMD (Trial Code NCT03067831); using muscle strength as a primary outcome, the results of which are anticipated in 2020.

The on-going examination of cellular candidates in preclinical models and in clinical trials, highlights that cell based therapies in the treatment of muscular dystrophies continues to be a dynamic and evolving field of research. In combination, with greater production of cellular material becoming more achievable owing to the ability to re-programme somatic cells to become iPSCs and produce certain cell lineages (Takahashi et al. 2007; Goudenege et al. 2012; Tedesco et al. 2012); It can be seen that the advancing field of gene-therapeutics and cell based strategies could supplement each other in the future advancement of regenerative therapeutic approaches.

7.3.2 Viral Delivery Strategies

Direct administration of gene editing reagents, may represent an alternative delivery strategy to cell engraftment approaches. Although improved in-silico guide design and Cas9 protein specificity would be required, if developed, this approach could hold therapeutic promise in the future. In addition to improvement of these parameters, an efficient delivery vector would also be of utmost importance to ensure the success of this approach; pertinent to this is the recent packaging of SpCas9 in a range of viral vectors including Adenoviral (Maggio et al. 2016), Lentiviral (Ousterout et al. 2015) and Adeno-Associated Virus(Wright et al. 2015). In the

case of the latter, the large size of the SpCas9 transgene (4.1kb), precluded its complete packaging into AAV, which have a genome of 4.7kb and upper packaging limit of 5kb (Wu et al. 2010b). Thus the SpCas9 transgene was split in half and reconstituted by dual AAV transduction experimentally (Wright et al. 2015). In addition, the continued examination of alternative Cas9 proteins have served to identify the smaller *Staphylococcus Aureus* Cas9 and Cpf1 derived from *Acidominococcus* and *Lachnospiraceae*, which can be readily packaged into AAV (Ran et al. 2015; Zetsche et al. 2015). In addition to overcoming the packaging constraints of AAV, these bacterial orthologues have unique PAM recognition sequences and specificity profiles and thus are promising expansions to the gene-editing tool kit.

The viral packaging of SpCas9 and SaCas9 served to rejuvenate interest in in-vivo correction strategies of the premature stop codon in the mdx mouse. Four studies were performed in pursuit of this: using AdV serotype 5 (Xu et al. 2016), AAV8 (Tabebordbar et al. 2016; Nelson et al. 2016) and AAV9 (Long et al. 2016). Initially, Wu et al demonstrated NHEJ mediated deletion of exons 21-23 in newborn mdx mouse pups, utilising a dual adenoviral transduction approach; in which vectors encoded spCas9-2A-GFP and guide RNAs respectively. Impressively, the injection of these two vectors in the gastrocnemius produced a 23kb genomic deletion and restored dystrophin expression to 50% of that observed in healthy mice (Xu et al. 2016). Interestingly, the subsequent three studies that followed utilised the SaCas9, which owing to its alternative guide requirements enabled precise NHEJ deletion of exon 23 in isolation. Nelson et al performed an Intramuscular injection of AAV8 into mdx mice, prompting the NHEJ mediated removal of exon 23. A 2% correction of exon 23 at the genomic level was detected, however an analysis at the transcriptome level highlighted that 59% of transcripts harboured a deletion of exon 23 (Nelson et al. 2016). It was posited that this was due to protection of these transcripts from nonsense mediated decay. Intriguingly, this low level of genomic correction also produced improvement in mdx muscle histology and functional assessments. Finally, Intraperitoneal and intravenous injections were also trialled showing improvement in cardiac muscles.

A similar study to Nelson's used mdx Ai9 mice; using the Ai9 reporter allele for initial visual identification of gene modification, as the removal of a premature stop codon in this sequence enabled dTomato expression; this was prior to progressing to the correction of the mdx mutation (Tabebordbar et al. 2016). Secondly, they trialled AAV9 serotypes by a range of delivery systems including: Intramuscular injection into the *tiabialis anterior*, intraperitoneal and intravenous injections. At an individual *tiabialis anterior* muscle level approximately 39% of mRNA transcripts, assessed by RT-PCR and sequencing harboured exon 23 deletion. This translated to improvement in neuronal nitric oxide synthase (nNOS) localisation and muscle

specific force. Promisingly, this extended to intraperitoneal and intravenous delivery, for which exon 23 excision varied from 3 -18% (Tabebordbar et al. 2016). Furthermore, upon crossing Ai9 mice Pax7+ZsGreen+ satellite cells, which allow for the identification of satellite cells by green fluorescence; it was discovered that exon 23 excision extended to satellite cells. In light of the high regenerative capacity of satellite cells in muscle physiologically, this is an important finding to the muscular dystrophy field. Further characterisation of the extent of satellite cell correction is warranted particularly, as the propensity for AAVs to transduce and correct satellite cells is asserted as a current limitation of using this viral vector for a gene editing strategy (Philpott et al. 2002; Counsell et al. 2017)

A final study used SaCas9 NHEJ-mediated removal of exon 23, to assess the efficiency of intramuscular, intraperitoneal and retro-orbital injections, using dystrophin presence in cardiac muscle as an end-point (Long et al. 2016). Intriguingly, this study found that all methods of delivery resulted in some restoration in vascular smooth muscle cells, attributed to the entry of the AAV9 into intravascular circulation. This stated retro-orbital was shown to provide the best dystrophin restoration, with a 2 fold improvement upon other methods in skeletal muscle providing a measure of $53.2 \pm 18.5\%$ and a marginal increase in cardiomyocytes. It was also observed that AAVs were unable to cross the blood brain barrier and address the neurological aspect of DMD.

These studies provide some important demonstrations of the therapeutic potential of gene-editing: i) Both SpCas9 and the bacterial orthologue SaCas9 can be vectorised using AdVs and AAVs, ii) The bacterial system utilised influences target-site availability and by extension the amount of genomic material that needs to be excised in an NHEJ-approach iii) Low levels of genomic correction can confer functional improvement particularly if the transcript is protected or correction occurs in cells with high muscle regenerative capacity iv) Numerous routes of administration are successful but influence muscle-group dystrophin restoration v) Alternative strategies may be required to provide correction of muscle and neurological clinical phenotypes.

In addition to the correction of the premature stop codon genetic defect in mdx mice, two pertinent studies highlighted the therapeutic prospect of using NHEJ-mediated deletion to remove the $\Delta 45-55$ mutational hotspot, which could be applicable to 62% of patients (Ousterout et al. 2015; Maggio et al. 2016). The former used a Lentiviral delivery system and showed low to undetectable levels of gene modification in patient myoblasts, requiring enrichment to detect the truncated dystrophin. In contrast, the latter demonstrated NHEJ-mediated deletion of exons 45-55 using an all-in-one construction of SpCas9 in an adenoviral

vector. Impressively, 10% of patient myoblasts were shown to harbour internal deletion of dystrophin, in the absence of selection (Maggio et al. 2016). These demonstrations remain at the *in vitro* demonstration stage and have yet to be tested in-vivo.

Notably, all of the studies presented used an NHEJ-mediated deletion strategy, arguably indicating that HDR-therapeutics/gene integration approaches are lagging behind in the gene-editing field currently. The most recent demonstration with CRISPR/Cas9 being the correction of mdx zygotes by the injection of mRNA guides and Cas9 alongside a ssODN exogenous donor (Long et al. 2014). Importantly, while the modification of Beta thalassemia in human embryos was recently undertaken using the Cas9 gene-editing system; this approach is associated with a plethora of ethical considerations and is currently not a treatment strategy for DMD patients (P. Liang et al. 2015).

Collectively these publications highlight the potential of a viral-based strategy, to deliver the gene-editing reagents required to achieve the near-universal therapeutic strategy presented. In the context of a HDR-mediated MD1 integration strategy, dual adenoviral transduction using 'guttled' adenoviruses could be envisaged; one vector encoding an all-in-one spCas9 construct and a second delivering the exogenous repair template encoding either the full length or $\Delta 45-55$ sequence optimised dystrophin variant. This could be pertinent to the continued advancement of the strategy presented, particularly as the $\Delta 45-55$ and full length sequence optimised dystrophin variants whilst characterised remained untested in the context of integration. The large packaging capacity of the gutted AdV being approximately 37kb makes it a suitable candidate for the delivery of the approximate 11kb full-length dystrophin cDNA alongside selection markers (Mitani et al. 1995; Gonçalves & de Vries 2006). Importantly, there is precedent of the gutted adenoviral delivering full length dystrophin in gene augmentation studies; with intramuscular injections into mdx mice demonstrating 25-30% dystrophin restoration 1 month post transduction and improved resilience to contraction induced injury (DelloRusso et al. 2002). In addition, this study utilised native dystrophin, it is likely the enhanced protein expression associated with sequence optimised dystrophin would produce a more favourable outcome.

One limitation to the use of adenoviral vectors therapeutically, is their immunogenicity. It has been shown that the components of the AdV: its capsid, dsDNA genome and transcriptome can all trigger innate immune responses in hosts (Machitani et al. 2011; Alonso-Padilla et al. 2016). Further innate immune responses have also been indicated resulting from the trafficking of the AdV from the cytoplasm to the nuclear pore post-entry into the cell (Teigler et al. 2014). In addition when delivered systemically, the most suitable route of administration for the

treatment of DMD the AdVs are sequestered by the liver, a major source of pro-inflammation. Importantly, improvements to AdV construction and experimental design have been developed. It has been shown construction of the adenoviral capsids from adenoviral serotypes other than 5 and the shielding of the virus with polymers, reduces inflammatory responses (Roberts et al. 2006; Capasso et al. 2014). Secondly, the sequestration of the AdV vector by the liver, reduces efficient transduction and increases the chance of an immune response. To improve upon this Piccolo et al showed that the scavenger receptors: SR-A and SREC-L that mediate AdV uptake by kupffer cells in the liver, could be blocked transiently with the use of peptides. This was shown to be well tolerated and keep interleukins at a stable level in preclinical models (Piccolo et al. 2014).

A contrasting approach also prompted by these studies, is a HITI NHEJ-mediated integration of the dystrophin transgenes. Provisional, studies could focus upon the use of AAV-SaCas9 vectors coupled to an AAV-MD1-RT with flanking guide sequences in reverse orientation to its occurrence in the genome. This could provide a higher efficiency as the NHEJ DNA repair pathway occurs throughout the cell cycle. The use of AAV serotype 8 or 9 would be beneficial as they have a more favourable immunity profile relative to AdVs, furthermore they have a natural tropism to cardiac and skeletal muscle (Zincarelli et al. 2008; Ferrand et al. 2015). Due to the limited <5kb packaging capacity of AAVs, the integration of the larger transgenes could be attempted but would likely be constrained to the use of AdVs or possibly Lentiviruses, as the delivery of full length dystrophin using this vector was recently shown (Counsell et al. 2017).

7.3.3 Non-viral Delivery Approaches:

An alternative approach to the delivery of gene-editing reagents is to use a non-viral carrier to protect /shield the biological entity from degradation (Sun et al. 2015). Lee et al 2017 elegantly demonstrated the delivery Cas9 RNP and an exogenous ssODN donor, both *in vitro* and *in vivo* with the use of DNA-coated gold nanoparticles. The construction of these nanoparticles was dependent upon: i) the affinity of the Cas9 Nuclease for the guide RNA and for the DNA coating the gold particle and ii) the electrostatic attraction between the Cas9 protein and the gold nanoparticle (Lee et al. 2017). Being concerned with the development of HDR-therapeutics, this study demonstrated *in vitro* the conversion of eGFP to BFP, in a HEK293T-eGFP cell line, with the use of Cas9 ribonucleotide proteins (RNPs) at 11.3% efficiency. This strategy was then extended to address the premature stop codon in exon23 of the mdx mouse. Intra-muscular injection of 4 week old mdx mice were undertaken post cardiotoxin treatment of tibialis anterior muscles. This was shown to restore 5.4% of dystrophin

expression, as ascertained 2 weeks post injection. Furthermore, some functional improvement was observed in treated mice with the four limb hang test.

A second study by Sun et al of non-viral delivery of Cas9 RNPs, illustrated the use of self-assembled DNA nanoclews. The DNA nanoclews, a yarn-like nanoparticle, were delivered to U2Os.eGFP cells; the reduction in fluorescence was used to indicate the presence of sequence modifications resulting from DNA cleavage. *In vitro*, this system indicated a 36% reduction in fluorescence. Importantly when extended to U2Os.eGFP tumour bearing mice, by a direct injection into the tumour this efficiency was shown to be 25%(Sun et al. 2015).

The use of non-viral delivery strategies represent an attractive alternative to the use of viral vectors, being well tolerated by hosts even upon repeat administration(Lee et al. 2017). Moreover, the use of Cas9 RNPs would be beneficial in a therapeutic strategy owing to their transient expression profile and engineering for enhanced specificity (S. Kim et al. 2014; Slaymaker et al. 2016; Kleinstiver et al. 2016). This would serve to reduce Genome-wide off-target effects, a concern that has served to limit the use of Cas9 in a therapeutic setting. However, in the context the near-universal therapeutic strategy presented, it would be challenging to deliver the large exogenous repair template required being approximately 9.8kb. Moreover, the reduced efficiency relative to viral vectors and the local delivery presented by the above strategies, present further hurdles to attaining therapeutic thresholds of dystrophin body-wide. Overall, these strategies remain promising for defined organs or tumours; however, further research is warranted to develop higher efficiency platforms amenable to systemic delivery.

7.4 Concluding Remarks

The study aimed to investigate a genome-editing strategy with near universal therapeutic applicability to Duchenne Muscular Dystrophy patients. By using the newly re-purposed CRISPR-Cas9 prokaryotic immune system, targeting intron 1 of the DMD gene, in combination with an exogenous repair template encoding microdystrophin; a proof of principle demonstration of this genome surgery approach was achieved in HEK293T cells. In addition, a provisional indication of microdystrophin integration was also observed in immortalised patient myoblasts. The research presented has extensive scope for development, alternative dystrophin transgenes, exogenous repair template constructs and delivery systems of gene-editing reagents. The efficacy could also be improved with altered guide design. Furthermore in light of on-going research in the field of cell based therapies and the ability to package Cas9 nucleases into viral and non-viral vectors it has high therapeutic potential, in the treatment of DMD.

Bibliography

- Aartsma-rus, A. et al., 2005. Functional Analysis of 114 Exon-Internal AONs for Targeted DMD Exon Skipping: Indication for Steric Hindrance of SR Protein Binding Sites. *Oligonucleotides*, 15(4), pp.284–197. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16396622> [Accessed November 24, 2017].
- Aartsma-Rus, A. et al., 2017. Development of Exon Skipping Therapies for Duchenne Muscular Dystrophy: A Critical Review and a Perspective on the Outstanding Issues. *Nucleic acid therapeutics*, 27(5), pp.251–259. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28796573> [Accessed November 22, 2017].
- Aartsma-Rus, A. et al., 2006. Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. *Muscle & nerve*, 34(2), pp.135–44. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16770791> [Accessed November 24, 2015].
- Aartsma-Rus, A. et al., 2009. Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations. *Human mutation*, 30(3), pp.293–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19156838> [Accessed January 27, 2014].
- Aartsma-Rus, A. & Krieg, A.M., 2017. FDA Approves Eteplirsen for Duchenne Muscular Dystrophy: The Next Chapter in the Eteplirsen Saga. *Nucleic acid therapeutics*, 27(1), pp.1–3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27929755> [Accessed November 24, 2017].
- Acharyya, S. et al., 2007. Interplay of IKK/NF-κB signaling in macrophages and myofibers promotes muscle degeneration in Duchenne muscular dystrophy. *Journal of Clinical Investigation*, 117(4), pp.889–901. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17380205> [Accessed November 24, 2017].
- Aiuti, A., Roncarolo, M.G. & Naldini, L., 2017. Gene therapy for ADA-SCID, the first marketing approval of an ex vivo gene therapy in Europe: paving the road for the next generation of advanced therapy medicinal products. *EMBO molecular medicine*, 9(6), pp.737–740. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28396566> [Accessed November 25, 2017].
- Allen, D.G., Whitehead, N.P. & Froehner, S.C., 2016. Absence of Dystrophin Disrupts Skeletal Muscle Signaling: Roles of Ca²⁺, Reactive Oxygen Species, and Nitric Oxide in the Development of Muscular Dystrophy. *Physiological Reviews*, 96(1), pp.253–305. Available at: <http://physrev.physiology.org/content/physrev/96/1/253.full.pdf> [Accessed November 19, 2017].
- Alonso-Padilla, J. et al., 2016. Development of Novel Adenoviral Vectors to Overcome Challenges Observed With HAdV-5–based Constructs. *Molecular Therapy*, 24(1), pp.6–16. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26478249> [Accessed November 17, 2017].
- Alton, E.W. et al., 2016. *A randomised, double-blind, placebo-controlled trial of repeated nebulisation of non-viral cystic fibrosis transmembrane conductance regulator (CFTR) gene therapy in patients with cystic fibrosis*, NIHR Journals Library. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27441329> [Accessed November 25, 2017].
- Alton, E.W.F.W. et al., 2015. Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial. *The Lancet Respiratory Medicine*, 3(9), pp.684–691. Available at:

- <http://www.thelancet.com/article/S2213260015002453/fulltext> [Accessed July 6, 2015].
- Anders, C. et al., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513(7519), pp.569–573. Available at: <http://www.nature.com/doi/10.1038/nature13579> [Accessed November 6, 2017].
- Anderson, K. & Meissner, G., 1995. T-tubule depolarization-induced SR Ca²⁺ release is controlled by dihydropyridine receptor- and Ca(2+)-dependent mechanisms in cell homogenates from rabbit skeletal muscle. *The Journal of general physiology*, 105(3), pp.363–83. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7769380> [Accessed June 11, 2018].
- Anthony, K. et al., 2011. Dystrophin quantification and clinical correlations in Becker muscular dystrophy: implications for clinical trials. *Brain*, 134(12), pp.3547–3559. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22102647> [Accessed July 28, 2017].
- Araki, K., Araki, M. & Yamamura, K. -i., 2006. Negative Selection with the Diphtheria toxin A fragment Gene Improves Frequency of Cre-Mediated Cassette Exchange in ES Cells. *Journal of Biochemistry*, 140(6), pp.793–798. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17043056> [Accessed June 28, 2017].
- Arpke, R.W. et al., 2013. A new immuno-, dystrophin-deficient model, the NSG-mdx(4Cv) mouse, provides evidence for functional improvement following allogeneic satellite cell transplantation. *Stem cells (Dayton, Ohio)*, 31(8), pp.1611–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23606600> [Accessed November 17, 2017].
- Arras, S.D.M. & Fraser, J.A., 2016. Chemical inhibitors of non-homologous end joining increase targeted construct integration in *Cryptococcus neoformans*. *PLoS ONE*, 11(9), pp.1–14.
- ASGCT, 2014. Gene Therapy Defined | ASGCT - American Society of Gene & Cell Therapy. Available at: http://www.asgct.org/about_gene_therapy/defined.php [Accessed February 11, 2014].
- Ashworth, J. et al., 2010. Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic acids research*, 38(16), pp.5601–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20435674> [Accessed November 26, 2017].
- Assenmacher, N. & Hopfner, K.P., 2004. MRE11/RAD50/NBS1: Complex activities. *Chromosoma*, 113(4), pp.157–166.
- Athanasopoulos, T. et al., 2011. Codon optimization of the microdystrophin gene for Duchene muscular dystrophy gene therapy. *Methods in molecular biology (Clifton, N.J.)*, 709, pp.21–37. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21194019> [Accessed July 16, 2015].
- Austin, R.C. et al., 1995. Cloning and characterization of alternatively spliced isoforms of Dp71. *Human molecular genetics*, 4(9), pp.1475–83. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8541829> [Accessed August 6, 2017].
- Banks, G.B. et al., 2007. Functional capacity of dystrophins carrying deletions in the N-terminal actin-binding domain. *Human Molecular Genetics*, 16(17), pp.2105–2113. Available at: <https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddm158> [Accessed November 5, 2017].
- Banks, G.B. et al., 2010. The polyproline site in hinge 2 influences the functional capacity of truncated dystrophins. G. A. Cox, ed. *PLoS genetics*, 6(5), p.e1000958. Available at: <http://dx.plos.org/10.1371/journal.pgen.1000958> [Accessed September 8, 2017].

- Bar, S. et al., 1990. A novel product of the Duchenne muscular dystrophy gene which greatly differs from the known isoforms in its structure and tissue distribution. *Biochemical Journal*, 272(2). Available at: <http://www.biochemj.org/content/272/2/557> [Accessed August 6, 2017].
- Barrangou, R. et al., 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, 315(5819), pp.1709–1712. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17379808> [Accessed November 27, 2017].
- Barzel, a et al., 2014. Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature*, 517(7534), pp.360–364.
- Beauchamp, J.R. et al., 1999. Dynamics of myoblast transplantation reveal a discrete minority of precursors with stem cell-like properties as the myogenic source. *The Journal of cell biology*, 144(6), pp.1113–22. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2150577&tool=pmcentrez&rendertype=abstract> [Accessed February 11, 2014].
- Belanto, J.J. et al., 2014. Microtubule binding distinguishes dystrophin from utrophin. *Proceedings of the National Academy of Sciences of the United States of America*, 111(15), pp.5723–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24706788> [Accessed November 8, 2017].
- Benchling, 2015. Benchling · Better tools, faster research. Available at: <https://benchling.com/> [Accessed November 6, 2017].
- Bengtsson, N.E. et al., 2017. Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy. *Nature Communications*, 8, p.14454. Available at: <http://www.nature.com/doi/10.1038/ncomms14454> [Accessed November 28, 2017].
- Betts, C. et al., 2012. Pip6-PMO, A New Generation of Peptide-oligonucleotide Conjugates With Improved Cardiac Exon Skipping Activity for DMD Treatment. *Molecular therapy. Nucleic acids*, 1(8), p.e38. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23344180> [Accessed December 16, 2017].
- Bhakta, M.S. et al., 2013. Highly active zinc-finger nucleases by extended modular assembly. *Genome research*, 23(3), pp.530–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23222846> [Accessed November 26, 2017].
- Bhattacharya, I. et al., 2017. Safety, Tolerability, Pharmacokinetics, and Pharmacodynamics of Domagrozumab (PF-06252616), an Antimycostatin Monoclonal Antibody, in Healthy Subjects. *Clinical Pharmacology in Drug Development*. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28881472> [Accessed November 29, 2017].
- Birnkrant, D.J. et al., 2010. The respiratory management of patients with duchenne muscular dystrophy: A DMD care considerations working group specialty article. *Pediatric Pulmonology*, 45(8), pp.739–748. Available at: <http://doi.wiley.com/10.1002/ppul.21254> [Accessed November 24, 2017].
- Bitinaite, J. et al., 1998. FokI dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, 95(18), pp.10570–5. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=85228&tool=pmcentrez&rendertype=abstract>.

- Bladen, C.L. et al., 2015. The TREAT-NMD DMD Global Database: analysis of more than 7,000 Duchenne muscular dystrophy mutations. *Human mutation*, 36(4), pp.395–402. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4405042&tool=pmcentrez&rendertype=abstract> [Accessed August 5, 2015].
- Blat, Y. & Blat, S., 2015. Drug Discovery of Therapies for Duchenne Muscular Dystrophy. *Journal of Biomolecular Screening*, 20(10), pp.1189–1203. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25975656> [Accessed November 20, 2017].
- Blot, W. & Handbook, N., 2017. Technical Note Western Blot Normalization Handbook. , (March).
- Boch, J. et al., 2009. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*, 326(5959), pp.1509–1512. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19933107> [Accessed November 27, 2017].
- Boch, J., 2011. TALEs of genome targeting. *Nature Biotechnology*, 29(2), pp.135–136. Available at: <http://www.nature.com/doifinder/10.1038/nbt.1767> [Accessed November 27, 2017].
- Boldrin, L., Zammit, P.S. & Morgan, J.E., 2015. Satellite cells from dystrophic muscle retain regenerative capacity. *Stem Cell Research*, 14(1), pp.14–20. Available at: <http://www.sciencedirect.com/science/article/pii/S1873506114001172> [Accessed November 28, 2017].
- Bostick, B. et al., 2008. Adeno-associated virus serotype-9 microdystrophin gene therapy ameliorates electrocardiographic abnormalities in mdx mice. *Human gene therapy*, 19(8), pp.851–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18666839> [Accessed September 8, 2017].
- Bothmer, A. et al., 2017. Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nature Communications*, 8(May 2016), p.13905. Available at: <http://www.nature.com/doifinder/10.1038/ncomms13905>.
- Brandan, E., Cabello-Verrugio, C. & Vial, C., 2008. Novel regulatory mechanisms for the proteoglycans decorin and biglycan during muscle formation and muscular dystrophy. *Matrix Biology*, 27(8), pp.700–708. Available at: <http://www.sciencedirect.com/science/article/pii/S0945053X08001054?via%3Dihub> [Accessed November 24, 2017].
- Briggs, A.W. et al., 2012. Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic acids research*, 40(15), p.e117. Available at: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks624> [Accessed November 27, 2017].
- Briggs, D. & Morgan, J.E., 2013. Recent progress in satellite cell/myoblast engraftment -- relevance for therapy. *The FEBS journal*, 280(17), pp.4281–93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23560812> [Accessed October 4, 2017].
- Brinkman, E.K. et al., 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic acids research*, 42(22), p.e168. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25300484> [Accessed July 3, 2017].
- Brinkmeyer-Langford, C. & Kornegay, J.N., 2013. Comparative Genomics of X-linked Muscular Dystrophies: The Golden Retriever Model. *Current genomics*, 14(5), pp.330–42. Available

at:

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3763684&tool=pmcentrez&rendertype=abstract>.

- Bulfield, G. et al., 1984. X chromosome-linked muscular dystrophy (mdx) in the mouse. *Proceedings of the National Academy of Sciences of the United States of America*, 81(4), pp.1189–92. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=344791&tool=pmcentrez&rendertype=abstract>.
- Buller, R.M. et al., 1981. Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication. *Journal of virology*, 40(1), pp.241–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6270377> [Accessed December 16, 2017].
- Bushby, K. et al., 2014. Ataluren treatment of patients with nonsense mutation dystrophinopathy. *Muscle & nerve*, 50(4), pp.477–87. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4241581&tool=pmcentrez&rendertype=abstract> [Accessed November 24, 2015].
- Bushby, K. et al., 2010a. Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and pharmacological and psychosocial management. *Lancet neurology*, 9(1), pp.77–93. Available at: <http://www.thelancet.com/journals/a/article/PIIS1474-4422%2809%2970271-6/fulltext> [Accessed January 25, 2014].
- Bushby, K. et al., 2010b. Diagnosis and management of Duchenne muscular dystrophy, part 2: implementation of multidisciplinary care. *Lancet neurology*, 9(2), pp.177–89. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19945914> [Accessed January 27, 2014].
- Byers, T.J., Lidov, H.G.W. & Kunkel, L.M., 1993. An alternative dystrophin transcript specific to peripheral nerve. *Nature Genetics*, 4(1), pp.77–81. Available at: <http://www.nature.com/doi/10.1038/ng0593-77> [Accessed August 6, 2017].
- Byrne, S.M. et al., 2015. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Research*, 43(3), pp.e21–e21. Available at: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku1246> [Accessed July 29, 2017].
- Calabrese, C.R. et al., 2004. Anticancer Chemosensitization and Radiosensitization by the Novel Poly(ADP-ribose) Polymerase-1 Inhibitor AG14361. *JNCI Journal of the National Cancer Institute*, 96(1), pp.56–67. Available at: <https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djh005> [Accessed June 19, 2017].
- Campbell, C. et al., 2017. Myostatin inhibitor ACE-031 treatment of ambulatory boys with Duchenne muscular dystrophy: Results of a randomized, placebo-controlled clinical trial. *Muscle & Nerve*, 55(4), pp.458–464. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27462804> [Accessed November 29, 2017].
- Cannarozzi, G. et al., 2010. A Role for Codon Order in Translation Dynamics. *Cell*, 141(2), pp.355–367. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0092867410001893> [Accessed December 12, 2017].
- Canver, M.C., Bauer, D.E. & Orkin, S.H., 2017. Functional interrogation of non-coding DNA through CRISPR genome editing. *Methods*, 121–122, pp.118–129. Available at: <http://www.sciencedirect.com/science/article/pii/S1046202317300014> [Accessed November 5, 2017].

- Capasso, C. et al., 2014. The Evolution of Adenoviral Vectors through Genetic and Chemical Surface Modifications. *Viruses*, 6(2), pp.832–855. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24549268> [Accessed November 17, 2017].
- Carlson, D.F. et al., 2012. Efficient TALEN-mediated gene knockout in livestock. *Proceedings of the National Academy of Sciences of the United States of America*, 109(43), pp.17382–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23027955> [Accessed July 14, 2017].
- Carroll, D., 2013. Staying on target with CRISPR-Cas. *Nature Biotechnology*, 31(9), pp.807–809. Available at: <http://www.nature.com/doifinder/10.1038/nbt.2684> [Accessed October 12, 2017].
- Cellectis Bioresearch, C., 2011. Cellectis bioresearch. Available at: <http://talen-hit.cellectis-bioresearch.com/> [Accessed February 11, 2014].
- Cerletti, M. et al., 2003. Dystrophic phenotype of canine X-linked muscular dystrophy is mitigated by adenovirus-mediated utrophin gene transfer. *Gene Therapy*, 10(9), pp.750–757. Available at: <http://www.nature.com/articles/3301941> [Accessed November 8, 2017].
- Cermak, T. et al., 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*, 39(12), pp.e82–e82. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21493687> [Accessed November 27, 2017].
- Chamberlain, J.S., 1997. Dystrophin levels required for genetic correction of Duchenne muscular dystrophy. *Basic and Applied Myology*, 7(3–4), pp.251–256. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.588.819> [Accessed July 28, 2017].
- Chamberlain, J.S. et al., 1988. Expression of the murine Duchenne muscular dystrophy gene in muscle and brain. *Science (New York, N.Y.)*, 239(4846), pp.1416–1418.
- Chamberlain, J.S. et al., 1987. Regional localization of the murine Duchenne muscular dystrophy gene on the mouse X chromosome. *Somatic cell and molecular genetics*, 13(6), pp.671–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2890215> [Accessed June 24, 2015].
- Chan, J.H., Lim, S. & Wong, W.F., 2006. ANTISENSE OLIGONUCLEOTIDES: FROM DESIGN TO THERAPEUTIC APPLICATION. *Clinical and Experimental Pharmacology and Physiology*, 33(5–6), pp.533–540. Available at: <http://doi.wiley.com/10.1111/j.1440-1681.2006.04403.x> [Accessed November 24, 2017].
- Chari, R. et al., 2015a. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nature Methods*, 12(9), pp.823–826. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26167643> [Accessed November 6, 2017].
- Chari, R. et al., 2015b. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nature methods*, 12(July), pp.1–7. Available at: <http://www.nature.com/doifinder/10.1038/nmeth.3473%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/26167643>.
- Chen, B. et al., 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155(7), pp.1479–1491. Available at: <http://dx.doi.org/10.1016/j.cell.2013.12.001>.
- Chen, P.R. & Lee, K., 2016. INVITED REVIEW: Inhibitors of myostatin as methods of enhancing

- muscle growth and development. *Journal of Animal Science*, 94(8), p.3125. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27695802> [Accessed November 29, 2017].
- Chevalier, B.S. et al., 2002. Design, activity, and structure of a highly specific artificial endonuclease. *Molecular cell*, 10(4), pp.895–905. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12419232> [Accessed November 26, 2017].
- Cho, S.W. et al., 2014a. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research*, 24(1), pp.132–41. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24253446> [Accessed November 6, 2017].
- Cho, S.W. et al., 2014b. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research*, 24(1), pp.132–41. Available at: <http://genome.cshlp.org/content/24/1/132.full?sid=9fe4b18b-ed1-4831-bae1-a71a70d94318> [Accessed January 20, 2014].
- Chu, V.T. et al., 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, 33(5), pp.543–548. Available at: <http://www.nature.com/doifinder/10.1038/nbt.3198>.
- Ciccia, A. & Elledge, S.J., 2010. The DNA damage response: making it safe to play with knives. *Molecular cell*, 40(2), pp.179–204. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20965415> [Accessed November 21, 2017].
- Ciccia, A., McDonald, N. & West, S.C., 2008. Structural and Functional Relationships of the XPF/MUS81 Family of Proteins. *Annual Review of Biochemistry*, 77(1), pp.259–287. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18518821> [Accessed November 23, 2017].
- Cirak, S. et al., 2011. Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study. *Lancet*, 378(9791), pp.595–605. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3156980&tool=pmcentrez&rendertype=abstract> [Accessed February 11, 2014].
- Cirak, S. et al., 2012. Restoration of the Dystrophin-associated Glycoprotein Complex After Exon Skipping Therapy in Duchenne Muscular Dystrophy. *Molecular Therapy*, 20(2), pp.462–467. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22086232> [Accessed November 24, 2017].
- Clarke, M.S., Khakee, R. & McNeil, P.L., 1993. Loss of cytoplasmic basic fibroblast growth factor from physiologically wounded myofibers of normal and dystrophic muscle. *Journal of cell science*, 106 (Pt 1), pp.121–33. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8270618> [Accessed November 19, 2017].
- Collins, F.S. et al., 2007. A New Partner for the International Knockout Mouse Consortium. *Cell*, 129(2), p.235. Available at: <http://www.sciencedirect.com/science/article/pii/S0092867407004643#bib1> [Accessed November 7, 2017].
- Cong, L. et al., 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, 339(6121), pp.819–23. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3795411&tool=pmcentrez&rendertype=abstract> [Accessed January 20, 2014].

- Cornu, T.I. et al., 2008. DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Molecular Therapy*, 16(2), pp.352–358. Available at: <http://www.sciencedirect.com/science/article/pii/S1525001616314022?via%3Dihub> [Accessed November 26, 2017].
- Cossu, G. et al., 2015. Intra-arterial transplantation of HLA-matched donor mesoangioblasts in Duchenne muscular dystrophy. *EMBO Molecular Medicine*, 7(12), pp.1513–1528. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26543057> [Accessed November 17, 2017].
- Counsell, J.R. et al., 2017. Lentiviral vectors can be used for full-length dystrophin gene therapy. *Scientific reports*, 7, p.44775. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28303972> [Accessed November 16, 2017].
- D'Angelo, M.G. et al., 2011. Neurocognitive Profiles in Duchenne Muscular Dystrophy and Gene Mutation Site. *Pediatric Neurology*, 45(5), pp.292–299. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22000308> [Accessed August 6, 2017].
- D'Souza, V.N. et al., 1995. A novel dystrophin isoform is required for normal retinal electrophysiology. *Human Molecular Genetics*, 4(5), pp.837–842. Available at: <https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/4.5.837> [Accessed August 6, 2017].
- Dang, Y. et al., 2015. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biology*, 16(1), p.280. Available at: <http://genomebiology.com/2015/16/1/280> [Accessed July 14, 2017].
- Decottignies, A., 2013. Alternative end-joining mechanisms: a historical perspective. *Frontiers in genetics*, 4, p.48. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23565119> [Accessed November 21, 2017].
- DelloRusso, C. et al., 2002. Functional correction of adult mdx mouse muscle using gutted adenoviral vectors expressing full-length dystrophin. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20), pp.12979–84. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12271128> [Accessed November 17, 2017].
- Deng, D. et al., 2012. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science (New York, N.Y.)*, 335(6069), pp.720–3. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3586824&tool=pmcentrez&rendertype=abstract> [Accessed January 23, 2014].
- Desmet, F.-O. et al., 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research*, 37(9), p.e67. Available at: <http://nar.oxfordjournals.org/content/early/2009/04/01/nar.gkp215.abstract> [Accessed December 10, 2014].
- van Deutekom, J.C. et al., 2007. Local dystrophin restoration with antisense oligonucleotide PRO051. *The New England journal of medicine*, 357(26), pp.2677–86. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18160687> [Accessed October 21, 2015].
- Díaz-Manera, J. et al., 2010. Partial dysferlin reconstitution by adult murine mesoangioblasts is sufficient for full functional recovery in a murine model of dysferlinopathy. *Cell Death and Disease*, 1(8), p.e61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21364666> [Accessed November 18, 2017].
- Disatnik, M.H. et al., 1998. Evidence of oxidative stress in mdx mouse muscle: studies of the

- pre-necrotic state. *Journal of the neurological sciences*, 161(1), pp.77–84. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9879685> [Accessed November 28, 2017].
- Doench, J.G. et al., 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2), pp.184–191. Available at: <http://dx.doi.org/10.1038/nbt.3437> [Accessed January 18, 2016].
- Doench, J.G. et al., 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology*, 32(12), pp.1262–7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4262738&tool=pmcentrez&rendertype=abstract> [Accessed September 4, 2014].
- Doran, P. et al., 2004. Subproteomics analysis of Ca²⁺-binding proteins demonstrates decreased calsequestrin expression in dystrophic mouse skeletal muscle. *European Journal of Biochemistry*, 271(19), pp.3943–3952. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15373840> [Accessed November 20, 2017].
- Doyon, Y. et al., 2010. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nature methods*, 7(6), pp.459–60. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20436476> [Accessed July 14, 2017].
- Dumont, N.A. et al., 2015. Dystrophin expression in muscle stem cells regulates their polarity and asymmetric division. *Nature Medicine*, 21(12), pp.1455–1463. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26569381> [Accessed September 8, 2017].
- Dunham, I. et al., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74. Available at: <http://www.nature.com/doi/10.1038/nature11247> [Accessed November 5, 2017].
- Durant, S. & Karran, P., 2003. Vanillins--a novel family of DNA-PK inhibitors. *Nucleic acids research*, 31(19), pp.5501–12. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14500812> [Accessed June 19, 2017].
- E. K. Brown, M.D.S.B.N.C.S.G.C.S.J.F.C.M.E.L.F.M.D., 1987. The mapping of a cDNA from the human X-linked muscular dystrophin gene to the mouse X chromosome. *Nature*, 330, p.1987.
- Edwards, R.H. et al., 1977. Fatigue of long duration in human skeletal muscle after exercise. *The Journal of physiology*, 272(3), pp.769–78. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/592214> [Accessed November 20, 2017].
- Elion, G.B., 1982. Mechanism of action and selectivity of acyclovir. *The American journal of medicine*, 73(1A), pp.7–13. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6285736> [Accessed November 16, 2017].
- Emery, A.E., 2002. The muscular dystrophies. *The Lancet*, 359(9307), pp.687–695. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11879882> [Accessed November 24, 2017].
- England, S.B. et al., 1990. Very mild muscular dystrophy associated with the deletion of 46% of dystrophin. *Nature*, 343(6254), pp.180–182.
- Ensembl, R. 90, 2017. Gene: DMD (ENSG00000198947) - Summary - Homo sapiens - Ensembl genome browser 90. Available at: https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000198947;r=X:31097677-33339441 [Accessed October 4, 2017].
- Epinat, J.-C. et al., 2003. A novel engineered meganuclease induces homologous

- recombination in yeast and mammalian cells. *Nucleic acids research*, 31(11), pp.2952–62. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12771221> [Accessed November 25, 2017].
- Ervasti, J.M. et al., 1990. Deficiency of a glycoprotein component of the dystrophin complex in dystrophic muscle. *Nature*, 345(6273), pp.315–319. Available at: <http://www.nature.com/doi/10.1038/345315a0> [Accessed November 19, 2017].
- Evans, N.P. et al., 2009. Immune-Mediated Mechanisms Potentially Regulate the Disease Time-Course of Duchenne Muscular Dystrophy and Provide Targets for Therapeutic Intervention. *PM&R*, 1(8), pp.755–768. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19695529> [Accessed November 24, 2017].
- Fath, S. et al., 2011. Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. *PLoS one*, 6(3), p.e17596. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3048298&tool=pmcentrez&rendertype=abstract> [Accessed July 16, 2015].
- Faysoil, A., Abasse, S. & Silverston, K., 2017. Cardiac Involvement Classification and Therapeutic Management in Patients with Duchenne Muscular Dystrophy. *Journal of Neuromuscular Diseases*, 4, pp.17–23. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5345647/pdf/jnd-4-jnd160194.pdf> [Accessed November 24, 2017].
- Fekairi, S. et al., 2009. Human SLX4 Is a Holliday Junction Resolvase Subunit that Binds Multiple DNA Repair/Recombination Endonucleases. *Cell*, 138(1), pp.78–89. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19596236> [Accessed November 23, 2017].
- Ferrand, M. et al., 2015. Serotype-specific Binding Properties and Nanoparticle Characteristics Contribute to the Immunogenicity of rAAV1 Vectors. *Molecular therapy : the journal of the American Society of Gene Therapy*, 23(6), pp.1022–1033. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25881000> [Accessed November 17, 2017].
- Finkel, R.S. et al., 2013. Phase 2a Study of Ataluren-Mediated Dystrophin Production in Patients with Nonsense Mutation Duchenne Muscular Dystrophy H. Sawada, ed. *PLoS ONE*, 8(12), p.e81302. Available at: <http://dx.plos.org/10.1371/journal.pone.0081302> [Accessed November 29, 2017].
- Foster, H. et al., 2008. Codon and mRNA sequence optimization of microdystrophin transgenes improves expression and physiological outcome in dystrophic mdx mice following AAV2/8 gene transfer. *Molecular therapy : the journal of the American Society of Gene Therapy*, 16(11), pp.1825–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18766174> [Accessed February 11, 2014].
- Frazier, K. et al., 1996. Stimulation of fibroblast cell growth, matrix production, and granulation tissue formation by connective tissue growth factor. *Journal of Investigative Dermatology*, 107(3), pp.404–411. Available at: <http://www.sciencedirect.com/science/article/pii/S0022202X15426867?via%3Dihub> [Accessed November 24, 2017].
- Fu, Y. et al., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology*, 31(9), pp.822–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23792628> [Accessed January 20, 2014].
- Fu, Y. et al., 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.

Nature biotechnology, 32(3), pp.279–84. Available at:
<http://dx.doi.org/10.1038/nbt.2808> [Accessed July 9, 2014].

- Gagnon, J.A. et al., 2014. Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs B. Riley, ed. *PLoS ONE*, 9(5), p.e98186. Available at: <http://dx.plos.org/10.1371/journal.pone.0098186> [Accessed November 6, 2017].
- Gallagher, K., Sambridge, M. & Drikkoningen, G., 1991. Genetic algorithms: An evolution from Monte Carlo Methods for strongly non-linear geophysical optimization problems. *Geophysical Research Letters*, 18(12), pp.2177–2180. Available at: <http://doi.wiley.com/10.1029/91GL02368> [Accessed July 14, 2015].
- Gao, W. et al., 2018. Risk analysis for genome editing-derived food safety in China. *Food Control*, 84, pp.128–137. Available at: <http://www.sciencedirect.com/science/article/pii/S095671351730378X> [Accessed November 26, 2017].
- Gilbert, L.A. et al., 2013. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, 154(2), pp.442–451. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23849981> [Accessed November 5, 2017].
- Godfrey, C. et al., 2015. How much dystrophin is enough: the physiological consequences of different levels of dystrophin in the *mdx* mouse. *Human Molecular Genetics*, 24(15), pp.4225–4237. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25935000> [Accessed July 28, 2017].
- Goemans, N.M. et al., 2016. Long-Term Efficacy, Safety, and Pharmacokinetics of Drisapersen in Duchenne Muscular Dystrophy: Results from an Open-Label Extension Study M. D. Cordero, ed. *PLOS ONE*, 11(9), p.e0161955. Available at: <http://dx.plos.org/10.1371/journal.pone.0161955> [Accessed November 24, 2017].
- Goemans, N.M. et al., 2011. Systemic administration of PRO051 in Duchenne’s muscular dystrophy. *The New England journal of medicine*, 364(16), pp.1513–22. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21428760> [Accessed February 11, 2014].
- Gonçalves, M.A.F. V. & de Vries, A.A.F., 2006. Adenovirus: from foe to friend. *Reviews in Medical Virology*, 16(3), pp.167–186. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16710837> [Accessed November 17, 2017].
- Górecki, D.C. et al., 1992. Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. *Human molecular genetics*, 1(7), pp.505–10. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1307251> [Accessed July 10, 2015].
- Goudenege, S. et al., 2012. Myoblasts derived from normal hESCs and dystrophic hiPSCs efficiently fuse with existing muscle fibers following transplantation. *Molecular therapy : the journal of the American Society of Gene Therapy*, 20(11), pp.2153–67. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22990676> [Accessed November 17, 2017].
- Graf, M. et al., 2000. Concerted action of multiple cis-acting sequences is required for Rev dependence of late human immunodeficiency virus type 1 gene expression. *Journal of virology*, 74(22), pp.10822–6. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=110961&tool=pmcentrez&rendertype=abstract> [Accessed July 16, 2015].
- Graf, M., Deml, L. & Wagner, R., 2004. Codon-optimized genes that enable increased

- heterologous expression in mammalian cells and elicit efficient immune responses in mice after vaccination of naked DNA. *Methods in molecular medicine*, 94, pp.197–210. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14959831> [Accessed July 16, 2015].
- Gray, S.J. et al., 2011. Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. *Human gene therapy*, 22(9), pp.1143–53. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21476867> [Accessed July 16, 2017].
- Grens, K., 2017. First In Vivo Human Genome Editing to Be Tested in New Clinical Trial | The Scientist Magazine®. Available at: <https://www.the-scientist.com/?articles.view/articleNo/49456/title/First-In-Vivo-Human-Genome-Editing-to-Be-Tested-in-New-Clinical-Trial/> [Accessed November 25, 2017].
- Guha, T.K., Wai, A. & Hausner, G., 2017. Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering. *Computational and Structural Biotechnology Journal*, 15, pp.146–160. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28179977> [Accessed November 26, 2017].
- Le Guiner, C. et al., 2017. Long-term microdystrophin gene therapy is effective in a canine model of Duchenne muscular dystrophy. *Nature Communications*, 8, p.16105. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28742067> [Accessed September 8, 2017].
- Guiraud, S., Chen, H., et al., 2015. Advances in genetic therapeutic strategies for Duchenne muscular dystrophy. *Experimental physiology*. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26140505> [Accessed July 14, 2015].
- Guiraud, S., Squire, S.E., et al., 2015. Second-generation compound for the modulation of utrophin in the therapy of DMD. *Human molecular genetics*, 24(15), pp.4212–24. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4492389&tool=pmcentrez&rendertype=abstract> [Accessed September 26, 2015].
- Gupta, A. et al., 2012. An optimized two-finger archive for ZFN-mediated gene targeting. *Nature methods*, 9(6), pp.588–90. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22543349> [Accessed November 26, 2017].
- Gurvich, O.L. et al., 2009. DMD exon 1 truncating point mutations: amelioration of phenotype by alternative translation initiation in exon 6. *Human mutation*, 30(4), pp.633–40. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19206170> [Accessed June 4, 2018].
- Guschin, D.Y. et al., 2010. A rapid and general assay for monitoring endogenous gene modification. *Methods in molecular biology (Clifton, N.J.)*, 649, pp.247–56. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20680839> [Accessed April 5, 2016].
- Gutschner, T. et al., 2016. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Reports*, 14(6), pp.1555–1566. Available at: <http://dx.doi.org/10.1016/j.celrep.2016.01.019>.
- Haas, M. et al., 2015. European Medicines Agency review of ataluren for the treatment of ambulant patients aged 5 years and older with Duchenne muscular dystrophy resulting from a nonsense mutation in the dystrophin gene. *Neuromuscular Disorders*, 25(1), pp.5–13. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25497400> [Accessed November 29, 2017].
- Hacein-Bey-Abina, S. et al., 2010. Efficacy of gene therapy for X-linked severe combined

- immunodeficiency. *The New England journal of medicine*, 363(4), pp.355–64.
- Hacein-Bey-Abina, S. et al., 2002. Sustained Correction of X-Linked Severe Combined Immunodeficiency by ex Vivo Gene Therapy. *New England Journal of Medicine*, 346(16), pp.1185–1193. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11961146> [Accessed November 25, 2017].
- Haeussler, M. et al., 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, 17(1), p.148. Available at: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1012-2> [Accessed October 12, 2017].
- Hakim, C.H. et al., 2017. A Five-Repeat Micro-Dystrophin Gene Ameliorated Dystrophic Phenotype in the Severe DBA/2J-mdx Model of Duchenne Muscular Dystrophy. *Molecular therapy. Methods & clinical development*, 6, pp.216–230. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28932757> [Accessed November 25, 2017].
- Harman, D., 1956. Aging: A Theory Based on Free Radical and Radiation Chemistry. *Journal of Gerontology*, 11(3), pp.298–300. Available at: https://www.uccs.edu/Documents/rmelamed/harman_1956_13332224.pdf [Accessed November 21, 2017].
- Harper, S.Q. et al., 2002. Modular flexibility of dystrophin: implications for gene therapy of Duchenne muscular dystrophy. *Nature medicine*, 8(3), pp.253–61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11875496> [Accessed August 3, 2015].
- Harris, J.B., 2003. Myotoxic phospholipases A2 and the regeneration of skeletal muscles. *Toxicon : official journal of the International Society on Toxinology*, 42(8), pp.933–45. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15019492> [Accessed November 17, 2017].
- Hauser, E. et al., 1995. Oxyradical Damage and Mitochondrial Enzyme Activities in the mdx Mouse. *Neuropediatrics*, 26(5), pp.260–262. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8552217> [Accessed November 28, 2017].
- Heilig, R. et al., 1987. Localization of the region homologous to the Duchenne muscular dystrophy locus on the mouse X chromosome. *Nature*, 328(6126), pp.168–170. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3600794> [Accessed August 6, 2017].
- Helliwell, T.R. et al., 1992. The dystrophin-related protein, utrophin, is expressed on the sarcolemma of regenerating human skeletal muscle fibres in dystrophies and inflammatory myopathies. *Neuromuscular disorders : NMD*, 2(3), pp.177–84. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1483043> [Accessed November 28, 2017].
- Heslop, L., Morgan, J.E. & Partridge, T.A., 2000. Evidence for a myogenic stem cell that is exhausted in dystrophic muscle. *Journal of Cell Science*, 113(12).
- Heyer, W.-D., Ehmsen, K.T. & Liu, J., 2010. Regulation of homologous recombination in eukaryotes. *Annual review of genetics*, 44, pp.113–39. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20690856> [Accessed June 21, 2017].
- Hirawat, S. et al., 2007. Safety, Tolerability, and Pharmacokinetics of PTC124, a Nonaminoglycoside Nonsense Mutation Suppressor, Following Single- and Multiple-Dose Administration to Healthy Male and Female Adult Volunteers. *The Journal of Clinical Pharmacology*, 47(4), pp.430–444. Available at: <http://doi.wiley.com/10.1177/0091270006297140> [Accessed November 8, 2017].

- Hoeijmakers, J.H.J., 2009. DNA Damage, Aging, and Cancer. *New England Journal of Medicine*, 361(15), pp.1475–1485. Available at: <http://www.nejm.org/doi/abs/10.1056/NEJMra0804615> [Accessed November 21, 2017].
- Hoffman, E.P., Brown, R.H. & Kunkel, L.M., 1987. Dystrophin: The protein product of the duchenne muscular dystrophy locus. *Cell*, 51(6), pp.919–928.
- Hollingworth, S., Zeiger, U. & Baylor, S.M., 2008. Comparison of the myoplasmic calcium transient elicited by an action potential in intact fibres of mdx and normal mice. *The Journal of physiology*, 586(21), pp.5063–75. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18772198> [Accessed November 20, 2017].
- Hopf, F.W. et al., 1996. A critical evaluation of resting intracellular free calcium regulation in dystrophic mdx muscle. *The American journal of physiology*, 271(4 Pt 1), pp.C1325–39. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8897840> [Accessed November 20, 2017].
- Horiguchi, M., Ota, M. & Rifkin, D.B., 2012. Matrix control of transforming growth factor- β function. *Journal of biochemistry*, 152(4), pp.321–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22923731> [Accessed November 23, 2017].
- Horvath, P. & Barrangou, R., 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science (New York, N.Y.)*, 327(5962), pp.167–70. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20056882> [Accessed January 21, 2014].
- Hou, Z. et al., 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proceedings of the National Academy of Sciences of the United States of America*, 110(39), pp.15644–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3785731&tool=pmcentrez&rendertype=abstract> [Accessed January 24, 2014].
- Howe, S.J. et al., 2008. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *The Journal of clinical investigation*, 118(9), pp.3143–50. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2496964&tool=pmcentrez&rendertype=abstract> [Accessed July 10, 2015].
- Hsu, P.D. et al., 2013a. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, 31(9), pp.827–832. Available at: <http://www.nature.com/doi/abs/10.1038/nbt.2647> [Accessed October 12, 2017].
- Hsu, P.D. et al., 2013b. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9), pp.827–32. Available at: <http://www.nature.com/doi/abs/10.1038/nbt.2647> [Accessed July 10, 2014].
- Huchet, C. et al., 2017. Exhaustive characterization of the newly developed Duchenne muscular dystrophy rat model: a unique animal model for DMD which mimics the human disease at both the muscular and the cardiac levels. *Neuromuscular Disorders*, 27, p.S247. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0960896617311197> [Accessed December 13, 2017].
- Huen, M.S.Y., Sy, S.M.H. & Chen, J., 2010. BRCA1 and its toolbox for the maintenance of genome integrity. *Nature Reviews Molecular Cell Biology*, 11(2), pp.138–148. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20029420> [Accessed November 23, 2017].
- Invitrogen, 2017. Zeocin Selection Manual User Guide.

- Ip, S.C.Y. et al., 2008. Identification of Holliday junction resolvases from humans and yeast. *Nature*, 456(7220), pp.357–361. Available at: <http://www.nature.com/doi/10.1038/nature07470> [Accessed November 23, 2017].
- Ishino, Y. et al., 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology*, 169(12), pp.5429–33. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=213968&tool=pmcentrez&endertype=abstract>.
- Iwata, Y. et al., 2009. Dominant-negative inhibition of Ca²⁺ influx via TRPV2 ameliorates muscular dystrophy in animal models. *Human Molecular Genetics*, 18(5), pp.824–834. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19050039> [Accessed November 20, 2017].
- De Jager, M. et al., 2001. Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Molecular Cell*, 8(5), pp.1129–1135. Available at: <http://www.sciencedirect.com/science/article/pii/S1097276501003811?via%3Dihub> [Accessed November 22, 2017].
- Jarmin, S. et al., 2014. New developments in the use of gene therapy to treat Duchenne muscular dystrophy. *Expert opinion on biological therapy*, 14(2), pp.209–30. Available at: <http://informahealthcare.com/doi/abs/10.1517/14712598.2014.866087?af=R&> [Accessed February 11, 2014].
- Jasin, M. & Berg, P., 1988. Homologous integration in mammalian cells without target gene selection. *Genes & development*, 2(11), pp.1353–63. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2850258> [Accessed June 28, 2017].
- Jasin, M. & Rothstein, R., 2013. Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harbor Perspectives in Biology*, 5(11), pp.a012740–a012740. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24097900> [Accessed May 24, 2017].
- Jeffreys, A.J., 1979. DNA sequence variants in the G^γ-, A^γ-, δ - and ϵ -globin genes of man. *Cell*, 18(1), pp.1–10. Available at: <http://linkinghub.elsevier.com/retrieve/pii/0092867479903489> [Accessed July 3, 2017].
- Jinek, M. et al., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096), pp.816–21. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22745249> [Accessed July 9, 2014].
- Jinek, M. et al., 2013. RNA-programmed genome editing in human cells. *eLife*, 2, p.e00471. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3557905&tool=pmcentrez&rendertype=abstract> [Accessed July 10, 2014].
- Jinek, M. et al., 2014. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science*, 343(6176), pp.1247997–1247997. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24505130> [Accessed November 6, 2017].
- Jones, D.P., 2006. Redefining Oxidative Stress. *Antioxidants & Redox Signaling*, 8(9–10), pp.1865–1879. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16987039> [Accessed November 28, 2017].
- Kabadi, A.M. et al., 2014. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Research*, 42(19), pp.e147–e147. Available at:

<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku749> [Accessed June 13, 2017].

- Kalvala, A. et al., 2010. Enhancement of gene targeting in human cells by intranuclear permeation of the *Saccharomyces cerevisiae* Rad52 protein. *Nucleic Acids Research*, 38(14), pp.e149–e149. Available at: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq486> [Accessed June 24, 2017].
- Karin, M., 1999. How NF- κ B is activated: the role of the I κ B kinase (IKK) complex. *Oncogene*, 18(49), pp.6867–6874. Available at: <https://www.nature.com/articles/1203219.pdf> [Accessed November 24, 2017].
- Kay, M.A., He, C.-Y. & Chen, Z.-Y., 2010. A robust system for production of minicircle DNA vectors. *Nature Biotechnology*, 28(12), pp.1287–1289. Available at: <http://www.nature.com/doifinder/10.1038/nbt.1708> [Accessed November 16, 2017].
- Kay, S. et al., 2007. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science (New York, N.Y.)*, 318(5850), pp.648–51. Available at: <http://www.sciencemag.org/content/318/5850/648> [Accessed January 27, 2014].
- Kent, W.J. et al., 2002. The human genome browser at UCSC. *Genome research*, 12(6), pp.996–1006. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12045153> [Accessed October 12, 2017].
- Khairallah, R.J. et al., 2012. Microtubules underlie dysfunction in duchenne muscular dystrophy. *Science signaling*, 5(236), p.ra56. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22871609> [Accessed November 28, 2017].
- Kharraz, Y. et al., 2014. Understanding the process of fibrosis in Duchenne muscular dystrophy. *BioMed research international*, 2014, p.965631. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24877152> [Accessed November 23, 2017].
- Kim, D. et al., 2015. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature Methods*, 12(3), pp.237–243. Available at: <http://www.nature.com/doifinder/10.1038/nmeth.3284> [Accessed October 18, 2017].
- Kim, H. & Kim, J.-S., 2014a. A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics*, advance on. Available at: <http://dx.doi.org/10.1038/nrg3686> [Accessed April 2, 2014].
- Kim, H. & Kim, J.-S., 2014b. A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics*, 15(5), pp.321–334. Available at: <http://www.nature.com/doifinder/10.1038/nrg3686> [Accessed November 26, 2017].
- Kim, J.M. et al., 2014. Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nature Communications*, 5, pp.1–7. Available at: <http://www.nature.com/doifinder/10.1038/ncomms4157>.
- Kim, S. et al., 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research*, 24(6), pp.1012–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4032847&tool=pmcentrez&rendertype=abstract> [Accessed July 9, 2014].
- Kim, Y.G., Cha, J. & Chandrasegaran, S., 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, 93(3), pp.1156–60. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8577732> [Accessed July 14, 2017].

- Kinali, M. et al., 2009. Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. *The Lancet. Neurology*, 8(10), pp.918–28. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19713152> [Accessed November 24, 2017].
- Kirkwood, T.B.L., 2005. Understanding the odd science of aging. *Cell*, 120(4), pp.437–447. Available at: <http://www.sciencedirect.com/science/article/pii/S0092867405001017?via%3Dihub> [Accessed November 21, 2017].
- Kleinstiver, B.P. et al., 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529(7587), pp.490–495. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26735016> [Accessed November 17, 2017].
- Klungland, A. et al., 1999. Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proceedings of the National Academy of Sciences of the United States of America*, 96(23), pp.13300–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10557315> [Accessed November 21, 2017].
- Koenig, M. et al., 1987. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*, 50(3), pp.509–17. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3607877> [Accessed June 1, 2015].
- Koenig, M. et al., 1988. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell*, 53(2), pp.219–28. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3282674> [Accessed August 6, 2017].
- Koo, T. et al., 2011. Long-term functional adeno-associated virus-microdystrophin expression in the dystrophic CXMDj dog. *The Journal of Gene Medicine*, 13(9), pp.497–506. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22144143> [Accessed September 8, 2017].
- Koo, T. et al., 2014. Triple trans-splicing adeno-associated virus vectors capable of transferring the coding sequence for full-length dystrophin protein into dystrophic mice. *Human gene therapy*, 25(2), pp.98–108. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24191945> [Accessed June 7, 2015].
- Kotin, R.M. et al., 1990. Site-specific integration by adeno-associated virus. *Proceedings of the National Academy of Sciences*, 87(6), pp.2211–2215. Available at: <http://www.pnas.org/content/87/6/2211.short> [Accessed February 11, 2014].
- Kottlors, M. & Kirschner, J., 2010. Elevated satellite cell number in Duchenne muscular dystrophy. *Cell and Tissue Research*, 340(3), pp.541–548. Available at: <http://link.springer.com/10.1007/s00441-010-0976-6> [Accessed November 28, 2017].
- Lai, Y. et al., 2009. Dystrophins carrying spectrin-like repeats 16 and 17 anchor nNOS to the sarcolemma and enhance exercise performance in a mouse model of muscular dystrophy. *The Journal of clinical investigation*, 119(3), pp.624–35. Available at: <http://www.jci.org/articles/view/36612> [Accessed September 8, 2017].
- Lai, Y. et al., 2005. Efficient in vivo gene expression by trans-splicing adeno-associated viral vectors. *Nature biotechnology*, 23(11), pp.1435–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16244658> [Accessed November 5, 2017].
- Lamb, B.M., Mercer, A.C. & Barbas, C.F., 2013. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Research*, 41(21), pp.9779–9785.

Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23980031> [Accessed November 27, 2017].

- Lander, E.S. et al., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822), pp.860–921. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11237011> [Accessed November 5, 2017].
- Larcher, T. et al., 2014. Characterization of dystrophin deficient rats: a new model for Duchenne muscular dystrophy. *PloS one*, 9(10), p.e110371. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4195719&tool=pmcentrez&rendertype=abstract> [Accessed September 26, 2015].
- Latres, E. et al., 2015. Myostatin blockade with a fully human monoclonal antibody induces muscle hypertrophy and reverses muscle atrophy in young and aged mice. *Skeletal Muscle*, 5(1), p.34. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26457176> [Accessed November 29, 2017].
- Lattanzi, A. et al., 2017. Correction of the Exon 2 Duplication in DMD Myoblasts by a Single CRISPR/Cas9 System. *Molecular therapy. Nucleic acids*, 7, pp.11–19. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28624187> [Accessed July 2, 2017].
- Leahy, J.J.J. et al., 2004. *Identification of a highly potent and selective DNA-dependent protein kinase (DNA-PK) inhibitor (NU7441) by screening of chromenone libraries*, Available at: <http://www.sciencedirect.com/science/article/pii/S0960894X04011837> [Accessed June 19, 2017].
- Lee, K. et al., 2017. Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nature Biomedical Engineering*, 1(11), pp.889–901. Available at: <http://www.nature.com/articles/s41551-017-0137-2> [Accessed November 17, 2017].
- Lee, S.-J. et al., 2010. Regulation of muscle mass by follistatin and activins. *Molecular endocrinology (Baltimore, Md.)*, 24(10), pp.1998–2008. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20810712> [Accessed November 29, 2017].
- Lee, S.-J., 2004. REGULATION OF MUSCLE MASS BY MYOSTATIN. *Annual Review of Cell and Developmental Biology*, 20(1), pp.61–86. Available at: <http://www.annualreviews.org/doi/10.1146/annurev.cellbio.20.012103.135836> [Accessed November 29, 2017].
- Li, D. et al., 2010. Sarcolemmal nNOS anchoring reveals a qualitative difference between dystrophin and utrophin. *Journal of Cell Science*, 123(12), pp.2008–2013. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20483958> [Accessed November 29, 2017].
- Li, H., Malhotra, S. & Kumar, A., 2008. Nuclear factor-kappa B signaling in skeletal muscle atrophy. *Journal of Molecular Medicine*, 86(10), pp.1113–1126. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18574572> [Accessed November 24, 2017].
- Li, H.L. et al., 2015. Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem cell reports*, 4(1), pp.143–54. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4297888&tool=pmcentrez&rendertype=abstract> [Accessed March 25, 2015].
- Li, L., Wu, L.P. & Chandrasegaran, S., 1992. Functional domains in Fok I restriction endonuclease. *Proceedings of the National Academy of Sciences of the United States of*

- America*, 89(10), pp.4275–9. Available at:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=49064&tool=pmcentrez&rendertype=abstract>.
- Li, X. et al., 1999. Synthetic muscle promoters: activities exceeding naturally occurring regulatory sequences. *Nature biotechnology*, 17(3), pp.241–245. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/10096290> [Accessed July 28, 2017].
- Li, Y. et al., 2004. Transforming growth factor-beta1 induces the differentiation of myogenic cells into fibrotic cells in injured skeletal muscle: a key event in muscle fibrogenesis. *The American journal of pathology*, 164(3), pp.1007–19. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/14982854> [Accessed November 23, 2017].
- Liang, F. et al., 1998. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, 95(9), pp.5172–7. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/9560248> [Accessed November 23, 2017].
- Liang, P. et al., 2015. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein & Cell*, 6(5), pp.363–372. Available at: <http://link.springer.com/10.1007/s13238-015-0153-5> [Accessed November 17, 2017].
- Liang, X. et al., 2015. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology*, 208, pp.44–53. Available at:
<http://linkinghub.elsevier.com/retrieve/pii/S016816561500200X> [Accessed June 13, 2017].
- Lidov, H.G.W., Selig, S. & Kunkel, L.M., 1995. Dp140: a novel 140 kDa CNS transcript from the dystrophin locus. *Human Molecular Genetics*, 4(3), pp.329–335. Available at:
<https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/4.3.329> [Accessed August 6, 2017].
- Lin, S. et al., 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife*, 3, p.e04766. Available at:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4383097&tool=pmcentrez&rendertype=abstract> [Accessed December 18, 2014].
- Lipton, B.H. & Schultz, E., 1979. Developmental fate of skeletal muscle satellite cells. *Science (New York, N.Y.)*, 205(4412), pp.1292–4. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/472747> [Accessed November 17, 2017].
- Liu, X. et al., 2016a. Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Scientific Reports*, 6(1), p.19675. Available at:
<http://www.nature.com/articles/srep19675> [Accessed October 18, 2017].
- Liu, X. et al., 2016b. Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Scientific Reports*, 6(1), p.19675. Available at:
<http://www.nature.com/articles/srep19675> [Accessed November 5, 2017].
- Liu, Z. et al., 2017. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific reports*, 7(1), p.2193. Available at:
<http://www.nature.com/articles/s41598-017-02460-2>
<http://www.ncbi.nlm.nih.gov/pubmed/28526819>
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5438344>.
- Long, C. et al., 2016. Postnatal genome editing partially restores dystrophin expression in a

- mouse model of muscular dystrophy. *Science (New York, N.Y.)*, 351(6271), pp.400–3. Available at: <http://www.sciencemag.org/cgi/doi/10.1126/science.aad5725> [Accessed November 16, 2017].
- Long, C. et al., 2014. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science (New York, N.Y.)*, p.science.1254445-. Available at: <http://www.sciencemag.org/content/early/2014/08/13/science.1254445> [Accessed August 17, 2014].
- Love, D.R. et al., 1989. An autosomal transcript in skeletal muscle with homology to dystrophin. *Nature*, 339(6219), pp.55–58. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2541343> [Accessed November 29, 2017].
- Love, D.R. et al., 1993. Dystrophin and dystrophin-related proteins: A review of protein and RNA studies. *Neuromuscular Disorders*, 3(1), pp.5–21.
- Lovering, R.M. & De Deyne, P.G., 2004. Contractile function, sarcolemma integrity, and the loss of dystrophin after skeletal muscle eccentric contraction-induced injury. *American journal of physiology. Cell physiology*, 286(2), pp.C230-8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14522817> [Accessed November 19, 2017].
- Lu-Nguyen, N.B. et al., 2015. Combination Antisense Treatment for Destructive Exon Skipping of Myostatin and Open Reading Frame Rescue of Dystrophin in Neonatal mdx Mice. *Molecular Therapy*, 23(8), pp.1341–1348. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25959011> [Accessed November 25, 2017].
- Lu, A. et al., 2012. NF-κB negatively impacts the myogenic potential of muscle-derived stem cells. *Molecular therapy : the journal of the American Society of Gene Therapy*, 20(3), pp.661–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22158056> [Accessed November 24, 2017].
- Ma, H. et al., 2014. Pol III Promoters to Express Small RNAs: Delineation of Transcription Initiation. *Molecular Therapy - Nucleic Acids*, 3(7536), p.e161. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S2162253116303018> [Accessed July 14, 2017].
- Machitani, M. et al., 2011. Adenovirus Vector-Derived VA-RNA-Mediated Innate Immune Responses. *Pharmaceutics*, 3(4), pp.338–353. Available at: <http://www.mdpi.com/1999-4923/3/3/338/> [Accessed November 17, 2017].
- Maggio, I. et al., 2014. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. *Scientific Reports*, 4, pp.1–11. Available at: <http://www.nature.com/articles/srep05105>.
- Maggio, I. et al., 2016. Adenoviral vectors encoding CRISPR/Cas9 multiplexes rescue dystrophin synthesis in unselected populations of DMD muscle cells. *Scientific Reports*, 6(1), p.37051. Available at: <http://www.nature.com/articles/srep37051>.
- Mah, J.K. et al., 2014. A systematic review and meta-analysis on the epidemiology of Duchenne and Becker muscular dystrophy. *Neuromuscular Disorders*, 24(6), pp.482–491. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24780148> [Accessed August 6, 2017].
- Mak, A.N.-S. et al., 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science (New York, N.Y.)*, 335(6069), pp.716–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3427646&tool=pmcentrez&rendertype=abstract> [Accessed January 23, 2014].

- Makarova, K.S. et al., 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, 1(1), p.7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16545108> [Accessed November 27, 2017].
- Makarova, K.S. et al., 2011. Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, 9(6), pp.467–477. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21552286> [Accessed November 27, 2017].
- Makrides, S., 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.*, 60(3), pp.512–538. Available at: <http://mmbr.asm.org/content/60/3/512.abstract> [Accessed July 16, 2015].
- Mali, P., Aach, J., et al., 2013. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology*, 31(9), pp.833–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23907171> [Accessed January 21, 2014].
- Mali, P., Yang, L., et al., 2013. RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, 339(6121), pp.823–6. Available at: <http://www.sciencemag.org/content/339/6121/823.abstract> [Accessed January 20, 2014].
- Mali, P., Esvelt, K.M. & Church, G.M., 2013. Cas9 as a versatile tool for engineering biology. *Nature methods*, 10(10), pp.957–63. Available at: <http://dx.doi.org/10.1038/nmeth.2649> [Accessed January 20, 2014].
- Malik, V. et al., 2010. Aminoglycoside-induced mutation suppression (stop codon readthrough) as a therapeutic strategy for Duchenne muscular dystrophy. *Therapeutic advances in neurological disorders*, 3(6), pp.379–89. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21179598> [Accessed November 29, 2017].
- Malina, A. et al., 2015. PAM multiplicity marks genomic target sites as inhibitory to CRISPR-Cas9 editing. *Nature communications*, 6, p.10124. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26644285> [Accessed October 18, 2017].
- Mamchaoui, K. et al., 2011. Immortalized pathological human myoblasts: towards a universal tool for the study of neuromuscular disorders. *Skeletal muscle*, 1, p.34. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22040608> [Accessed November 11, 2017].
- Manning, J. & O'Malley, D., 2015. What has the mdx mouse model of duchenne muscular dystrophy contributed to our understanding of this disease? *Journal of muscle research and cell motility*, 36(2), pp.155–67. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25669899> [Accessed April 26, 2015].
- Mao, Z. et al., 2008. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell cycle (Georgetown, Tex.)*, 7(18), pp.2902–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18769152> [Accessed November 21, 2017].
- Maresca, M. et al., 2013. Obligate ligation-gated recombination (ObLiGaRe): Custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Research*, 23(3).
- Marie, C. et al., 2010. pFARs, Plasmids free of antibiotic resistance markers, display high-level

- transgene expression in muscle, skin and tumour cells. *The Journal of Gene Medicine*, 12(4), pp.323–332. Available at: <http://doi.wiley.com/10.1002/jgm.1441> [Accessed September 6, 2017].
- Martin, E.A. et al., 2012. Tadalafil alleviates muscle ischemia in patients with Becker muscular dystrophy. *Science translational medicine*, 4(162), p.162ra155. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23197572> [Accessed August 6, 2017].
- Maruyama, T. et al., 2016. Corrigendum: Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature Biotechnology*, 34(2), pp.210–210. Available at: <http://www.nature.com/doifinder/10.1038/nbt0216-210c>.
- Maruyama, T. et al., 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature Biotechnology*, 33(5), pp.538–542. Available at: <http://www.nature.com/doifinder/10.1038/nbt.3190> [Accessed November 6, 2017].
- Massague, J., Seoane, J. & Wotton, D., 2005. Smad transcription factors. *Genes & Development*, 19(23), pp.2783–2810. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16322555> [Accessed November 24, 2017].
- Massouridès, E. et al., 2015. Dp412e: a novel human embryonic dystrophin isoform induced by BMP4 in early differentiated cells. *SKELETAL MUSCLE*, 5(1), p.40. Available at: <http://www.skeletalmusclejournal.com/content/5/1/40> [Accessed August 6, 2017].
- de Massy, B., Weisberg, R.A. & Studier, F.W., 1987. Gene 3 endonuclease of bacteriophage T7 resolves conformationally branched structures in double-stranded DNA. *Journal of Molecular Biology*, 193(2), pp.359–376. Available at: <http://www.sciencedirect.com/science/article/pii/0022283687902245?via%3Dihub> [Accessed December 6, 2017].
- McDonald, C.M. et al., 2013. THE 6-minute walk test and other endpoints in Duchenne muscular dystrophy: Longitudinal natural history observations over 48 weeks from a multicenter study. *Muscle & Nerve*, 48(3), pp.343–356. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23681930> [Accessed November 29, 2017].
- McDonald, C.M. et al., 2017. Ataluren in patients with nonsense mutation Duchenne muscular dystrophy (ACT DMD): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet*, 390(10101), pp.1489–1498. Available at: <http://www.sciencedirect.com.ezproxy01.rhul.ac.uk/science/article/pii/S0140673617316112> [Accessed November 8, 2017].
- McDonald, C.M. et al., 2010. The 6-minute walk test as a new outcome measure in Duchenne muscular dystrophy. *Muscle & Nerve*, 41(4), pp.500–510. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19941337> [Accessed November 29, 2017].
- McPherron, A.C., Lawler, A.M. & Lee, S.-J., 1997. Regulation of skeletal muscle mass in mice by a new TGF- β superfamily member. *nature*, 387(6628), pp.83–90. Available at: <http://www.nature.com/doifinder/10.1038/387083a0> [Accessed November 29, 2017].
- McPherron, A.C. & Lee, S.J., 1997. Double muscling in cattle due to mutations in the myostatin gene. *Proceedings of the National Academy of Sciences of the United States of America*, 94(23), pp.12457–61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9356471> [Accessed November 29, 2017].

- McWilliam, H. et al., 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic acids research*, 41(Web Server issue), pp.W597-600. Available at: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt376> [Accessed October 12, 2017].
- Meckler, J.F. et al., 2013. Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic acids research*, 41(7), pp.4118–28. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3627578&tool=pmcentrez&rendertype=abstract> [Accessed January 23, 2014].
- Mendell, J.R. et al., 2015. A Phase 1/2a Follistatin Gene Therapy Trial for Becker Muscular Dystrophy. *Molecular Therapy*, 23(1), pp.192–201. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25322757> [Accessed November 29, 2017].
- Mendell, J.R. et al., 2010. Dystrophin Immunity in Duchenne’s Muscular Dystrophy. *New England Journal of Medicine*, 363(15), pp.1429–1437. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20925545> [Accessed November 25, 2017].
- Mendell, J.R. et al., 2010. Dystrophin immunity in Duchenne’s muscular dystrophy. *The New England journal of medicine*, 363(15), pp.1429–37. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3014106&tool=pmcentrez&rendertype=abstract> [Accessed November 25, 2015].
- Mendell, J.R. et al., 2013. Eteplirsen for the treatment of Duchenne muscular dystrophy. *Annals of neurology*, 74(5), pp.637–47. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23907995> [Accessed November 19, 2015].
- Mendell, J.R. et al., 2012. Gene therapy for muscular dystrophy: lessons learned and path forward. *Neuroscience letters*, 527(2), pp.90–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22609847> [Accessed November 25, 2017].
- Mendell, J.R. et al., 2016. Longitudinal effect of eteplirsen versus historical control on ambulation in Duchenne muscular dystrophy. *Annals of Neurology*, 79(2), pp.257–271. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26573217> [Accessed November 24, 2017].
- Mendell, J.R., Sahenk, Z. & Rodino-Klapac, L.R., 2017. Clinical trials of exon skipping in Duchenne muscular dystrophy. *Expert Opinion on Orphan Drugs*, 5(9), pp.683–690. Available at: <http://www.tandfonline.com/action/journalInformation?journalCode=ieod20> [Accessed November 24, 2017].
- Meng, J. et al., 2016. Autologous skeletal muscle derived cells expressing a novel functional dystrophin provide a potential therapy for Duchenne Muscular Dystrophy. *Scientific Reports*, 6(1), p.19750. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26813695> [Accessed September 8, 2017].
- Mercuri, E. & Muntoni, F., 2013. Muscular dystrophies. *Lancet*, 381(9869), pp.845–60. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23465426> [Accessed February 2, 2014].
- Messina, S. et al., 2006. Lipid peroxidation inhibition blunts nuclear factor-kappaB activation, reduces skeletal muscle degeneration, and enhances muscle function in mdx mice. *The American journal of pathology*, 168(3), pp.918–26. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16507907> [Accessed November 28, 2017].

- Miller, J.C. et al., 2011. A TALE nuclease architecture for efficient genome editing. *Nature biotechnology*, 29(2), pp.143–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21179091> [Accessed January 21, 2014].
- Miller, J.C. et al., 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnology*, 25(7), pp.778–785. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17603475> [Accessed November 26, 2017].
- Mitani, K. et al., 1995. Rescue, propagation, and partial purification of a helper virus-dependent adenovirus vector. *Proceedings of the National Academy of Sciences of the United States of America*, 92(9), pp.3854–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7731995> [Accessed November 17, 2017].
- Mojica, F.J. et al., 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology*, 36(1), pp.244–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10760181> [Accessed November 27, 2017].
- Monaco, A.P. et al., 1988. An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics*, 2(1), pp.90–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3384440> [Accessed November 28, 2017].
- Monaco, A.P., 1989. Dystrophin, the protein product of the Duchenne/Becker muscular dystrophy gene. *Trends in Biochemical Sciences*, 14(10), pp.412–415. Available at: <http://linkinghub.elsevier.com/retrieve/pii/0968000489902909> [Accessed June 26, 2017].
- Moore, M. et al., 2015. Exciting developments in CRISPR/Cas9-mediated approaches for Duchenne MD. *Cell and Gene Therapy Insights*, 1(2). Available at: http://insights.bio/cell-and-gene-therapy-insights/?bio_journals=exciting-developments-in-crisprcas9-mediated-approaches-for-duchenne-muscular-dystrophy [Accessed November 7, 2017].
- Moreno-Mateos, M.A. et al., 2015. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nature Methods*, 12(10), pp.982–988. Available at: <http://www.nature.com/doi/10.1038/nmeth.3543> [Accessed October 18, 2017].
- Morgan, J.E. et al., 1993. Long-term persistence and migration of myogenic cells injected into pre-irradiated muscles of mdx mice. *Journal of the Neurological Sciences*, 115(2), pp.191–200. Available at: <http://www.sciencedirect.com/science/article/pii/0022510X9390224M> [Accessed November 17, 2017].
- Morgan, J.E. et al., 1994. Myogenic Cell Lines Derived from Transgenic Mice Carrying a Thermolabile T Antigen: A Model System for the Derivation of Tissue-Specific and Mutation-Specific Cell Lines. *Developmental Biology*, 162(2), pp.486–498. Available at: <http://www.sciencedirect.com/science/article/pii/S0012160684711031?via=ihub> [Accessed November 11, 2017].
- Morgan, J.E., Coulton, G.R. & Partridge, T.A., 1987. Muscle precursor cells invade and repopulate freeze-killed muscles. *Journal of muscle research and cell motility*, 8(5), pp.386–96. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3480896> [Accessed November 17, 2017].
- Morgan, J.E. & Zammit, P.S., 2010. Direct effects of the pathogenic mutation on satellite cell function in muscular dystrophy. *Experimental Cell Research*, 316, pp.3100–3108. Available at: https://ac.els-cdn.com/S0014482710002752/1-s2.0-S0014482710002752-main.pdf?_tid=ede1793e-d492-11e7-8c34-

00000aab0f6c&acdnat=1511911498_20cff737497fab566238d717ede00e3f [Accessed November 28, 2017].

- Moscou, M.J. & Bogdanove, A.J., 2009. A simple cipher governs DNA recognition by TAL effectors. *Science (New York, N.Y.)*, 326(5959), p.1501. Available at: <http://www.sciencemag.org/cgi/doi/10.1126/science.1178817> [Accessed November 27, 2017].
- Moure, C.M., Gimble, F.S. & Quioco, F.A., 2003. The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *Journal of molecular biology*, 334(4), pp.685–95. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14636596> [Accessed November 26, 2017].
- Moynahan, M.E., Pierce, A.J. & Jasin, M., 2001. BRCA2 is required for homology-directed repair of chromosomal breaks. *Molecular Cell*, 7(2), pp.263–272. Available at: <http://www.sciencedirect.com/science/article/pii/S1097276501001745> [Accessed November 23, 2017].
- Mulsant, P. et al., 1988. Phleomycin resistance as a dominant selectable marker in CHO cells. *Somatic cell and molecular genetics*, 14(3), pp.243–52. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2453083> [Accessed July 29, 2017].
- Muntoni, F. et al., 1993. Deletion of the Dystrophin Muscle-Promoter Region Associated with X-Linked Dilated Cardiomyopathy. *New England Journal of Medicine*, 329(13), pp.921–925. Available at: <http://www.nejm.org/doi/abs/10.1056/NEJM199309233291304> [Accessed October 17, 2017].
- Murphy, R.M. et al., 2013. Ca²⁺-dependent proteolysis of junctophilin-1 and junctophilin-2 in skeletal and cardiac muscle. *The Journal of Physiology*, 591(3), pp.719–729. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23148318> [Accessed November 20, 2017].
- Mussolino, C. et al., 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research*, 39(21), pp.9283–93. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3241638&tool=pmcentrez&rendertype=abstract> [Accessed January 23, 2014].
- Musunuru, K., 2017. The hope and hype of CRISPR-Cas9 genome editing: A review. *JAMA Cardiology*, 2(8), pp.914–919. Available at: <http://cardiology.jamanetwork.com/article.aspx?doi=10.1001/jamacardio.2017.1713> [Accessed November 23, 2017].
- Nakamae, K. et al., 2017. Establishment of expanded and streamlined pipeline of PITCh knock-in – a web-based design tool for MMEJ-mediated gene knock-in, PITCh designer, and the variations of PITCh, PITCh-TG and PITCh-KIKO. *Bioengineered*, 8(3), pp.1–7. Available at: <https://www.tandfonline.com/doi/full/10.1080/21655979.2017.1313645>.
- Nakamura, A. & Takeda, S., 2011. Mammalian models of Duchenne Muscular Dystrophy: pathological characteristics and therapeutic applications. *Journal of biomedicine & biotechnology*, 2011, p.184393. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21274260> [Accessed October 4, 2017].
- Nakamura, K. et al., 2014. Generation of muscular dystrophy model rats with a CRISPR/Cas system. *Scientific reports*, 4, p.5635. Available at: <http://www.nature.com/srep/2014/140709/srep05635/full/srep05635.html> [Accessed September 26, 2015].

- Nayerossadat, N., Maedeh, T. & Ali, P.A., 2012. Viral and nonviral delivery systems for gene delivery. *Advanced biomedical research*, 1, p.27. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23210086> [Accessed December 16, 2017].
- NCBI Homologene, 2015. HomoloGene - NCBI. Available at: http://www.ncbi.nlm.nih.gov/homologene?cmd=Retrieve&dopt=AlignmentScores&list_ids=20856 [Accessed June 24, 2015].
- Negrone, E. et al., 2016. Cellular Therapies for Muscular Dystrophies: Frustrations and Clinical Successes. *Human gene therapy*, 27(2), pp.117–26. Available at: <http://online.liebertpub.com/doi/10.1089/hum.2015.139> [Accessed November 17, 2017].
- Nelson, C.E. et al., 2016. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science (New York, N.Y.)*, 351(6271), pp.403–7. Available at: <http://www.sciencemag.org/cgi/doi/10.1126/science.aad5143> [Accessed November 16, 2017].
- Neri, M. et al., 2007. Dystrophin levels as low as 30% are sufficient to avoid muscular dystrophy in the human. *Neuromuscular Disorders*, 17(11–12), pp.913–918.
- NICE, 2016. Evaluation consultation document – ataluren for treating Duchenne muscular dystrophy with a nonsense mutation in the dystrophin gene Ataluren for treating Duchenne muscular dystrophy with a nonsense mutation in the dystrophin gene. Available at: <https://www.nice.org.uk/guidance/hst3/documents/evaluation-consultation-document> [Accessed November 7, 2017].
- Nielsen, S., Yuzenkova, Y. & Zenkin, N., 2013. Mechanism of Eukaryotic RNA Polymerase III Transcription Termination. *Science*, 340(6140), pp.1577–1580. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23812715> [Accessed July 14, 2017].
- Nishimasu, H. et al., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156(5), pp.935–949. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24529477> [Accessed November 27, 2017].
- Nishio, H. et al., 1994. Identification of a novel first exon in the human dystrophin gene and of a new promoter located more than 500 kb upstream of the nearest known promoter. *The Journal of clinical investigation*, 94(3), pp.1037–42. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=295157&tool=pmcentrez&endertype=abstract> [Accessed July 10, 2015].
- Nudel, U., 2005. Alternative Promoters : Duchenne Muscular Dystrophy (DMD) Gene. *eLS*, (Dmd), pp.1–5. Available at: <http://doi.wiley.com/10.1038/npg.els.0005026> [Accessed August 6, 2017].
- Nudel, U. et al., 1989. Duchenne muscular dystrophy gene product is not identical in muscle and brain. *Nature*, 337(6202), pp.76–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2909892> [Accessed July 9, 2015].
- Optimised CRISPR Design, 2013. Optimized CRISPR Design. Available at: <http://crispr.mit.edu/> [Accessed July 14, 2015].
- Orlowski, J., Boniecki, M. & Bujnicki, J.M., 2007. I-Ssp6803I: the first homing endonuclease from the PD-(D/E)XK superfamily exhibits an unusual mode of DNA recognition. *Bioinformatics*, 23(5), pp.527–530. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17242028> [Accessed November 26, 2017].

- Ousterout, D.G., Kabadi, A.M., Thakore, P.I., Perez-Pinera, P., et al., 2015. Correction of dystrophin expression in cells from duchenne muscular dystrophy patients through genomic excision of exon 51 by zinc finger nucleases. *Molecular therapy : the journal of the American Society of Gene Therapy*, 23(3), pp.523–32. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4351462&tool=pmcentrez&rendertype=abstract> [Accessed September 6, 2015].
- Ousterout, D.G., Kabadi, A.M., Thakore, P.I., Majoros, W.H., et al., 2015. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nature communications*, 6, p.6244. Available at: <http://www.nature.com/ncomms/2015/150218/ncomms7244/full/ncomms7244.html> [Accessed February 19, 2015].
- Ousterout, D.G. et al., 2013a. Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. *Molecular therapy : the journal of the American Society of Gene Therapy*, 21(9), pp.1718–26. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3776627&tool=pmcentrez&rendertype=abstract> [Accessed September 5, 2015].
- Ousterout, D.G. et al., 2013b. Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. *Molecular therapy : the journal of the American Society of Gene Therapy*, 21(9), pp.1718–26. Available at: [/pmcc/articles/PMC3776627/?report=abstract](http://pmcc/articles/PMC3776627/?report=abstract) [Accessed February 1, 2014].
- van Overbeek, M. et al., 2016. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Molecular Cell*, 63(4), pp.633–646. Available at: <http://www.sciencedirect.com/science/article/pii/S1097276516303252> [Accessed November 27, 2017].
- Partridge, T., 1978. Evidence of fusion between host and donor myoblasts in skeletal muscle grafts. *Nature*, 273(5660), pp.306–308. Available at: <http://www.nature.com/doi/10.1038/273306a0> [Accessed November 17, 2017].
- Partridge, T.A. et al., 1989. Conversion of mdx myofibres from dystrophin-negative to -positive by injection of normal myoblasts. *Nature*, 337(6203), pp.176–179. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2643055> [Accessed November 17, 2017].
- Pattanayak, V. et al., 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology*, 31(9), pp.839–43. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23934178> [Accessed January 22, 2014].
- Paull, T.T., 2015. Mechanisms of ATM Activation. *Annual Review of Biochemistry*, 84(1), pp.711–738. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25580527> [Accessed November 22, 2017].
- Pavletich, N.P. & Pabo, C.O., 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science (New York, N.Y.)*, 252(5007), pp.809–17. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2028256> [Accessed November 26, 2017].
- Périé, S. et al., 2014. Autologous Myoblast Transplantation for Oculopharyngeal Muscular Dystrophy: a Phase I/IIa Clinical Study. *Molecular Therapy*, 22(1), pp.219–225. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23831596> [Accessed November 17, 2017].
- Phillips, M.F. & Quinlivan, R., 2008. Calcium antagonists for Duchenne muscular dystrophy. *The Cochrane database of systematic reviews*, (4), p.CD004571. Available at:

<http://www.ncbi.nlm.nih.gov/pubmed/18843663> [Accessed November 28, 2017].

- Philpott, N.J. et al., 2002. Efficient integration of recombinant adeno-associated virus DNA vectors requires a p5-rep sequence in cis. *Journal of virology*, 76(11), pp.5411–21. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11991970> [Accessed November 17, 2017].
- Piccolo, P. et al., 2014. SR-A and SREC-I binding peptides increase HDAd-mediated liver transduction. *Gene therapy*, 21(11), pp.950–7. Available at: <http://www.nature.com/doi/10.1038/gt.2014.71> [Accessed November 17, 2017].
- Pichavant, C. et al., 2010. Expression of dog microdystrophin in mouse and dog muscles by gene therapy. *Molecular therapy : the journal of the American Society of Gene Therapy*, 18(5), pp.1002–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2890099&tool=pmcentrez&rendertype=abstract> [Accessed July 14, 2015].
- Pini, V. et al., 2017. Genome Editing and Muscle Stem Cells as a Therapeutic Tool for Muscular Dystrophies. *Current Stem Cell Reports*, 3(2), pp.137–148. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28616376> [Accessed November 17, 2017].
- Plotkin, J.B., Robins, H. & Levine, A.J., 2004. Tissue-specific codon usage and the expression of human genes. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34), pp.12588–91. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=515101&tool=pmcentrez&rendertype=abstract> [Accessed May 29, 2015].
- Popplewell, L. et al., 2013. Gene correction of a duchenne muscular dystrophy mutation by meganuclease-enhanced exon knock-in. *Human gene therapy*, 24(7), pp.692–701. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23790397> [Accessed February 11, 2014].
- Prakash, V., Moore, M. & Yáñez-Muñoz, R.J., 2016. Current Progress in Therapeutic Gene Editing for Monogenic Diseases. *Molecular therapy : the journal of the American Society of Gene Therapy*, 24(3), pp.465–74. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26765770> [Accessed July 2, 2017].
- Di Primio, C. et al., 2005. Potentiation of gene targeting in human cells by expression of *Saccharomyces cerevisiae* Rad52. *Nucleic Acids Research*, 33(14), pp.4639–4648. Available at: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gki778> [Accessed June 24, 2017].
- Qiu, P. et al., 2004. Mutation detection using Surveyor nuclease. *BioTechniques*, 36(4), pp.702–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15088388> [Accessed May 30, 2015].
- Raab, D. et al., 2010. The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. *Systems and synthetic biology*, 4(3), pp.215–25. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2955205&tool=pmcentrez&rendertype=abstract> [Accessed July 14, 2015].
- Rahdar, M. et al., 2015. Synthetic CRISPR RNA-Cas9-guided genome editing in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, 112(51), pp.E7110–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26589814> [Accessed July 14, 2017].

- Rajagopal, N. et al., 2016. High-throughput mapping of regulatory DNA. *Nature Biotechnology*, 34(2), pp.167–174. Available at: <http://www.nature.com/doifinder/10.1038/nbt.3468> [Accessed November 5, 2017].
- Ramirez, C.L. et al., 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature Methods*, 5(5), pp.374–375. Available at: <http://www.nature.com/doifinder/10.1038/nmeth0508-374> [Accessed November 26, 2017].
- Ran, F. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11), pp.2281–308. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24157548> <http://www.nature.com/nprot/journal/v8/n11/abs/nprot.2013.143.html>.
- Ran, F.A., Hsu, P.D., Lin, C.-Y., et al., 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154(6), pp.1380–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3856256&tool=pmcentrez&rendertype=abstract> [Accessed July 10, 2014].
- Ran, F.A., Hsu, P.D., Wright, J., et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), pp.2281–2308. Available at: <http://www.nature.com/doifinder/10.1038/nprot.2013.143> [Accessed November 27, 2017].
- Ran, F.A. et al., 2015. In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, 520(7546), pp.186–91. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4393360&tool=pmcentrez&rendertype=abstract> [Accessed April 1, 2015].
- Rando, T.A. et al., 1998. Muscle cells from mdx mice have an increased susceptibility to oxidative stress. *Neuromuscular disorders : NMD*, 8(1), pp.14–21. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9565986> [Accessed November 28, 2017].
- Ranzani, M. et al., 2013. Lentiviral vector-based insertional mutagenesis identifies genes associated with liver cancer. *Nature methods*, 10(2), pp.155–61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23314173> [Accessed July 29, 2017].
- Regalado, A., 2016. The World's Most Expensive Medicine Is a Bust. *MIT Technology Review*, pp.1–11. Available at: <https://www.technologyreview.com/s/601165/the-worlds-most-expensive-medicine-is-a-bust/> [Accessed November 25, 2017].
- Relizani, K. et al., 2017. Efficacy and Safety Profile of Tricyclo-DNA Antisense Oligonucleotides in Duchenne Muscular Dystrophy Mouse Model. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5498286/pdf/main.pdf> [Accessed November 24, 2017].
- Remy, S. et al., 2014. Efficient gene targeting by homology-directed repair in rat zygotes using TALE nucleases. *Genome research*, 24(8), pp.1371–83. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24989021> [Accessed July 14, 2017].
- Renjini, R. et al., 2012. Oxidative Damage in Muscular Dystrophy Correlates with the Severity of the Pathology: Role of Glutathione Metabolism. *Neurochemical Research*, 37(4), pp.885–898. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22219131> [Accessed November 28, 2017].
- Reyon, D. et al., 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nature*

- Biotechnology*, 30(5), pp.460–465. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/22484455> [Accessed November 27, 2017].
- Reza, M. et al., 2016. Optimization of Internally Deleted Dystrophin Constructs. *Human gene therapy methods*, 27(5), pp.174–186. Available at:
<http://online.liebertpub.com/doi/10.1089/hgtb.2016.026> [Accessed September 8, 2017].
- Ricotti, V. et al., 2016. Safety , Tolerability , and Pharmacokinetics of SMT C1100 , a 2-Arylbenzoxazole Utrophin Modulator , following Single- and Multiple- Dose Administration to Pediatric Patients with Duchenne Muscular Dystrophy. Available at:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4824384/pdf/pone.0152840.pdf>
 [Accessed November 8, 2017].
- te Riele, H., Maandag, E.R. & Berns, A., 1992. Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proceedings of the National Academy of Sciences of the United States of America*, 89(11), pp.5128–32. Available at:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=49242&tool=pmcentrez&rendertype=abstract> [Accessed February 13, 2014].
- Roberts, D.M. et al., 2006. Hexon-chimaeric adenovirus serotype 5 vectors circumvent pre-existing anti-vector immunity. *Nature*, 441(7090), pp.239–243. Available at:
<http://www.nature.com/doi/10.1038/nature04721> [Accessed November 17, 2017].
- Rodino-Klapac, L.R., Mendell, J.R. & Sahenk, Z., 2013. Update on the Treatment of Duchenne Muscular Dystrophy. *Current Neurology and Neuroscience Reports*, 13(3), p.332. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23328943> [Accessed November 25, 2017].
- Römer, P. et al., 2007. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science (New York, N.Y.)*, 318(5850), pp.645–8. Available at:
<http://www.sciencemag.org/content/318/5850/645.abstract> [Accessed February 11, 2014].
- Romero, N.B. et al., 2004. Phase I Study of Dystrophin Plasmid-Based Gene Therapy in Duchenne/Becker Muscular Dystrophy. *Human Gene Therapy*, 15(11), pp.1065–1076. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15610607> [Accessed December 12, 2017].
- Rose, J.A. & Koczot, F., 1972. Adenovirus-Associated Virus Multiplication VII. Helper Requirement for Viral Deoxyribonucleic Acid and Ribonucleic Acid Synthesis. *JOURNAL OF VIROLOGY*, pp.1–8. Available at: <http://jvi.asm.org/content/10/1/1.full.pdf> [Accessed December 16, 2017].
- Sakuma, T. et al., 2015. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nature Protocols*, 11(1), pp.118–133. Available at:
<http://www.nature.com/doi/10.1038/nprot.2015.140> [Accessed September 13, 2017].
- Sampaolesi, M. et al., 2003. Cell Therapy of α -Sarcoglycan Null Dystrophic Mice Through Intra-Arterial Delivery of Mesoangioblasts. *Science*, 301(5632), pp.487–492. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/12855815> [Accessed November 18, 2017].
- Sampaolesi, M. et al., 2006. Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs. *Nature*, 444(7119), pp.574–579. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/17108972> [Accessed November 18, 2017].

- San Filippo, J. et al., 2008. Mechanism of Eukaryotic Homologous Recombination. *Annual Review of Biochemistry*, 77(1), pp.229–257. Available at: <http://www.annualreviews.org/doi/10.1146/annurev.biochem.77.061306.125255> [Accessed November 23, 2017].
- Sander, J.D. et al., 2011. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature Methods*, 8(1), pp.67–69. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21151135> [Accessed November 26, 2017].
- Schmid-Burgk, J.L. et al., 2012. A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes. *Nature Biotechnology*, 31(1), pp.76–81. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23242165> [Accessed November 27, 2017].
- Schuelke, M. et al., 2004. Myostatin Mutation Associated with Gross Muscle Hypertrophy in a Child. *New England Journal of Medicine*, 350(26), pp.2682–2688. Available at: <http://www.nejm.org/doi/abs/10.1056/NEJMoa040933> [Accessed November 9, 2017].
- Schwartz, F. et al., 1991. A dominant positive and negative selectable gene for use in mammalian cells (gene fusion/herpes simplex virus thymidine kinase/bacterial neomycin phosphotransferase). *Genetics*, 88, pp.10416–10420. Available at: <http://www.pnas.org/content/88/23/10416.full.pdf> [Accessed June 28, 2017].
- Seth, P. et al., 2008. Identification of an intronic splicing enhancer essential for the inclusion of FGFR2 exon IIIc. *The Journal of biological chemistry*, 283(15), pp.10058–67. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18256031> [Accessed July 29, 2017].
- sgRNA Designer, 2017. sgRNA Designer: CRISPRko. Available at: <https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design> [Accessed November 6, 2017].
- Shin, J. et al., 2013. Wasting mechanisms in muscular dystrophy. *The international journal of biochemistry & cell biology*, 45(10), pp.2266–79. Available at: <http://www.sciencedirect.com/science/article/pii/S1357272513001386> [Accessed February 11, 2014].
- Sicinski, P. et al., 1989. The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. *Science*, 244(4912), pp.1578–1580.
- Singh, P. et al., 2016. Translational Pharmacokinetic/Pharmacodynamic Analysis of MYO-029 Antibody for Muscular Dystrophy. *Clinical and translational science*, 9(6), pp.302–310. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27700008> [Accessed November 29, 2017].
- Skuk, D. et al., 2006. Dystrophin Expression in Muscles of Duchenne Muscular Dystrophy Patients After High-Density Injections of Normal Myogenic Cells. *Journal of Neuropathology and Experimental Neurology*, 65(4), pp.371–386. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16691118> [Accessed November 17, 2017].
- Skuk, D. et al., 2002. Efficacy of Myoblast Transplantation in Nonhuman Primates Following Simple Intramuscular Cell Injections: Toward Defining Strategies Applicable to Humans. *Experimental Neurology*, 175(1), pp.112–126. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12009764> [Accessed November 18, 2017].
- Skuk, D. et al., 2007. First test of a “high-density injection” protocol for myogenic cell transplantation throughout large volumes of muscles in a Duchenne muscular dystrophy

- patient: eighteen months follow-up. *Neuromuscular Disorders*, 17(1), pp.38–46. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17142039> [Accessed November 17, 2017].
- Slaymaker, I.M. et al., 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science*, 351(6268).
- Song, F. & Stieger, K., 2017. Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. *Molecular Therapy - Nucleic Acids*, 7, pp.53–60. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S2162253117301336> [Accessed July 29, 2017].
- Spinazzola, J.M. & Kunkel, L.M., 2016. Pharmacological therapeutics targeting the secondary defects and downstream pathology of Duchenne muscular dystrophy. *Expert opinion on orphan drugs*, 4(11), pp.1179–1194. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28670506> [Accessed November 29, 2017].
- Stern, L.Z. et al., 1982. Drug trial of superoxide dismutase in Duchenne's muscular dystrophy. *Archives of neurology*, 39(6), pp.342–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7046702> [Accessed November 28, 2017].
- Stoddard, B.L., 2014. Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mobile DNA*, 5(1), p.7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24589358> [Accessed November 26, 2017].
- Streubel, J. et al., 2012. TAL effector RVD specificities and efficiencies. *Nature biotechnology*, 30(7), pp.593–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22781676> [Accessed January 29, 2014].
- Sugio, A. et al., 2007. Two type III effector genes of *Xanthomonas oryzae* pv. *oryzae* control the induction of the host genes *OstFIIA 1* and *OstTFX1* during bacterial blight of rice. *Proceedings of the National Academy of Sciences*, 104(25), pp.10720–10725. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17563377> [Accessed November 27, 2017].
- Summit Therapeutics Plc, 2016. FE Investegate | Summit Therapeutics plc Announcements | Summit Therapeutics plc: Summit Reports Positive Interim Data from Phase... Available at: <https://www.investegate.co.uk/summit-therapeutics-plc--summ--/gnw/summit-reports-positive-interim-data-from-phase---/20160330120004H8564/> [Accessed November 29, 2017].
- Sun, W. et al., 2015. Self-assembled DNA nanoclews for the efficient delivery of CRISPR-Cas9 for genome editing. *Angewandte Chemie (International ed. in English)*, 54(41), pp.12029–33. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26310292> [Accessed November 17, 2017].
- Suzuki, K. et al., 2016. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, 540(7631), pp.144–149. Available at: <http://www.nature.com/doifinder/10.1038/nature20565>.
- Szcepek, M. et al., 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature biotechnology*, 25(7), pp.786–93. Available at: <http://www.nature.com/doifinder/10.1038/nbt1317> [Accessed November 26, 2017].
- Tabebordbar, M. et al., 2016. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science*, 351(6271), pp.407–411. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26721686> [Accessed November 16, 2017].
- Taglia, A. et al., 2015. Clinical features of patients with dystrophinopathy sharing the 45-55

- exon deletion of DMD gene. *Acta myologica : myopathies and cardiomyopathies : official journal of the Mediterranean Society of Myology*, 34(1), pp.9–13. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26155064> [Accessed December 12, 2017].
- Takahashi, K. et al., 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, 131(5), pp.861–872. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0092867407014717> [Accessed November 18, 2017].
- Tebas, P. et al., 2014. Gene Editing of *CCR5* in Autologous CD4 T Cells of Persons Infected with HIV. *New England Journal of Medicine*, 370(10), pp.901–910. Available at: <http://www.nejm.org/doi/10.1056/NEJMoa1300662> [Accessed November 25, 2017].
- Tedesco, F.S. et al., 2012. Transplantation of Genetically Corrected Human iPSC-Derived Progenitors in Mice with Limb-Girdle Muscular Dystrophy. *Science Translational Medicine*, 4(140), p.140ra89-140ra89. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22745439> [Accessed November 18, 2017].
- Teigler, J.E., Kagan, J.C. & Barouch, D.H., 2014. Late endosomal trafficking of alternative serotype adenovirus vaccine vectors augments antiviral innate immunity. *Journal of virology*, 88(18), pp.10354–63. Available at: <http://jvi.asm.org/cgi/doi/10.1128/JVI.00936-14> [Accessed November 17, 2017].
- Thierry, A. & Dujon, B., 1992. Nested chromosomal fragmentation in yeast using the meganuclease I-Sce I: a new method for physical mapping of eukaryotic genomes. *Nucleic acids research*, 20(21), pp.5625–31. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1333585> [Accessed November 26, 2017].
- Thomas, M. et al., 2000. Myostatin, a Negative Regulator of Muscle Growth, Functions by Inhibiting Myoblast Proliferation. *Journal of Biological Chemistry*, 275(51), pp.40235–40243. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10976104> [Accessed November 29, 2017].
- Tidball, J.G. & Spencer, M.J., 2000. Calpains and muscular dystrophies. *International Journal of Biochemistry and Cell Biology*, 32(1), pp.1–5. Available at: <https://www.sciencedirect.com/science/article/pii/S1357272599000953?via%3Dihub> [Accessed November 20, 2017].
- Tinsley, J. et al., 1998. Expression of full-length utrophin prevents muscular dystrophy in mdx mice. *Nature Medicine*, 4(12), pp.1441–1444. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9846586> [Accessed November 29, 2017].
- Tinsley, J., Robinson, N. & Davies, K.E., 2015. Safety, tolerability, and pharmacokinetics of SMT C1100, a 2-arylbenzoxazole utrophin modulator, following single- and multiple-dose administration to healthy male adult volunteers. *Journal of clinical pharmacology*, 55(6), pp.698–707. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25651188> [Accessed November 9, 2017].
- Tinsley, J.M. et al., 2011. Daily treatment with SMT C1100, a novel small molecule utrophin upregulator, dramatically reduces the dystrophic symptoms in the mdx mouse P. Dent, ed. *PLoS ONE*, 6(5), p.e19189. Available at: <http://dx.plos.org/10.1371/journal.pone.0019189> [Accessed November 8, 2017].
- Torrente, Y. et al., 2004. Human circulating AC133 + stem cells restore dystrophin expression and ameliorate function in dystrophic skeletal muscle. , 114(2).

- Transgenomic, I., 2010. User Guide for the Transgenomic SURVEYOR® Mutation Detection Kit for Standard Gel Electrophoresis. , (482106), pp.1–34. Available at: <http://www.idtdna.com/pages/docs/default-source/user-guides-and-protocols/userguide-surveyor-standard.pdf> [Accessed December 10, 2017].
- Traynor, K., 2017. Deflazacort approved for Duchenne muscular dystrophy. *American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists*, 74(6), p.368. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28274975> [Accessed November 28, 2017].
- Tsai, S.Q. et al., 2014. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology*, 33(2), pp.187–197.
- Tupler, R., Perini, G. & Green, M.R., 2001. Expressing the human genome. *Nature*, 409(6822), pp.832–833. Available at: <http://www.nature.com/doi/10.1038/35057011> [Accessed November 26, 2017].
- Turner, P. et al., 1993. Proteolysis results in altered leak channel kinetics and elevated free calcium in mdx muscle. *The Journal of Membrane Biology*, 133(3), pp.243–251. Available at: <http://link.springer.com/10.1007/BF00232023> [Accessed November 20, 2017].
- Turner, P.R. et al., 1991. Increased calcium influx in dystrophic muscle. *The Journal of cell biology*, 115(6), pp.1701–12. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1661733> [Accessed November 20, 2017].
- Tycko, J., Myer, V.E. & Hsu, P.D., 2016. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Molecular Cell*, 63(3), pp.355–370. Available at: <http://www.sciencedirect.com/science/article/pii/S1097276516303318?via%3Dihub> [Accessed October 18, 2017].
- Vallese, D. et al., 2013. The Rag2^{-/-}Il2rb^{-/-}Dmd^{-/-} mouse: a novel dystrophic and immunodeficient model to assess innovating therapeutic strategies for muscular dystrophies. *Molecular therapy : the journal of the American Society of Gene Therapy*, 21(10), pp.1950–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23975040> [Accessed October 4, 2017].
- Velic, D. et al., 2015. DNA damage signalling and repair inhibitors: The long-sought-after achilles heel of cancer. *Biomolecules*, 5(4), pp.3204–3259.
- Verhaart, I.E.C. & Aartsma-Rus, A., 2012. Verhaart I, Aartsma-Rus A. AON-Mediated Exon Skipping for Duchenne Muscular Dystrophy. AON-Mediated Exon Skipping for Duchenne Muscular Dystrophy. *Neuromuscular Disorders. Intech*, pp.55–80. Available at: <http://dx.doi.org/10.5772/33938> [Accessed November 24, 2017].
- Villalta, S.A., Deng, B., et al., 2011. IFN-γ Promotes Muscle Damage in the mdx Mouse Model of Duchenne Muscular Dystrophy by Suppressing M2 Macrophage Activation and Inhibiting Muscle Cell Proliferation. *The Journal of Immunology*, 187(10), pp.5419–5428. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22013114> [Accessed November 24, 2017].
- Villalta, S.A., Rinaldi, C., et al., 2011. Interleukin-10 reduces the pathology of mdx muscular dystrophy by deactivating M1 macrophages and modulating macrophage phenotype. *Human Molecular Genetics*, 20(4), pp.790–805. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21118895> [Accessed November 24, 2017].
- Voit, T. et al., 2014. Safety and efficacy of drisapersen for the treatment of Duchenne muscular dystrophy (DEMAND II): an exploratory, randomised, placebo-controlled phase 2 study.

- The Lancet. Neurology*, 13(10), pp.987–96. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/25209738> [Accessed November 19, 2015].
- Vossen, R.H.A.M. et al., 2009. High-resolution melting analysis (HRMA): more than just sequence variant screening. *Human mutation*, 30(6), pp.860–6. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/19418555> [Accessed May 29, 2015].
- Vouillot, L., Thélie, A. & Pollet, N., 2015. Comparison of T7E1 and Surveyor Mismatch Cleavage Assays To Detect Mutations Triggered by Engineered Nucleases. *G3 (Bethesda, Md.)*, 5(3), pp.407–15. Available at:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4349094&tool=pmcentrez&rendertype=abstract> [Accessed February 3, 2015].
- Wagner, K.R. et al., 2008. A Phase I/II trial of MYO-029 in Adult Subjects with Muscular Dystrophy. *Ann Neurol*, 63, pp.561–571. Available at:
https://s3.amazonaws.com/objects.readcube.com/articles/downloaded/wiley/f135fcc1fe099a8128fb8bbdfbd1335b2db7651f54c233357c3ddc133e470b02.pdf?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIS5LBPCM5JPOCDGQ%2F20171109%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20171109T192244Z&X-Amz-Expires=103035&X-Amz-SignedHeaders=host&X-Amz-Signature=aebf406a9ee6debe1b318967f52868aefec821bb655dbe84fe6dcf4b1f7bf3b7 [Accessed November 9, 2017].
- Wagner, K.R. et al., 2008. A phase I/II trial of MYO-029 in adult subjects with muscular dystrophy. *Annals of Neurology*, 63(5), pp.561–571. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/18335515> [Accessed November 9, 2017].
- Wagner, K.R. et al., 2002. Loss of myostatin attenuates severity of muscular dystrophy in mdx mice. *Annals of Neurology*, 52(6), pp.832–836. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/12447939> [Accessed November 29, 2017].
- Wang, B., Li, J. & Xiao, X., 2000. Adeno-associated virus vector carrying human minidystrophin genes effectively ameliorates muscular dystrophy in mdx mouse model. *Proceedings of the National Academy of Sciences*, 97(25), pp.13714–13719. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/11095710> [Accessed September 8, 2017].
- Wang, L. et al., 2017. Enhancing Targeted Genomic DNA Editing in Chicken Cells Using the CRISPR/Cas9 System W. Shen, ed. *PLOS ONE*, 12(1), p.e0169768. Available at:
<http://dx.plos.org/10.1371/journal.pone.0169768> [Accessed June 24, 2017].
- Waterston, R.H. et al., 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), pp.520–62. Available at:
<http://dx.doi.org/10.1038/nature01262> [Accessed January 2, 2015].
- Wein, N. et al., 2014. Translation from a DMD exon 5 IRES results in a functional dystrophin isoform that attenuates dystrophinopathy in humans and mice. *Nature Medicine*, 20(9), pp.992–1000. Available at: <http://www.nature.com/doifinder/10.1038/nm.3628>.
- Welch, E.M. et al., 2007. PTC124 targets genetic disorders caused by nonsense mutations. *Nature*, 447(7140), pp.87–91. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/17450125> [Accessed February 5, 2014].
- West, S.C., 2003. Molecular views of recombination proteins and their control. *Nature Reviews Molecular Cell Biology*, 4(6), pp.435–445. Available at:
<http://www.nature.com/doifinder/10.1038/nrm1127> [Accessed November 23, 2017].

- Weterings, E. & Chen, D.J., 2008. The endless tale of non-homologous end-joining. *Cell Research*, 18(1), pp.114–124. Available at: <http://www.nature.com/doifinder/10.1038/cr.2008.3> [Accessed November 22, 2017].
- Whewy, J.M. & Roberts, R.G., 2003. The dystrophin lymphocyte promoter revisited: 4.5-megabase intron, or artifact? *Neuromuscular disorders : NMD*, 13(1), pp.17–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12467728> [Accessed July 14, 2015].
- White, M.F. et al., 1997. Review Article: Recognition and Manipulation of Branched DNA Structure by Junction-resolving Enzymes. Available at: https://ac.els-cdn.com/S0022283697910974/1-s2.0-S0022283697910974-main.pdf?_tid=988ae118-da1a-11e7-8920-00000aab0f01&acdnat=1512519531_9479c0ecd91ace46f18090d529a4fa48 [Accessed December 6, 2017].
- Whitehead, N.P. et al., 2010. Skeletal Muscle NADPH Oxidase Is Increased and Triggers Stretch-Induced Damage in the mdx Mouse M. A. Tarnopolsky, ed. *PLoS ONE*, 5(12), p.e15354. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21187957> [Accessed November 28, 2017].
- Whitehead, N.P. et al., 2006. Streptomycin reduces stretch-induced membrane permeability in muscles from mdx mice. *Neuromuscular Disorders*, 16(12), pp.845–854. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17005404> [Accessed November 20, 2017].
- Whitmore, C. & Morgan, J., 2014. What do mouse models of muscular dystrophy tell us about the DAPC and its components? *International Journal of Experimental Pathology*, 95(6), pp.365–377. Available at: <http://doi.wiley.com/10.1111/iep.12095>.
- Williams, R.S., Williams, J.S. & Tainer, J.A., 2007. Mre11–Rad50–Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin template This paper is one of a selection of papers published in this Special Issue, entitled 28th International West Coast Chromatin and Chromosome Conference, and has undergone the Journal’s usual peer review process. *Biochemistry and Cell Biology*, 85(4), pp.509–520. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17713585> [Accessed November 22, 2017].
- Wilton, S.D. et al., 2007. Antisense Oligonucleotide-induced Exon Skipping Across the Human Dystrophin Gene Transcript. *Molecular Therapy*, 15(7), pp.1288–1296. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17285139> [Accessed November 24, 2017].
- Wong, N., Liu, W. & Wang, X., 2015. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome biology*, 16, p.218. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26521937> [Accessed November 6, 2017].
- Wright, A. V et al., 2015. Rational design of a split-Cas9 enzyme complex. *Proceedings of the National Academy of Sciences of the United States of America*, 112(10), pp.2984–9. Available at: <http://www.pnas.org/content/112/10/2984.abstract>.
- Wright, J.B. & Sanjana, N.E., 2016. CRISPR Screens to Discover Functional Noncoding Elements. *Trends in Genetics*, 32(9), pp.526–529. Available at: <http://dx.doi.org/10.1016/j.tig.2016.06.004>.
- Wu, Z., Yang, H. & Colosi, P., 2010a. Effect of genome size on AAV vector packaging. *Molecular therapy : the journal of the American Society of Gene Therapy*, 18(1), pp.80–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19904234> [Accessed November 16, 2017].

- Wu, Z., Yang, H. & Colosi, P., 2010b. Effect of genome size on AAV vector packaging. *Molecular therapy : the journal of the American Society of Gene Therapy*, 18(1), pp.80–6. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2839202&tool=pmcentrez&rendertype=abstract> [Accessed February 6, 2014].
- Wynn, T.A., 2008. Cellular and molecular mechanisms of fibrosis. *Journal of Pathology*, 214(2), pp.199–210. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18161745> [Accessed November 24, 2017].
- Xiang, G. et al., 2017. Temperature effect on CRISPR-Cas9 mediated genome editing. *Journal of Genetics and Genomics*, 44(4), pp.199–205. Available at: <http://dx.doi.org/10.1016/j.jgg.2017.03.004>.
- Xu, H. et al., 2015. Sequence determinants of improved CRISPR sgRNA design. *Genome research*, 25(8), pp.1147–57. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26063738> [Accessed November 6, 2017].
- Xu, L. et al., 2016. CRISPR-mediated Genome Editing Restores Dystrophin Expression and Function in mdx Mice. *Molecular Therapy*, 24(3), pp.564–569. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26449883> [Accessed November 16, 2017].
- Ylä-Herttua, S., 2012. Endgame: glybera finally recommended for approval as the first gene therapy drug in the European union. *Molecular therapy : the journal of the American Society of Gene Therapy*, 20(10), pp.1831–2. Available at: <http://dx.doi.org/10.1038/mt.2012.194> [Accessed November 20, 2015].
- Yoshimi, K. et al., 2016. ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nature Communications*, 7, p.10431. Available at: <http://www.nature.com/doi/10.1038/ncomms10431>.
- Yu, X. et al., 2016. Improved delivery of Cas9 protein/gRNA complexes using lipofectamine CRISPRMAX. *Biotechnology letters*, 38(6), pp.919–29. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26892225> [Accessed June 13, 2017].
- Yue, Y. et al., 2016. 100-fold but not 50-fold dystrophin overexpression aggravates electrocardiographic defects in the mdx model of Duchenne muscular dystrophy. *Molecular therapy. Methods & clinical development*, 3, p.16045. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27419194> [Accessed July 29, 2017].
- Yue, Y., Liu, M. & Duan, D., 2006. C-Terminal-Truncated Microdystrophin Recruits Dystrobrevin and Syntrophin to the Dystrophin-Associated Glycoprotein Complex and Reduces Muscular Dystrophy in Symptomatic Utrophin/Dystrophin Double-Knockout Mice. *Molecular Therapy*, 14(1), pp.79–87. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16563874> [Accessed September 8, 2017].
- Zatz, M. et al., 1991. Serum creatine-kinase (CK) and pyruvate-kinase (PK) activities in Duchenne (DMD) as compared with Becker (BMD) muscular dystrophy. *Journal of the neurological sciences*, 102(2), pp.190–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2072118> [Accessed November 24, 2017].
- Zetsche, B. et al., 2015. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, 163(3), pp.759–771. Available at: <http://www.cell.com/article/S0092867415012003/fulltext> [Accessed September 26, 2015].

- Zhang, F. et al., 2011. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature*, 29(2), pp.149–154.
- Zhang, J.-P. et al., 2017. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biology*, 18(1), p.35. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28219395> [Accessed November 7, 2017].
- Zhang, T. et al., 2017. Production of Guide RNAs in vitro and in vivo for CRISPR Using Ribozymes and RNA Polymerase II Promoters. *Bio-protocol*, 7(4). Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28603751> [Accessed November 6, 2017].
- Zhao, L. et al., 2007. The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *The EMBO Journal*, 26(9), pp.2432–2442. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17410205> [Accessed November 26, 2017].
- Zhu, X. et al., 2015. An Efficient Genotyping Method for Genome-modified Animals and Human Cells Generated with CRISPR/Cas9 System. *Scientific Reports*, 4(1), p.6420. Available at: <http://www.nature.com/articles/srep06420> [Accessed December 10, 2017].
- Zincarelli, C. et al., 2008. Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Molecular therapy : the journal of the American Society of Gene Therapy*, 16(6), pp.1073–80. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1525001616317324> [Accessed November 17, 2017].
- Zuris, J.A. et al., 2014. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature Biotechnology*, 33(1), pp.73–80. Available at: <http://www.nature.com/doi/10.1038/nbt.3081>.