Development of molecular medicine approaches for neuromuscular disease



Britt Hanson

DEPARTMENT OF PHYSIOLOGY, ANATOMY AND GENETICS

EXETER COLLEGE, UNIVERSITY OF OXFORD

Thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2021

Declaration

The work presented in this thesis was undertaken in Professor Matthew Wood's laboratory at the Department of Physiology, Anatomy and Genetics, University of Oxford, between 2017 and 2021. I, Britt Hanson, hereby declare that all work presented is my own with the following exceptions. Assistance with injections of CRISPR-AAV into dKO mice and the preparation of tissue for immunohistochemistry staining for the experiments in results chapter 1 was provided by Sofia Stenler and Nina Ahlskog. The bioinformatics pipeline and uORF prediction analyses in chapter 3 were designed and carried out by Thomas Roberts. The validation of putative uORFs in the panel of oncogenes and TSGs in chapter 3 was done in collaboration with Matthew Drake and Nenad Svrzikapa. This work has not been submitted for any other degree at this university or any other institute of learning. My thesis includes text excerpts and figures of my first-author articles published throughout my DPhil as listed in **Section 8.1**. Where required, permission from individual journals to use figures and excerpts of text in this thesis has been obtained. These are accepted manuscripts of articles published by Taylor & Francis Group in RNA Biology, 2021, available online:
https://www.tandfonline.com/doi/full/10.1080/15476286.2021.1874161,
and by Cell Press in Molecular Therapy, 2020, available online:
https://www.sciencedirect.com/science/article/abs/pii/S1525001620305396.

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to Prof Matthew Wood for accommodating me in his lab, and providing me with endless support and advice throughout my DPhil. Thank you for allowing me the freedom to pursue research topics of my own interest and for encouraging me to be ambitious. Secondly, an enormous thanks must go to Dr Thomas Roberts for his unwavering supervision and concern for my academic progression throughout my DPhil. I have learnt an immense amount from you about a broad range of topics (both academic and otherwise) and I look forward to working with you in the future. Thank you for supporting me with funding for my research, and for spending an infinite amount of time helping me to refine my thesis and paper writing skills. Thank you Matthew and Tom for providing me with the opportunity to be an academic co-founder of a spin-out company so early on in my career.

Thank you to my generous funders, DPAG, Exeter College, the Clarendon Fund, the Medical Sciences Graduate School (Goodger and Schorstein Research scholarship), and the Oppenheimer Memorial Trust, without whom I would not have been afforded the opportunity to undertake my DPhil from the University of Oxford and to live so comfortably abroad for the past four years.

Thank you to the entire Wood lab - members past and present - who have provided me with a family away from home both in and out of the lab. To Amarjit Bhomra, thank you for making me feel so welcome and at home in Oxford from the moment I arrived. Thank you to Corinne Betts and Sofia Stenler for providing me with hands-on help and training with the animal work, for your endless patience and for sharing your invaluable expertise with me. To Nina Ahlskog, thank you for your additional support in the lab, particularly towards the end of my studies, and for assisting me so willingly to implement greener practice into our lab. Thank you to Nenad Svrzikapa for the interesting discussions, for your unwavering help throughout, and for being an integral part of starting Orfonyx. Thank you to Mariana Conceição and Imre Mäger for reading and advising me on my EV chapter, and for the incredible advice and support that you gave me during my DPhil. To Yulia Lomonosova, you have been the greatest source of happiness and motivation to me, and I look forward to many more years of tennis sessions and pub outings to come. To all of the other Wood lab Postdocs, you have each been a fantastic mentor to me and I have drawn a great deal of inspiration from you all. To Katarzyna

Chwalenia, I owe you my deepest gratitude for the hours you have spent helping me in the lab, and for being a bright and welcome presence in the office and outside the walls of Le Gros Clark. Thank you to Ambra Speciale, Ruth Ellerington, Ania Kordala, and Larissa Goli for an unusually extended (and wonderful) period of support and companionship during my studies, considering the rapid turnover of the Wood lab. Thank you also to all of the other students and RAs who have kept me smiling and motivated right from the start to the end. Thanks must also go to my many friends outside of lab, both in the UK and back home in South Africa, for your constant support and encouragement. To Megan Bybee in particular, for the numerous phone calls and constant checking in and for always making me feel infinitely closer to home.

To my academic mentor outside of the lab, Joey Riepsaame, thank you for your endless CRISPR advice and academic support over the years. To Lewis Fry, I am so grateful for the experiences we have shared, beginning at the Cold Spring Harbour conference in 2019. I look forward to seeing where your career takes you beyond the Oxford bubble, and hopefully getting the opportunity to work together at a later stage. To my long-term mentor, Marco Weinberg, since 2015 your contribution to my career progression has been incredibly valuable. Thank you for encouraging me to dream big, and for connecting me with the Wood lab to pursue my dreams of obtaining a DPhil at one of the leading institutions in the world. Without your advice and support I would never have imagined this to be possible.

To my wonderful partner, Matthew Ellis. There are no words that can fully describe or show adequate appreciation for the incredible presence that you have been in my life since I arrived in Oxford. Being able to share all of the experiences with you has made it even more wonderful, and these are moments that I will cherish forever. Thank you for keeping my spirits high, or swiftly lifting them when they plummeted, especially during the past few months while completing the DPhil.

A final and most important thank you goes to my fantastic family for not only supporting me emotionally and financially while I have been in Oxford, but for actively encouraging me to pursue my dreams and motivating me to keep going through the toughest times. I know the past four years have been incredibly difficult for you too as we have been so distant from one another, but I can assure you that I have taken none of this experience for granted and I have greatly appreciated your eagerness to live these moments with me through any possible means. I value each of you endlessly.

Abstract

Neuromuscular diseases are a broad group of debilitating disorders that impair muscle functioning. Duchenne muscular dystrophy (DMD) is the most common hereditary form of neuromuscular disease, for which there is currently no cure. DMD is caused by null mutations in the dystrophin gene (*DMD*), which encodes an essential cytoskeletal protein required for myofibre integrity and function. The leading DMD therapeutic strategy is antisense oligonucleotide-mediated 'exon skipping', however, this requires lifelong repeat administration. Single intervention strategies for the permanent correction of disease-causing mutations in the genomic DNA (such as CRISPR/Cas9, clustered regularly interspaced short palindromic repeat/CRISPR-associated protein 9) are important avenues for DMD therapeutic development. An AAV9-delivered CRISPR/Cas9 system was employed to excise a disease-causing *Dmd* exon in the severely affected double knockout (dKO) mouse (lacking dystrophin and its developmental paralogue, utrophin). Dystrophin protein expression was partially restored in the muscle of dKO mice, although no extension in lifespan was observed. Within-fibre patchiness of dystrophin expression was observed in longitudinal skeletal muscle sections, an as of yet unreported consequence of CRISPR-mediated *Dmd* gene editing in myofibres. This provides an insight into the myonuclear domain structure of dystrophic fibres, and raises concerns about the possible functionality of gene edited myofibres that contain interspersed dystrophin-positive and -negative regions.

The development of non-viral therapeutic strategies focused on improving muscle architecture are also of particular interest for neuromuscular diseases. Extracellular vesicles (EVs) are membrane-enclosed nanoparticles involved in natural cell-cell communication, believed to transfer pro-myogenic and anti-fibrotic factors in the muscle niche. Both myotube- and myoblast-derived EVs enhanced myogenic differentiation at lower, rather than higher, doses in recipient C2C12 muscle cell culture. Highly pro-myogenic effects observed in non-vesicular transfer experiments were attributed to the co-purification of insulin within the isolation medium, highlighting important methodological considerations for the EV field.

Strategies that enable the specific activation of gene expression by translational modulation are currently lacking. Upstream open reading frames (uORFs) present within the 5′ untranslated regions of mRNA transcripts function in *cis* to repress translation of the downstream primary ORF. The *in situ* disruption of uORFs would be therapeutically beneficial for upregulating expression of a disease-relevant protein or protective factor. Validation of a

panel of bioinformatically predicted uORFs revealed that not all are functional, including two uORFs with relevance to neuromuscular disease (i.e. *SMN2* and *UTRN* for SMA and DMD respectively). A complex interplay of uORF features was found to influence uORF activity, where initiation of uORF translation was essential for functionality. The biophysical properties of poly-amino acid residues in synthetic uORF micropeptides were found to be major determinants of uORF activity. A synthetic poly-acid uORF sequence was highly repressive, and truncations of this sequence resulted in predictable levels of transgene attenuation. Such uORFs could be useful in biotechnology for the development of tunable, inducible, and tissue-specific transgene expression systems. Finally, CRISPR/Cas13b-mediated RNA editing was applied for the upregulation of a reporter transgene by uORF start codon disruption.

Table of Contents

List of Figures ..... **xii**

List of Tables ..... **xv**

List of Abbreviations ..... **xvi**

1 Introduction **1**

1.1 Neuromuscular disease 1

1.2 Skeletal muscle structure and function 1

1.3 Duchenne muscular dystrophy 2

1.3.1 DMD genetics 2

DMD pathology 6

1.3.2 Animal models of muscular dystrophy 7

1.4 Genetic therapies for DMD 9

1.4.1 Mini- and micro-dystrophin gene therapy 10

1.4.2 Pre-mRNA targeted exon skipping 14

1.4.3 CRISPR/Cas9 genome engineering 17

1.4.3.1 Dual-cut-mediated exon excision 23

1.4.3.2 Single-cut-mediated exon reframing and splice signal disruption 25

1.4.3.3 Homology directed repair 28

1.4.3.4 Genomic DNA base editing 32

1.4.3.5 RNA editing 35

1.5 The skeletal muscle secretome 36

1.6 The role of EVs in cellular communication 38

1.6.1 Exosome biogenesis and secretion 39

1.6.2 Therapeutic exploitation of EVs 44

1.7 Protein translation as a mechanism of gene regulation 48

1.7.1 Canonical eukaryotic translation 48

1.7.2 *Trans* mechanisms of translational control 50

1.7.3 The discovery and annotation of small open reading frames 51
1.7.4 Upstream open reading frame (uORF)-mediated *cis* translational control 53
1.7.5 Medical relevance of uORFs 57

1.8 Thesis hypotheses and aims 61

2 Materials and methods 62

2.1 Cell culture 62

2.2 Golden Gate Assembly 62

2.2.1 Phosphorylation and annealing of ssODNs 62
2.2.2 Backbone digestion 62
2.2.3 Ligation and bacterial transformation 63

2.3 Ligation, transformation, and plating 63

2.4 DNA purification 63

2.4.1 PCR products and restriction enzyme-digested plasmid backbones 63
2.4.2 Plasmid vector DNA 64
2.4.3 Genomic DNA obtained from mouse tissue 64

2.5 RNA purification 64

2.5.1 Maxwell RSC simplyRNA Tissue Assay 64

2.6 Reverse transcription quantitative PCR (RT-qPCR) 65

2.6.1 cDNA synthesis 65
2.6.2 qPCR using SYBR Green 65

2.7 Western blotting for proteins obtained from cell culture 65

3 Results I - *In vivo* dystrophin rescue by genomic exon deletion using the *Staphylococcus aureus* CRISPR/Cas9 system 69

3.1 Introduction 69

3.2 Hypothesis and aims 72

3.3 Materials and methods 73

3.3.1 Cell culture 73

3.3.2 Animals 73
3.3.2.1 Genotyping dKO mice 74
3.3.3 CRISPR/Cas9 plasmid design 75
3.3.4 Dual sgRNA screening in murine 10T1/2 fibroblast cells 76
3.3.4.1 Transfections 76
3.3.4.2 Genomic DNA extraction, PCR, and Sanger sequencing 77
3.3.5 dKO treatment with CRISPR-AAVs 78
3.3.5.1 Injections and tissue harvesting 78
3.3.5.2 Genomic DNA extraction, PCR, and Sanger sequencing 78
3.3.5.3 RNA extraction and RT-qPCR 79
3.3.5.4 Protein isolation and Western blotting 80
3.3.5.5 Immunohistochemistry staining 81
3.4 Results 83
3.4.1 Selection of dual sgRNAs for *Dmd* exon 23 excision in murine fibroblasts 83
3.4.2 Dystrophin restoration in dKO mice by IP administration of CRISPR-AAVs 85
3.4.3 Dystrophin restoration in dKO mice by IV administration of CRISPR-AAVs 94
3.4.4 Dystrophin restoration by CRISPR-AAV treatment was insufficient to extend the lifespan of the dKO mice 99
3.4.5 CRISPR-AAV treated skeletal muscle exhibits within-fibre patchiness of dystrophin expression 101
3.5 Discussion 103
3.5.1 Application of the dual-cut CRISPR strategy *in vivo* 104
3.5.2 The effect of constitutive low levels of dystrophin expression in dKO mice 106
3.5.3 Other splice correction strategies assessed in the dKO mouse 107
3.5.4 Uniformity of dystrophin expression 109
3.5.5 Non-productive gene editing outcomes 110
3.5.6 Limitations and future directions 114
3.5.7 Conclusions 117
4 Results II - Investigation of the pro-myogenic potential of extracellular vesicles and secreted proteins derived from differentiating myotubes 118
4.1 Introduction 118

4.2 Hypothesis and aims 122

4.3 Materials and methods 123

4.3.1 Cell culture 123
4.3.2 Immunofluorescence staining of myotubes 123
4.3.3 Small molecule inhibition of EV secretion and uptake 125
4.3.4 siRNA knock down of exosome biogenesis markers *Rab27a* and *Rab27b* 125
4.3.5 Crude fractionation of myotube conditioned media 127
4.3.6 Transfer of ultrafiltration (UF)-derived CM fractions to C2C12 myoblasts 128
4.3.7 Pure isolation of extracellular vesicles by ultrafiltration size exclusion liquid chromatography (UF-SEC) 129

4.3.7.1 Collection of C2C12 myoblast and myotube conditioned cell culture media 129
4.3.7.2 Ultrafiltration of myoblast- and myotube-CM by tangential flow filtration (TFF) and spin filtration 130
4.3.7.3 Size exclusion liquid chromatography 130

4.3.8 Characterisation of UF-SEC purified EVs 131

4.3.8.1 Nanoparticle tracking analysis (NTA) 131
4.3.8.2 Western blotting 131

4.3.9 Transfer of EVs to C2C12 myoblasts 132
4.3.10 Transfer of non-vesicular factors from C2C12 myotubes isolated by UF-SEC 132
4.3.11 Treatment of C2C12 cultures with BSA, insulin, and transferrin 133

4.4 Results 134

4.4.1 Small molecule inhibitors of exosome release and EV uptake impairs myogenic differentiation 134
4.4.2 Combined siRNA knockdown of *Rab27a* and *Rab27b* exosome biogenesis factors inhibits myogenic differentiation 136
4.4.3 EV-containing myotube-conditioned media enhances myogenic differentiation 139
4.4.4 Myotube-derived EVs exert a pro-myogenic effect at low doses 143
4.4.5 The choice of isolation medium determines the effect of myotube-derived secreted protein transfer on myogenic differentiation 148
4.4.6 Insulin within Opti-MEM confounds myogenic differentiation experiments 155

4.5 Discussion 158

4.5.1 The functional role of EVs in the transfer of pro-myogenic signals 160

4.5.2 The effective dose of EVs in transfer experiments 162
4.5.3 Conclusions 164

5 Results III - Elucidating the functional role of upstream open reading frames 165

5.1 Introduction 165

5.2 Hypothesis and aims 167

5.3 Materials and Methods 168

5.3.1 Cell culture 168
5.3.2 uORF prediction 168
5.3.3 Luciferase assays 169

5.3.3.1 Dual-Glo Luciferase Assay System 169
5.3.3.2 Nano-Glo HiBiT Lytic Detection System 170

5.3.4 RT-qPCR 171
5.3.5 Transgene attenuation by synthetic uORFs 171
5.3.6 CRISPR/ADAR-mediated RNA editing 172

5.4 Results 174

5.4.1 Prediction of uORFs in the human and mouse transcriptomes 174
5.4.2 Validation of predicted uORF functionality 178
5.4.3 Analysis of uORF structure and function 193

5.4.3.1 The effect of uORF length and Kozak consensus on uORF activity 196
5.4.3.2 The effect of uORF amino acid composition on uORF activity 202
5.4.3.3 The effect of uORF nucleic acid sequence and structure on uORF activity 214
5.4.3.4 The effect of uORF position within the 5′ UTR on uORF activity 218

5.4.4 The *HOXA11* uORF acts only in *cis* 223
5.4.5 Applications of uORFs 225

5.4.5.1 uORF-mediated transgene attenuation 225
5.4.5.2 uORF knockout by CRISPR/ADAR-mediated RNA editing 228

5.5 Discussion 234

5.5.1 Putative uORF functionality 234

5.5.2 A hierarchy of uORF feature effects 235

5.5.2.1 Initiation of uORF translation 236

5.5.2.2 uORF amino acid identity and length 238

5.5.2.3 uORF position and Kozak sequence context 241

5.5.3 Future prospects of uORF research 243

5.5.3.1 Mechanism of uORF-mediated translational repression 243

5.5.3.2 Biotechnological applications of uORFs for transgene expression 244

5.5.3.3 Disruption of uORFs in human disease 245

5.5.4 Conclusions 246

6 Discussion and conclusions 248

7 References 256

8 Appendix 307

8.1 Manuscripts accepted for publication 307

8.2 Conferences attended 308

List of Figures

Figure 1.1 Human dystrophin gene and Dp427 isoform domain structure. 5

Figure 1.2 SpCas9 and SaCas9 primary protein sequence and complex with sgRNA and genomic target site. 22

Figure 1.3 CRISPR/Cas9 dual-cut-mediated exon excision for *Dmd* reading frame restoration. 24

Figure 1.4 CRISPR/Cas9 single-cut-mediated reframing or splice signal disruption for *Dmd* reading frame restoration. 27

Figure 1.5 CRISPR/Cas9-mediated homology directed repair for *Dmd* reading frame restoration. 31

Figure 1.6 CRISPR-mediated base editing for *Dmd* reading frame restoration. 34

Figure 1.7 Extracellular vesicle biogenesis. 43

Figure 1.8 Ribosomal fates during scanning of uORF-containing mRNA transcripts. 56

Figure 3.1 Design and validation of a dual sgRNA CRISPR/SaCas9 system for *Dmd* exon 23 excision in cell culture. 84

Figure 3.2 Dystrophin restoration in dKO mice following neonatal IP injection of CRISPR-AAVs. 87

Figure 3.3 Dystrophin protein is undetectable by Western blot in saline-treated dKO mouse tissues. 90

Figure 3.4 Dystrophin restoration in tissue sections from IP-treated dKO muscle. 92

Figure 3.5 Dystrophin restoration in dKO mice following neonatal IV injection of CRISPR-AAVs. 96

Figure 3.6 Dystrophin restoration in tissue sections from IV-treated dKO muscle 98

Figure 3.7 Treatment of dKO mice with CRISPR-AAVs by IP or IV injection has no effect on lifespan. 100

Figure 3.8 CRISPR-mediated correction of dystrophin results in within-fibre patchiness in longitudinal sections of the TA muscle. 102

Figure 3.9 Possible editing outcomes resulting from dual sgRNA CRISPR-mediated exon excision. 112

Figure 4.1 Inhibition of EV release and uptake impairs myoblast differentiation. 135

Figure 4.2 Combined siRNA-mediated knock down of *Rab27a* and *Rab27b* exosome biogenesis factors reduces myogenic differentiation. 138

Figure 4.3 EV-containing myotube-conditioned media enhances myoblast differentiation. 141

Figure 4.4 Myoblast- and myotube-derived EVs enhance myogenesis at lower doses. 146

Figure 4.5 Low molecular weight myotube-derived secreted proteins enhance myogenic differentiation when purified from Opti-MEM CM. 151

Figure 4.6 A non-specific increase in extracellular protein does not enhance myogenic differentiation. 152

Figure 4.7 Pro-myogenic factors in Opti-MEM confound results from myotube-derived secreted protein transfer experiments. 154

Figure 4.8 Insulin, but not transferrin, enhances myogenic differentiation in C2C12 cell cultures. 157

Figure 5.1 Properties of computationally predicted uORFs. 177

Figure 5.2 Generation of a dual-luciferase reporter vector containing a gene-of-interest 5′ UTR cloning site. 180

Figure 5.3 Generation of pDLR GOI 5′ UTR vectors. 183

Figure 5.4 Validation of predicted uORFs in a panel of oncogenes and TSGs. 187

Figure 5.5 Validation of predicted uORFs in *SMN2* and *UTRN* genes with relevance to neuromuscular disease. 192

Figure 5.6 Generation of *HOXA11* uORF cloning site in the pDLR vector. 194

Figure 5.7 Generation of *HOXA11* mutant uORF pDLR vectors. 195

Figure 5.8 Effect of uORF length and Kozak sequence context on uORF activity. 199

Figure 5.9 Detection of HiBiT-tagged *HOXA11*-uORF micropeptides. 201

Figure 5.10 The effect of single amino acid changes on *HOXA11* uORF activity. 203

Figure 5.11 The effect of homopolymeric amino acid chains on uORF activity. 206

Figure 5.12 The effect of truncating homopolymeric amino acid chains on uORF activity. 209

Figure 5.13 The effect of complex *HOXA11* uORF functional mutants on activity. 212

Figure 5.14 The effect of uORF nucleic acid sequence and structure on uORF activity. ...216

Figure 5.15 The effect of position on uORF activity within a uORF-naïve 5′ UTR. 219

Figure 5.16 Activity of the *SMN2* uORF depends on surrounding sequence context. 222

Figure 5.17 Translation initiation at a uORF acts on pORF in *cis*. 224

Figure 5.18 Transgene attenuation with truncations of hyper-functional uORFs. 227

Figure 5.19 Design and screening of a panel of crRNAs for RNA editing of the *HOXA11* uORF start codon. 231

Figure 5.20 Non-canonical start codon mutants of the *HOXA11* WT and hyper-functional uORFs. 233

List of Tables

Table 1.1 Mini- and micro-dystrophin clinical trials currently in progress 13

Table 1.2 Studies investigating the role of muscle cell EVs on myogenic differentiation and methods of EV isolation employed 46

Table 1.3 Disease-causing uORF mutations 59

Table 2.1 Plus and minus strand oligonucleotide sequences for sgRNAs and crRNAs 67

Table 2.2 Primary antibody information 68

Table 2.3 Secondary antibody information 68

Table 3.1 TaqMan primer and probe sequences for dKO mouse genotyping 75

Table 3.2 Target sequences of SaCas9 sgRNAs for targeted removal of *Dmd* exon 23 76

Table 3.3 Primer sequences for genotyping of *Dmd* exon 23 excision 77

Table 3.4 TaqMan primer and probe sequences for *Dmd* exon 23 exon skipping RT-qPCR 80

Table 4.1 ON-TARGETplus SMARTpool siRNAS targeted to *Rab27a* and *Rab27b* 125

Table 4.2 Primer sequences for RT-qPCR of *Rab27a*, *Rab27b*, and *Rplp0* reference gene 126

Table 5.1 Sequencing primer for confirmation of pDLR vector creation 169

Table 5.2 Primer sequences for RT-qPCR detection of *RLuc* and *FLuc* mRNA transcripts 171

Table 5.3 Sequences of CRISPR/ADAR crRNAs targeted to the *HOXA11 5′* UTR 173

Table 5.4 5′ UTR and translated uORF sequences of a panel of oncogenes and TSGs 188

List of Abbreviations

AAV	Adeno-associated virus
ABD	Actin binding domain
ADAR	Adenosine deaminase acting on RNA
ASO	Antisense oligonucleotide
BMD	Becker muscular dystrophy
bp	Base pairs
Cas9	CRISPR associated protein 9
cDNA	Complementary DNA
CM	Conditioned media
CPP	Cell-penetrating peptide
CRISPR	Clustered regularly interspaced short palindromic repeat
crRNA	Crispr RNA
$CXMD_J$	Canine X-linked muscular dystrophy in Japan
DAPC	Dystrophin-associated glycoprotein complex
dKO	*Utrophin/dystrophin* double knock out mouse
DLR	Dual-luciferase reporter
DMD	Duchenne muscular dystrophy
DMD	Dystrophin gene (human)
Dmd	Dystrophin gene (mouse)
DNA	Deoxyribonucleic acid
DSB	Double-strand break
ECM	Extracellular matrix
eEF	Eukaryotic elongation factor
eIF	Eukaryotic initiation factor
eRF	Eukaryotic peptide chain release factor
ESE	Exonic splicing enhancer
EV	Extracellular vesicle
gDNA	Genomic DNA
GOI	Gene-of-interest
GRMD	Golden retriever muscular dystrophy
GTP	Guanosine triphosphate

HDR	Homology directed repair
IF	Immunofluorescence
IHC	Immunohistochemistry
ILV	Intraluminal vesicle
IP	Intraperitoneal
IV	Intravenous
kDa	Kilodaltons
lncRNA	Long non-coding RNA
MB-EVs	EVs derived from myoblast (undifferentiated) cell cultures
MHC	Myosin heavy chain
miRNA	MicroRNA
MISEV	Minimal guidelines from the International Society for Extracellular Vesicles
MT-EVs	EVs derived from myotube (differentiated) cell cultures
MVB	Multivesicular body
myomiR	Muscle-specific miRNA
NHEJ	Non-homologous end joining
NMD	Nonsense-mediated decay
NOS1	Neuronal nitric oxide synthase
nt	Nucleotide(s)
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
pDLR	Dual-luciferase reporter vector
PMO	Phosphorodiamidate morpholino oligonucleotide
poly(A)	Polyadenylation
pORF	Primary ORF/coding DNA sequence
PPMO	Peptide-phosphorodiamidate morpholino oligonucleotide
RNA	Ribonucleic acid
RNAi	RNA interference
RT-qPCR	Reverse transcription quantitative PCR
SA	Splice acceptor
SaCas9	Staphylococcus aureus Cas9
SEC	Size exclusion liquid chromatography
sgRNA	Single guide RNA

siRNA	Small interfering RNA
SMA	Spinal muscular atrophy
SMN2	Survival motor neuron 2 gene (human)
smORFs	Small open reading frames
SNP	Single nucleotide polymorphism
ssODN	Single-stranded oligodeoxynucleotide
TIS	Translation initiation site
tracrRNA	Trans-activating RNA
tRNA	Transfer RNA
TSG	Tumour suppressor gene
TSS	Transcription start site
U	Units
UF	Ultrafiltration
uORF	Upstream open reading frame
UTR	Untranslated region
UTRN	Utrophin gene (human)
vg	Vector genomes
WT	Wild-type

1 Introduction

1.1 Neuromuscular disease

Neuromuscular diseases are a heterogenous group of hereditary and acquired disorders affecting various components of the neuromuscular unit, including skeletal and cardiac muscle, motor neurons, peripheral nerves, and the neuromuscular junctions. These disorders are highly diverse with respect to the genetic aberration or cause, as well as the age of onset, and disease severity. With recent advancements to molecular biology and genetic tools for genome-wide sequencing and bioinformatics analysis, an immense number of genes have been identified as being involved in neuromuscular disease [1], and rare variants are continually being discovered.

Muscular dystrophy is a major group within the neuromuscular disorders, encompassing primary diseases of the muscle tissue wherein dystrophic changes are visible by muscle biopsy [2]. Of these, Duchenne muscular dystrophy (DMD, OMIM #300377) is the most common form of congenital muscle-wasting disease in children. DMD is often referred to as a dystrophinopathy as the causal mutations arise in the dystrophin gene (*DMD*), which encodes a protein that is integral to muscle fibre stability and functioning. DMD is highly debilitating, progressive, and ultimately fatal, affecting ~1:3,500-1:5,000 boys worldwide [3]. An estimated two thirds of patients inherit a deleterious mutation from their mother, often an unknowing carrier, while the remainder of cases are believed to result from *de novo* germline mutations [4]. As such, it is not possible to eradicate DMD by genetic counselling alone, and so treatment and optimal management of this disease remain urgent unmet clinical needs.

1.2 Skeletal muscle structure and function

Skeletal muscle is the largest human tissue, accounting for ~40% of the total body mass [5]. While the major function of skeletal muscle is to facilitate locomotion, it has also been classified as a critical endocrine organ by the release of muscle-specific cytokines (i.e.

myokines) that serve to maintain energy homeostasis [6–9]. Myogenesis is the process whereby mononuclear myoblasts (i.e. muscle cells) undergo differentiation and fusion to generate multinucleated myotubes. Myogenesis supports the process of muscle development during embryogenesis, as well as regeneration of postnatal injured or diseased muscle tissue.

Myofibres are formed by the maturation of myotubes and are the basic contractile units of the muscle [10]. Myofibres are characterised by the expression of late stage markers of differentiation, such as dystrophin, muscle creatine kinase, desmin, titin, various troponins, and fibre type-specific isoforms of myosin heavy chain (MHC) [11]. The sarcolemma (i.e. muscle cell membrane) of individual myofibres is surrounded by the basal lamina which is a fibrous network of extracellular proteins comprised of type IV collagen and laminin polymers, linked via glycoproteins (e.g. dystroglycan) and heparan sulphate proteoglycans (e.g. nidogen and perlecan) [12,13]. Laminin is a heterotrimeric protein consisting of α, β, and γ subunits [14]. The laminin-α2-containing isoform (laminin-211) is the predominant variant within the muscle basal lamina, although other isoforms are present at neuromuscular junctions, and within the muscle tissue between myofibres and vasculature [15,16].

1.3 Duchenne muscular dystrophy

1.3.1 DMD genetics

DMD is an X-linked disease as it is caused by loss-of-function mutations in the *DMD* gene located on the X chromosome (Xp21.2-p21.1). *DMD* is the largest known human gene, spanning ~2.3 megabases of genomic DNA [3,17]. The *DMD* gene is comprised of 79 exons (**Figure 1.1A**) [3,17]. Seven independent, tissue-specific promoters drive the expression of different dystrophin isoforms which are characterised based on their length. The predominant variant in skeletal muscle is the largest isoform and encodes a 3,685 amino acid, 427 kDa rod-shaped protein (Dp427) [17–21]. Dp427 is also expressed in the brain (in the cortex and

hippocampal neurons) [22] and cerebellar Purkinje fibres [23]. Five other shorter isoforms have been detected in other tissues around the body, including the retina (Dp260 [24,25], and Dp140 [26]), brain (Dp140 [26], and Dp40 [27,28]), kidneys (Dp140 [26]), and peripheral nerves (Dp116 [29]). The shortest isoform (Dp71) is ubiquitously expressed, although with higher levels in the central nervous system [30,31]. The natural function of these variant dystrophin products is incompletely understood. The Dp427 isoform is comprised of an N-terminal domain which is the primary actin-binding region (exons 1-8, ABD1), a central rod domain (exons 9-63), a cysteine-rich domain (exons 64-69), a C-terminal domain (exons 70-79), as well as four hinge regions (**Figure 1.1B**). The central rod domain consists of 24 spectrin repeat units, each ~110 amino acids in length and consisting of triple α-helices which fold into small rod-like structures (~5 nm long) [18,32]. In skeletal muscle cells a microtubule lattice is created by the interaction between dystrophin spectrin repeats 20-23 and the intracellular microtubules [33,34]. A second actin binding domain (ABD2) exists within the central region of the rod domain that co-ordinates with ABD1 to facilitate the strong lateral association of dystrophin with intracellular actin filaments [35]. The dystrophin rod domain is made flexible through the interruption of the spectrin repeats by four proline-rich 'hinge' regions [36].

Dystrophin provides a crucial structural link in myofibres by the interaction with cytoskeletal actin via its N-terminus, and connection to the extracellular matrix (ECM) via the dystrophin-associated glycoprotein complex (DAPC) [3,18]. Other DAPC components include the dystroglycan complex (i.e. α-dystroglycan, and β-dystroglycan), the sarcoglycan complex (α-, β-, γ-, and δ-sarcoglycan), dystrobrevin, syntrophin, sarcospan, and neuronal nitric oxide synthase (NOS1) [37]. The α-Dystroglycan subunit of the dystroglycan complex binds to laminin within the ECM in a calcium-dependent manner, and is anchored to the sarcolemma by the β-dystroglycan subunit [38–40]. The fourth hinge region within the central rod domain of dystrophin is required for intracellular binding of the C-terminus of β-dystroglycan [41] (**Figure 1.1B**).

Dystrophin also plays an important role in myofibre signalling, such as for the production of nitric oxide mediated by NOS1 through interactions with spectrin repeats 16 and 17 [42–44] (**Figure 1.1B**).

Satellite cells are the resident muscle stem cells that are located between the basal lamina and the sarcolemma, and which typically remain quiescent until activated by exercise-induced injury or contractile damage [45]. The regenerative function of satellite cells mediates the plasticity of muscle despite this being a largely terminally differentiated tissue [45]. Asymmetric division of activated satellite cells provides myoblasts for myofibre growth and repair, as well as new satellite cells to replenish the pool of stem cells for future rounds of activation and regeneration [46]. Satellite cell exhaustion as a result of dystrophin-deficient myofibre sensitivity to contractile damage has been proposed to be a major contributor towards the debilitation of dystrophic muscle tissue over time [47]. There is evidence to suggest, however, that satellite cells from aged dystrophic muscle retain their myogenic potential when removed from the dystrophic niche, suggesting that this 'exhaustion' is a functional fatigue as a consequence of compromised muscle environment rather than premature senescence or loss of stem cell potential [48,49].

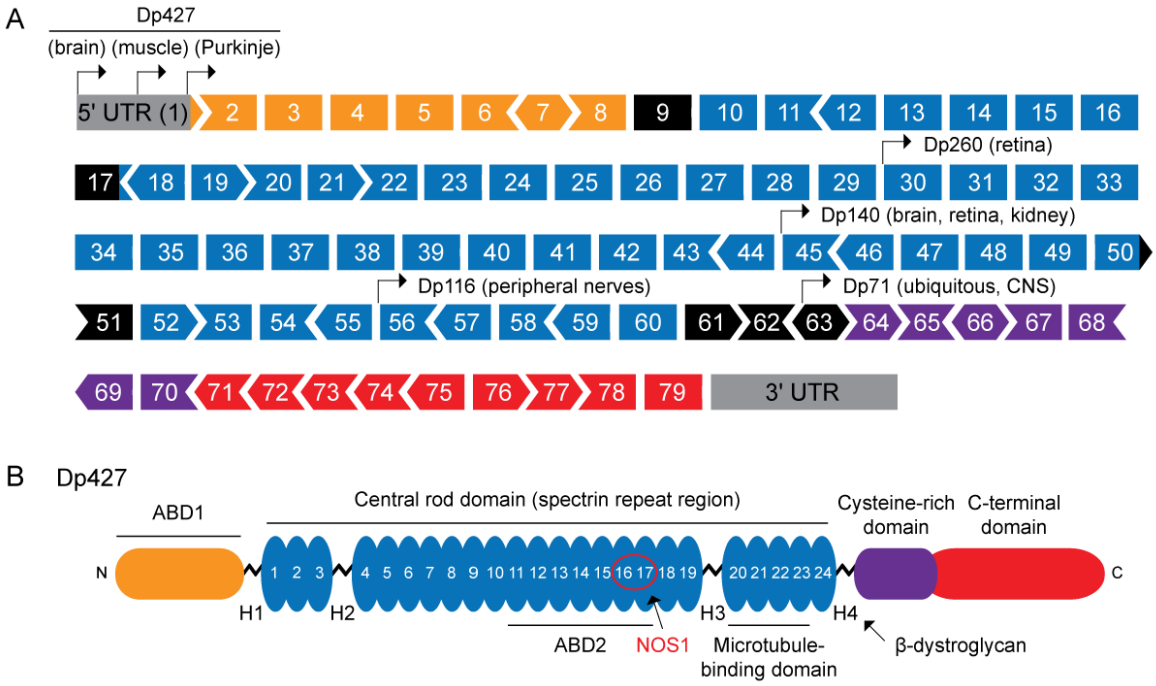


Figure 1.1 Human dystrophin gene and Dp427 isoform domain structure.

(**A**) The human dystrophin gene (*DMD*) contains 79 exons and expression of the isoforms (named according to the mass of the resulting protein product in kDa) is driven by seven independent promoters. Dp427 is the largest isoform, and is expressed in the brain, muscle, and Purkinje fibres. Dp260 is expressed in the retina, Dp140 in the brain, retina, and kidneys, Dp116 peripheral nerves, and Dp71 is ubiquitous with higher levels of expression in the central nervous system. Exon 1 is non-coding, and is contained within the 5′ untranslated region (UTR). The shape of the exons represents the dystrophin mRNA reading frame such that when a codon is split across two exons this is indicated via an arrowhead. This panel of the figure was adapted from [50]. (**B**) DMD is a 3,685 amino acid, 427 kDa protein, and is comprised of an actin-binding N-terminal domain (ABD1), four hinge regions (H1-4), 24 spectrin repeats within the rod domain, a second actin binding domain (ABD2), a microtubule-binding domain, a cysteine-rich domain, and a C-terminal domain. The hinge regions are required for dystrophin protein flexibility. H4 and neighbouring motifs (EF hands) interact with β-dystroglycan (a component of the dystrophin-associated glycoprotein complex, DAPC). Spectrin repeats 16 and 17 interact with neuronal nitric oxide synthase (NOS1).

DMD pathology

The genetic loss of dystrophin increases the fragility of the sarcolemma by disassembly of the DAPC, thereby sensitising it to contractile damage [51]. Similarly, mutations in other DAPC components are causative of other variants of muscular dystrophy [52]. Sarcolemmal damage leads to chronic cycles of myonecrosis and compensatory regeneration, together with persistent immune cell infiltration and thus, over time, muscle quality declines as a consequence of extensive fibrosis and fat deposition [53–55]. Boys affected with DMD typically present with delays in the development of certain motor skills (e.g. independent ambulation and standing from a supine position) by three years of age. Progressive loss of peripheral muscle function results in DMD patients becoming wheelchair-bound, typically by the age of twelve [56]. Affected individuals suffer from diaphragm-related respiratory complications, cardiomyopathy, and ultimately succumb to the disease by the age of thirty [57–59]. In general, the severity of DMD results from null mutations in the muscle-specific Dp427 isoform, however, concomitant changes in expression of the shorter isoforms have also been linked to certain other disease phenotypes, particularly those relating to cognitive impairment [60–62]. Debilitating bone-related effects are common in children with DMD as a result of reduced weight-bearing physical activity, including osteopaenia, fragility fractures, scoliosis, and contractures [63].

Utrophin (*UTRN*) is the autosomally-encoded (6q24), 395 kDa developmental paralogue of dystrophin, sharing 80% sequence homology in the C-terminal domain [64] and 65% nucleic acid identity overall [65]. Utrophin and dystrophin also exhibit similar intron-exon arrangements, however, expression is regulated in a spatiotemporally distinct manner [66]. During embryonic development, utrophin is localised to the sarcolemma but is progressively replaced by dystrophin up until birth [66]. Thereafter, utrophin is localised primarily at the neuromuscular and myotendinous junctions [24], although expression is observed in the

sarcolemma of regenerating myofibres, and in other non-muscle tissues [67]. Utrophin is also upregulated at the sarcolemma of skeletal muscle in DMD patients and *mdx* mice (i.e. the most widely-used murine model of muscular dystrophy) that lack functional dystrophin protein [68,69]. There are conflicting reports as to whether utrophin upregulation is fully compensatory and correlates with disease severity [70,71], however, exogenously-induced utrophin upregulation using a small molecule (i.e. SMT C1100/ezutromid) [72] or gene transfer using an adenovirus vector [73] was found to reduce skeletal muscle pathology in *mdx* mice. Therapeutics aimed at utrophin upregulation are currently being investigated for DMD treatment, although efforts are currently being made to improve the chemistry of ezutromid which failed at the 48-week timepoint in a Phase II clinical trial (NCT02858362), despite promising efficacy after 24-weeks [75].

1.3.2 Animal models of muscular dystrophy

More than 60 animal models of dystrophy have been reported to date [76]. Some of these were discovered fortuitously as a result of spontaneous mutations, while others have been engineered for some other purpose, for example, modelling of DMD. The most commonly used animal models for DMD research are the murine and canine models of muscular dystrophy. The *mdx* mouse (the most widely-used murine model) resulted from a spontaneous C>T mutation in *Dmd* exon 23 of a C57BL/10 mouse which was identified by elevated serum creatine kinase levels, a characteristic signature of dystrophic pathology [77,78]. The single nucleotide polymorphism (SNP) introduces a premature termination codon into exon 23, thereby disrupting dystrophin protein production. Despite a lack of dystrophin expression in these mice, disease pathology is relatively mild and the *mdx* lifespan is only marginally decreased relative to wild-type (WT) mice [79]. The pathology of the *mdx* diaphragm is similarly severe to that of DMD patients [80], while cardiomyopathy is delayed [54,81–84], creating a major disparity between the pathology of this murine model and DMD patients. Four chemical variants of the *mdx*

mouse have been generated using the mutagen N-ethyl-N-nitrosourea [85], whereby genetic mutations result in the loss of full-length dystrophin but differential expression of the other dystrophin isoforms [86]. These lines include *mdx*^{2cv}, *mdx*^{3cv} (which expresses ~5% dystrophin protein) [87], as well as *mdx*^{4cv} and *mdx*^{5cv} which both exhibit fewer revertant fibres [88] (i.e. rare occurrences of dystrophin-positive fibres in dystrophin null skeletal muscle tissue).

Compensatory upregulation of other genes are thought to account for the milder dystrophic pathology of murine models on the *mdx* background. For instance, utrophin and α7β1-integrin both perform similar scaffolding functions to dystrophin through linking of the ECM with the myofibre cytoskeleton. Several double knockout mouse models have thus been created to more closely resemble the DMD patient phenotype. Dystrophin knockout mutations have been combined with a null mutation in the cases of both utrophin [89,90], or α7β1-integrin [91,92]. Alternatively, double knockout mice have been generated to investigate the involvement of other key players in muscle structure, myogenesis, or inflammation (i.e. knock out of desmin [93], MYOD1 [94], dystrobrevin [95], dysferlin [96], or IL10 [97]). The utrophin/dystrophin double knockout mouse was employed for *in vivo* experiments in this thesis, and is hereafter referred to as the dKO mouse. These mice exhibit severe musculoskeletal defects leading to pronounced kyphosis (curvature of the spine), contractures in the forelimbs, and difficulty breathing as a result of diaphragm-related pathology [98]. Furthermore, these dKO mice have a reduced capacity for bone healing which is similar to that of DMD patients [98], making this a highly attractive murine model of muscular dystrophy.

The canine models are highly representative of human dystrophic pathology, experiencing a similar time of disease onset and reduction in lifespan to that observed in DMD patients [99–103]. Originally the most commonly used canine model was the golden retriever muscular dystrophy model (GRMD, otherwise known as canine X-linked muscular dystrophy in Japan, CXMD$_J$ [104]) which harbours a splice acceptor mutation at the border between intron

6 and exon 7 of the dystrophin gene [100,101]. Another model that is also currently prevalent and which has greater relevance to human DMD disease mutations is the deltaE50-MD model which carries a splice site mutation resulting in the exclusion of *DMD* exon 50 during mRNA maturation [105]. While there are many other canine models of muscular dystrophy, the major drawbacks of using these for preclinical research are the cost, access to appropriate animal facilities, and ethical considerations.

1.4 Genetic therapies for DMD

To date there is no cure for DMD. The majority of patients are treated with corticosteroids (e.g. prednisone/prednisolone, and deflazacort), which were introduced for symptomatic treatment of DMD in the 1990s and have been reported to improve dystrophic pathology and prolong ambulation [106]. A recent systematic literature review, however, highlights the challenges with drawing statistically significant evidence for the efficacy of corticosteroid use owing to a number of factors that cannot easily be controlled for (e.g. patient age at initiation of corticosteroid use, variable dose regimens, use in combination with other therapies, as well as limited numbers of corticosteroid-naïve control patients) [56]. Furthermore, osteotoxicity of corticosteroid therapies has been shown to exacerbate the bone-related phenotypes experienced by patients [107]. Other side effects of corticosteroid therapies include the development of cushingoid features, behavioural abnormalities, weight gain, and hypertension [106,108]. Physical therapies or interventions are also commonly used in DMD patients to improve their quality of life (e.g. speech, occupational, and physical therapy, and in later stages the use of a ventilator and/or pacemaker device). Investigation into the repurposing of existing FDA-approved drugs (e.g. tamoxifen) is also being carried out [109]. Importantly, while these approaches mentioned above are necessary and useful for the ongoing management of patient symptoms, they do not address the genetic cause of DMD. As such, several different genetic therapies are in various stages of development (and with some recent approvals discussed below) and testing in an

effort to uncover a more targeted therapeutic strategy for the alleviation of DMD disease pathology.

1.4.1 Mini- and micro-dystrophin gene therapy

DMD-causing mutations are highly diverse, posing a significant challenge for the development of broadly applicable therapies. DMD cannot be treated by gene replacement therapy using the full-length cDNA due to the enormity of the dystrophin gene (~2.3 Mb) and its transcript (~14 kb). The spectrin repeat units within the rod domain of dystrophin are, however, dispensable for protein function. This is exemplified by the related condition Becker muscular dystrophy (BMD, OMIM #300376) which is typically less severe, and presents with later onset, relative to DMD [110,111]. Both DMD and BMD are caused by mutations in the dystrophin gene, however, BMD-causing mutations typically do not disrupt the translation reading frame, leading to the production of an internally truncated dystrophin protein that retains some functionality [112]. Indeed, in some rare cases BMD patients are effectively asymptomatic [113,114].

To address the issue of DMD gene therapy, compact mini-/micro-dystrophin transgenes (<6 kb or <4 kb in length respectively) that lack large portions of the central rod domain have been developed [115–117]. The original design of the mini-dystrophin gene was inspired by a particular BMD patient with mild pathology despite lacking 46% of the dystrophin gene (i.e. from H2 through to the first half of spectrin repeat 19 was absent) [113]. Since the early 2000s, various iterative improvements have been made to the mini- and micro-dystrophin gene constructs to enable packaging within adeno-associated viral (AAV) vectors as these have a limited genome capacity of ~4.7 kb [118,119]. AAV is a single-stranded DNA *Parvovirus* that infects human cells but does not cause human pathology [120]. AAV vector genomes are typically maintained as episomal chromatin with the potential for long-term, persistent transgene expression [121]. AAVs are attractive gene therapy vectors as multiple serotypes exist with a

broad range of tissue tropisms (including for skeletal and cardiac muscle), they are able to transduce both dividing and non-dividing cells, and exhibit low pathogenicity in man [122,123]. Furthermore the non-integrative lifecycle of AAV reduces the risk of insertional oncogenesis, a major advantage over other commonly used viral vectors, such as lentivirus [124].

The AAV-deliverable mini- and micro-dystrophin therapies currently being tested in clinical trials are summarised in **Table 1.1**. Another gene therapy that provides the *GALGT2* gene (driven by the muscle-specific creatine kinase promoter) via the rAAVrh74 vector is also being investigated in a Phase I/II clinical trial in DMD patients (rAAVrh74.MCK.GALGT2, NCT03333590) [125]. GALGT2 is a β1,4 N-acetylgalactosaminyltransferase which, when overexpressed in skeletal muscle glycosylates α-dystroglycan and promotes recruitment of dystroglycan binding partners, as well as GALGT2 interacting proteins (usually localised to the neuromuscular and myotendinous junctions), to the sarcolemma thereby stabilising the membrane [125–128]. rAAVrh74.MCK.GALGT2 is an exploratory surrogate gene therapy approach as it does not replace the defective *DMD* gene itself, but is more broadly applicable to the group of muscular dystrophies (including DMD as well as limb girdle muscular dystrophy 2A, and congenital muscular dystrophy 1a) [125]. Immunity against novel dystrophin protein (i.e. to non-self antigens) in dystrophin-deficient animals has highlighted a possible limitation of mini- and micro-dystrophin gene therapies in humans [129]. To circumvent this, mini- and micro-utrophin gene therapies are also in development as utrophin protein is expressed in DMD patients and therefore is unexpected to elicit immune reactivity [115,130]. Achieving mini- and micro-dystrophin/utrophin transgene design that is compact yet which retains the important functional attributes of the central domain of dystrophin remains an ongoing challenge [131]. For this reason, other strategies that involve the direct manipulation of the dystrophin mRNA transcript or genomic DNA *in situ* are highly attractive therapeutic

options for DMD. These include pre-mRNA splice modulation via exon skipping, and gene editing respectively.

Table 1.1 Mini- and micro-dystrophin clinical trials currently in progress

Company	Clinicaltrials.gov ID	Phase	Recruitment ages/cohorts	Genetic therapy	AAV serotype	Ref
Solid Biosciences	NCT03368742	I/II	Ages 4-17 years 16 participants	Systemic IV infusion of SGT-001 micro-dystrophin (CK8 muscle-specific promoter)	AAV9	[132]
Sarepta Therapeutics	NCT03375164	I/II	3 months-3 years and 4-7 years 4 participants	Limb-specific IV infusion of SRP-9001 micro-dystrophin (MHCK7 muscle-specific promoter)	rAAVrh74	[133]
	NCT04626674	I	4-7 years 20 participants One dose	Systemic IV infusion of SRP-9001 micro-dystrophin (MHCK7 muscle-specific promoter)		
Pfizer	NCT03362502	Ib	4-12 years 30 participants	Systemic IV infusion of PF-06939926 mini-dystrophin (MCK7 muscle-specific promoter)	AAV9	[134]
	NCT04281485	III	5-12 years 99 participants			
Généthon/AFM-Téléthon and Sarepta Therapeutics	-	I	6-10 years	Systemic IV infusion of GNT 0004 (MD1) micro-dystrophin (Spc5.12 promoter)	rAAV2/8	[133,135]

1.4.2 Pre-mRNA targeted exon skipping

DMD-causing mutations occur throughout the *DMD* gene body, with the majority being whole exon deletions [136]. There are, however, 'hotspots' where mutations tend to cluster, such as at exons 45-55. This region has attracted particular interest, as the simultaneous removal of these eleven exons would be applicable to ~63% of DMD patients [137]. Furthermore, there are several reports of BMD patients who carry natural deletions of this region [137,138]. Exon skipping mediated by antisense oligonucleotides (ASOs) is currently the leading therapeutic strategy for the treatment of DMD, which aims to restore the dystrophin reading frame via RNA-targeted splice correction. Exon skipping is the modulation of alternative splicing in order that a specific exon, or combination of exons, are selectively excluded (or 'skipped') from the mature mRNA transcript [139]. Exon skipping aims to restore the reading frame of the mutant dystrophin transcript, and thereby convert the phenotype of DMD to BMD-like [140]. Exon skipping is most commonly achieved through the use of steric block ASOs, which are short (15–35 nt), single-stranded nucleic acid polymers that bind to splicing signals in the pre-mRNA via complementary base pairing, and thereby mask them from the spliceosome [141]. The leading approach in the case of DMD is to target removal of exon 51, which would theoretically be applicable for the treatment of ~13% of patients [140]. For example, patients carrying a deletion in *DMD* exon 52 can be treated by skipping exon 51, as exons 50 and 53 are in-frame with each other, leading to the generation of a partially truncated dystrophin protein that retains the majority of its functionality.

Multiple exon skipping in the case of DMD is technically challenging as skipping efficiencies are typically very low, variable between exons, and dependent on the order of intron removal [142]. It is worth noting the extraordinarily long distance that RNA polymerase II must travel between these exons (it has been estimated that it takes 16 hours to transcribe the *DMD* locus [143]) and that the splice modulation events at each exon must therefore be temporally

coordinated. Nevertheless, a recent report showed that a cocktail of 10 ASOs was able to induce exon 45–55 skipping in the *mdx52* mouse model of DMD [144]. Notably, high concentrations of oligonucleotides were required which raises potential toxicity issues, and the level of exon skipping was modest relative to that of single exon skipping [144].

To date, there are four FDA-approved exon skipping drugs for DMD: eteplirsen which targets exon 51, casimersen which targets exon 45, and golodirsen and viltolarsen which both target exon 53. Eteplirsen, casimersen, and golodirsen were developed by Sarepta Therapeutics (Cambridge, MA, USA), whereas viltolarsen was developed by NS-Pharma (Tokyo, Japan). All of these drugs are comprised of phosphorodiamidate morpholino oligonucleotide (PMO), a non-natural nucleotide analogue with a charge neutral phosphorodiamidate backbone and a six-membered morpholine ring (which substitutes for the arabinose in the equivalent RNA), and are administered via the intravenous route. PMOs have repeatedly been found to be remarkably safe, with very few adverse effects in clinical trials [145,146], and very high doses (up to 3 g/kg) are tolerated in mice [147]. In contrast, a rival compound, drisapersen (developed by Biomarin Pharmaceutical Inc., San Rafael, CA, USA), was discontinued due to lack of efficacy at Phase III [148,149]. This drug consists of a fully 2′-*O*-methyl substituted phosphorothioate-backbone RNA oligomer, and exhibited dose-limiting renal toxicity, which contributed to its inability to achieve efficacy goals.

Despite their low toxicity, effective delivery of PMOs to muscle (and especially to the heart, an organ that is critical to the lethality of DMD) remains a major obstacle to successful therapy. This is exemplified by the very low efficacy of the FDA-approved exon skipping drugs. Indeed, the approval of eteplirsen in September 2016 proved to be highly controversial [150–152]. The levels of dystrophin protein expression measured in muscle biopsies of eteplirsen clinical trial participants was ~1% of that expressed in healthy muscle [153], and is of debatable therapeutic relevance. Similarly, viltolarsen was recently shown to achieve ~5% dystrophin

correction [154]. Importantly, data from animal studies suggest that ~15% of wild-type (WT) dystrophin expression is sufficient to protect against dystrophic muscle pathology [155]. A further challenge for the treatment of DMD is the development of ASOs that target other exons (and therefore patient mutation types). For example, Sarepta is currently developing further therapeutic products targeting exons 43, 44, 50, 52, and 55 [156]. Furthermore, newer technologies are being developed to improve tissue targeting and uptake by the conjugation of PMOs to cell-penetrating peptides (referred to as PPMOs) by our group [157–160] and others [161,162].

An alternative to ASOs is the use of expressed exon skipping triggers (i.e. U1 and U7 small nuclear RNA expression systems) which can be delivered using AAV vectors. Notably, AAV-mediated dystrophin restoration (either by U7-mediated exon skipping or micro-dystrophin gene therapy) has been augmented by pre-treatment with PPMO [163]. The effects of AAV-mediated gene transfer are dependent on second strand DNA synthesis of the viral genome, meaning that there is a ~7–10 day delay between injection of the virus and the manifestation of therapeutic effects [164,165]. Importantly, the turnover of muscle fibres associated with dystrophic pathology means that viral genomes are progressively lost over time, including during this lag period [124,166]. In contrast, the correction of dystrophin transcripts mediated by ASOs, such as PPMO conjugates, begins immediately after cellular entry of the oligonucleotides. As such, transient dystrophin correction mediated by PPMO pre-treatment has been shown to stabilise muscle turnover and thereby enhance the effects of subsequently administered AAV therapies [163].

While ASOs are highly promising for the treatment of DMD, this approach necessitates lifelong repeat administration, constituting an enormous cost to healthcare systems and patients. Importantly, technologies that also involve RNA-based target recognition and engagement (i.e. CRISPR/Cas9, clustered regularly interspaced short palindromic repeat/CRISPR-associated protein 9), but which are aimed at permanently correcting disease-

causing mutations at the genomic DNA level, are rapidly emerging as novel approaches for treating DMD [167,168]. These exciting developments have the potential to supersede, or at least complement, the more established ASO strategies.

1.4.3 CRISPR/Cas9 genome engineering

Rational genome engineering involves the highly targeted manipulation of a user-defined target site within the genome. Previous generations of genome engineering technologies involved the use of ZFNs (zinc finger nucleases) and TALENs (transcription activator-like effector nucleases). While these are both highly effective tools for manipulating the human genome, they are cumbersome to design and expensive to produce, requiring that a unique synthetic protein be engineered for each target DNA site in order to induce the desired change [169,170]. In contrast, CRISPR/Cas9 is a simple and versatile technology that can be easily directed to recognise a specific genomic DNA site through careful design of a non-protein-coding guide RNA sequence, which programmes the complex [171–173].

CRISPR/Cas9 is derived from a prokaryotic natural defence system found in ~40% of bacteria and ~90% of archaea, which protects against foreign invading nucleic acid (e.g. from bacteriophages) in a sequence-specific manner [174–176]. The natural CRISPR genomic locus is comprised of a CRISPR array containing 'spacer' sequences interspersed between invariant precursor-CRISPR RNA (pre-crRNA) direct repeats, a non-coding trans-activating crRNA (tracrRNA), and the Cas9 gene [176,177]. 'Immunological memory' is acquired first by adaptation to the foreign invading nucleic acid by the integration of short fragments (known as protospacers) into the spacer region within the host genome CRISPR locus [173]. Thereafter, the spacer, a tracrRNA, and Cas9 endonuclease are expressed, and subsequent reinfecting sequences recognised by the crRNA component. Complementary base pairing of the Cas9:crRNA:tracrRNA complex with the invading target initiates double-strand break (DSB)

formation by the Cas9 endonuclease and inactivation of the foreign nucleic acid species [177]. A limiting factor of target recognition is the presence of a Cas9 orthologue-specific short protospacer adjacent motif (PAM) at the 3′ end of the target site in order to avoid self-recognition and genome destruction [177]. This natural system has since been repurposed and systematically engineered for applications in mammalian systems in culture [172] and *in vivo* [178] revolutionising the field of gene editing for research, genetic therapy, and biotechnology. The advantage of the CRISPR/Cas9 system is that the protein component is invariant, while the target sequence can be easily designed against many sites within the genome, facilitating rapid and inexpensive screening for activity.

The most widely adopted CRISPR/Cas9 systems are those derived from *Streptococcus pyogenes* (SpCas9) and *Staphylococcus aureus* (SaCas9), as these exhibit the most efficient *in vivo* cleavage activity reported to date [171,172,178]. A single guide RNA (sgRNA) is required to direct the Cas9 protein to the desired genomic target site for editing [173]. The Cas9 endonucleases are comprised of recognition (REC) and nuclease (NUC) lobes which are connected by an arginine-rich bridge helix (BH) [179,180] (**Figure 1.2A**). The REC lobe is required for recognition of a cognate sgRNA molecule with orthologue-specific structural features (i.e. referred to as the scaffold RNA) (**Figure 1.2B**). The sgRNA is generated as a chimera of the target site-specific 17-18 nt (SpCas9, **Figure 1.2B**) or 21-22 nt (SaCas9, **Figure 1.2C**) gRNA sequence, and the scaffold RNA [173,178]. In the natural bacterial system, the invariant RNA is encoded separately and hybridises post-transcription, however, this has been simplified for synthetic genome engineering by the fusion of the crRNA and tracrRNA sequences into a single non-coding RNA component. The scaffold sequences are comprised of a repeat:anti-repeat duplex, tetraloop region and two (SaCas9) or three (SpCas9) stem loops [173,178]. Extension of the tetraloop by five base pairs, and the inclusion of an adenine to uracil (A-U) flip in the anti-repeat region, improved recognition of the sgRNA by the wedge (WED) and REC domains

of SaCas9 (this is often referred to as the 'modified tracrRNA') [181] (**Figure 1.2A**). The Cas9 phosphate lock loop (PLL) plays a role in ATP-independent strand separation and heteroduplex formation between the gRNA and target DNA sequences, and the PAM interacting (PI) domain recognises this orthologue-specific motif on the non-target strand [179,180]. Cas9 unwinds the target DNA in a directional manner away from the PAM site whereby consecutive mismatches in the heteroduplex attenuates helicase activity [182]. The N-terminus of the REC lobe plays a role in identifying full complementarity between the first 8 bp (i.e. the 'seed' region) of the gRNA:DNA heteroduplex [179,180]. The REC C-terminus is required for PAM-distal heteroduplex identification before the two Cas9 catalytic domains become fully activated [179,180]. HNH and RuvC (which is comprised of three separate domains, RuvCI-III) are the highly conserved catalytic domains contained within the NUC lobe, responsible for targeted DNA cleavage of the gRNA complementary and non-complementary strands respectively [179,180]. The resulting DSB is introduced within the target recognition sequence, three nucleotides in from the PAM-proximal end [183] (**Figure 1.2B and C**). This lesion is resolved by one or more DNA repair pathways, including error-prone non-homologous end joining (NHEJ) and high-fidelity homology directed repair (HDR) [173,184,185]. CRISPR/Cas9 thus has the potential to disrupt/modify existing genes, or to introduce recombinant DNA. Alternatively, non-catalytic CRISPR/Cas9 systems can be combined with activator or repressor complexes to alter gene expression levels without permanently disrupting the sequence [186].

SpCas9 recognises a 5′-NGG trinucleotide PAM, a motif that occurs frequently throughout the human genome and thus enables broad targeting [171,172] (**Figure 1.2B**). The common occurrence of the SpCas9-specific canonical PAM sites, the potential for non-canonical 5′-NAG PAM recognition, and a high level of tolerance for mismatches between the spacer and similar off-target genomic sequences increase the risk of SpCas9 cleavage promiscuity [173,187,188]. In contrast, the SaCas9 endonuclease has a longer PAM recognition

requirement (5′-NNGRRT, where R is either A or G), which significantly reduces the number of potential cleavage sites within the human genome relative to SpCas9, thereby minimising unwanted off-target gene editing activity [178] (**Figure 1.2C**). SaCas9 also has the advantage of occupying a smaller DNA footprint than SpCas9 (3.3 kb compared with 4.1 kb) and is thus more easily packaged within the limited capacity of the AAV genome (~4.7 kb) [189]. SaCas9 is currently the orthologue-of-choice for the majority of *in vivo* gene editing studies, including for DMD, owing to its improved safety profile and compact nature [178]. Nevertheless, SaCas9 and its cognate sgRNA(s) have often still been co-administered separately, using two AAV vectors, for *in vivo* DMD CRISPR applications [181,190–192].

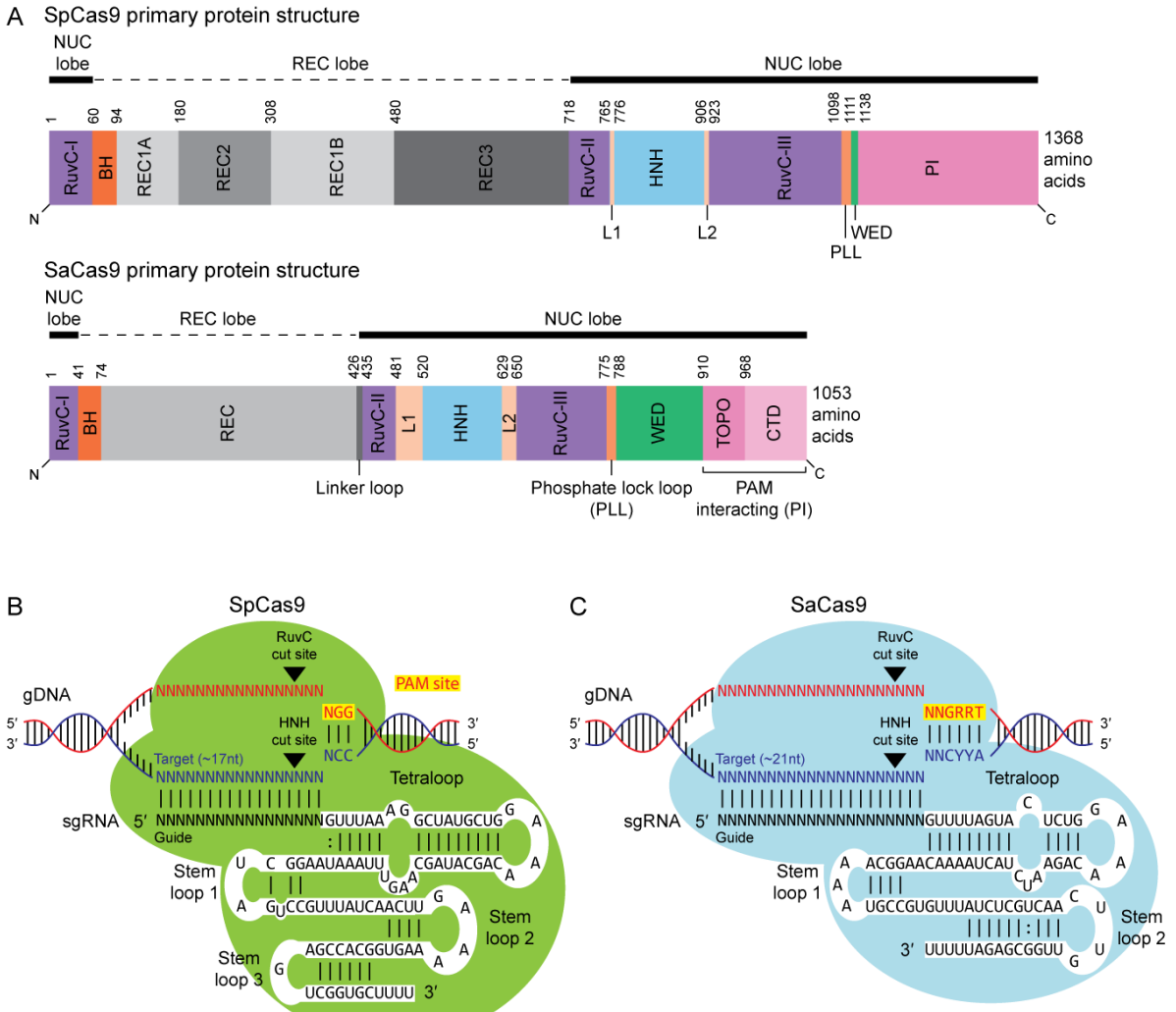

A
SpCas9 primary protein structure
NUC lobe
REC lobe
NUC lobe
RuvC-I
BH
REC1A
REC2
REC1B
REC3
RuvC-II
HNH
RuvC-III
PI
1368 amino acids
N
C
L1
L2
WED
PLL
SaCas9 primary protein structure
NUC lobe
REC lobe
NUC lobe
RuvC-I
BH
REC
RuvC-II
L1
HNH
L2
RuvC-III
WED
TOPO
CTD
1053 amino acids
Linker loop
Phosphate lock loop (PLL)
PAM interacting (PI)
B
SpCas9
RuvC cut site
PAM site
gDNA
NNNNNNNNNNNNNNNNNNN
NGG
HNH cut site
NCC
Target (~17nt)
Tetraloop
sgRNA
Guide
Stem loop 1
Stem loop 2
Stem loop 3
C
SaCas9
RuvC cut site
gDNA
NNNNNNNNNNNNNNNNNNNNN
NNGRRT
HNH cut site
NNCYYA
Target (~21nt)
Tetraloop
sgRNA
Guide
Stem loop 1
Stem loop 2

Figure 1.2 SpCas9 and SaCas9 primary protein sequence and complex with sgRNA and genomic target site.

(**A**) Schematic of primary protein domain structures for SpCas9 (1,368 amino acids) and SaCas9 (1,053 amino acids). Both Cas9 proteins are bilobed, consisting of the nuclease (NUC) lobe with three RuvC motifs (RuvC-I, RuvC-II, and RuvC-III) and the HNH domain, as well as the recognition (REC) lobe. The REC lobe of SpCas9 is divided further into four subdomains. BH, bridge helix; L1 and L2, linker 1 and 2 regions; PLL, phosphate lock loop; WED, wedge domain; PI, PAM interacting domain; TOPO, topoisomerase homology domain; CTD, C-terminal domain. Amino acid start positions of each domain are indicated. (**B** and **C**) Schematics of (**B**) SpCas9 and (**C**) SaCas9 in complex with their cognate single guide RNAs (sgRNAs), targeted to a genomic DNA (gDNA) target site. The target is defined by a variable guide sequence within the sgRNA that is typically ~17-18 nt for SpCas9 or ~21-22 nt for SaCas9. Protospacer-adjacent motif (PAM) sequences are located on the non-target strand and have the motif 5′-NGG-3′ for SpCas9 and 5′-NNGRRT-3′ for SaCas9, which are highlighted in yellow (N indicates any nucleotide, and R is guanine or adenine). RuvC and HNH cleavage domain cut sites are indicated by arrowheads on the non-target and target strands respectively (3 nt upstream of the PAM sequence). The invariant sgRNA scaffold sequences (highlighted in white) are comprised of a tetraloop region and either three or two stem loops for SpCas9 and SaCas9, respectively.

1.4.3.1 Dual-cut-mediated exon excision

Multiplex gene editing is possible, whereby multiple target sites are edited simultaneously through the provision of two or more sgRNAs [172,193]. A dual sgRNA strategy can be used to excise a genomic region containing an exon, thereby achieving permanent 'exon skipping' at the DNA level. To this end, two sgRNAs are used that target introns flanking the target exon(s) of interest. In some cases, the intervening sequence will be excised, and the cleaved genomic ends of the flanking regions are joined by the host cell NHEJ DNA repair pathway. Subsequently, the edited genomic DNA now lacks the excised exon, such that when the locus is transcribed, the upstream and downstream exons will be joined by the spliceosome. Multiple studies have demonstrated the utility of the exon excision strategy *in vivo* by dual sgRNA-mediated targeting of the introns flanking the *mdx* mouse *Dmd* exon 23 (which harbours a nonsense mutation leading to premature termination of dystrophin translation) [181,190–192,194] (**Figure 1.3A**). A similar study has applied this approach to the removal of *Dmd* exons 52 and 53 in the *mdx*^{4cv} murine model [195] (**Figure 1.3B**). Every new *DMD* transcript generated from the successfully edited locus will lack the targeted exon, leading to correction of the translation reading frame, and production of a functional, albeit truncated, dystrophin protein [181,190–192,194,195] (**Figure 1.3**). Repair by the NHEJ pathway typically results in indel (insertion/deletion) formation at the resolved DSB site. These indels are not expected to have a detrimental effect on the *DMD* open reading frame as they occur within the non-coding recombinant intron, provided that a cryptic splice site is not introduced.

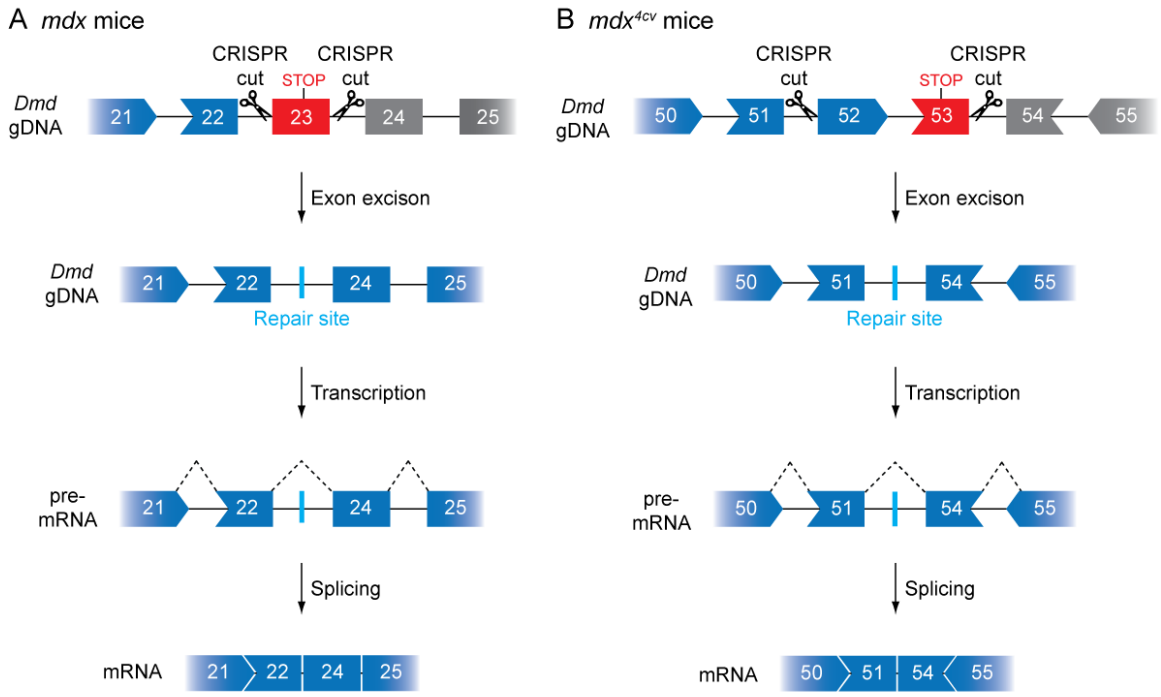


Figure 1.3 CRISPR/Cas9 dual-cut-mediated exon excision for *Dmd* reading frame restoration.

The CRISPR systems introduces double strand breaks (DSBs) at the cut sites and the intervening region is removed. Small indels (insertions and/or deletions, shown in light blue) are typically introduced at the repair site by the error-prone non-homologous end joining (NHEJ) DNA repair pathway. Every transcript derived from the edited locus will lack the excised exon(s), and the reading frame of the mature messenger RNA (mRNA) is thereby restored. The dual-cut CRISPR/Cas9 approach has been applied for removal of (**A**) exon 23 in the *mdx* mouse [181,190–192,194], and (**B**) exons 52 and 53 in the *mdx*^{4cv} mouse [195]. In-frame exons are in blue, out-of-frame exons are in grey, and exons containing a premature stop codon (i.e. DMD-causing) are in red.

1.4.3.2 Single-cut-mediated exon reframing and splice signal disruption

An alternative to excising whole exons is the use of a single-cut strategy to disrupt splice signal sequences, such as 3′ splice acceptors (SAs) and exonic splicing enhancers (ESEs) (**Figure 1.4**). Cas9 targeting of a PAM site adjacent to an SA or ESE site, results in indel formation via the NHEJ DNA repair pathway. These indels can ablate the neighbouring splice signal sequence, such that it is no longer recognised by the spliceosome, leading to permanent exon skipping. Alternatively, indel formation occurring within a mutation-containing exon has the potential for exon reframing and open reading frame restoration, depending on the size of the indel. In instances where the DSB is repaired in a 'scarless' manner (i.e. with no indel), the guide recognition site is reconstituted, and the CRISPR complex is able to induce cleavage at the same site again. However, a disadvantage of this approach is that where an indel does not restore the reading frame, the sgRNA target site will be lost and the *DMD* gene edited in a non-productive manner.

The single-cut strategy for splice signal disruption and exon reframing has been assessed for targeting of a putative ESE in exon 51 in ΔExon50 mice [196] (**Figure 1.4A**). Approximately half of all indels introduced at the target site consisted of a single adenosine insertion (which would accomplish productive reading frame restoration) [196]. The remainder of the indels consisted of deletions with a range of sizes spanning the putative ESE site and, in some cases, also encompassing the proximal SA site of exon 51, resulting in the further promotion of productive exon 51 skipping. Widespread dystrophin expression was observed in the heart, diaphragm and peripheral skeletal muscles after treatment, increasing between 4- and 8-weeks post injection [196]. Expanding on these promising murine studies, the single-cut system was further tested in a ΔExon50 canine model of DMD (**Figure 1.4A**), with only a single nucleotide difference in the sgRNA target site relative to the mouse model [197]. The single-cut approach has also been applied to exon 45 in the ΔExon44 murine model [198,199].

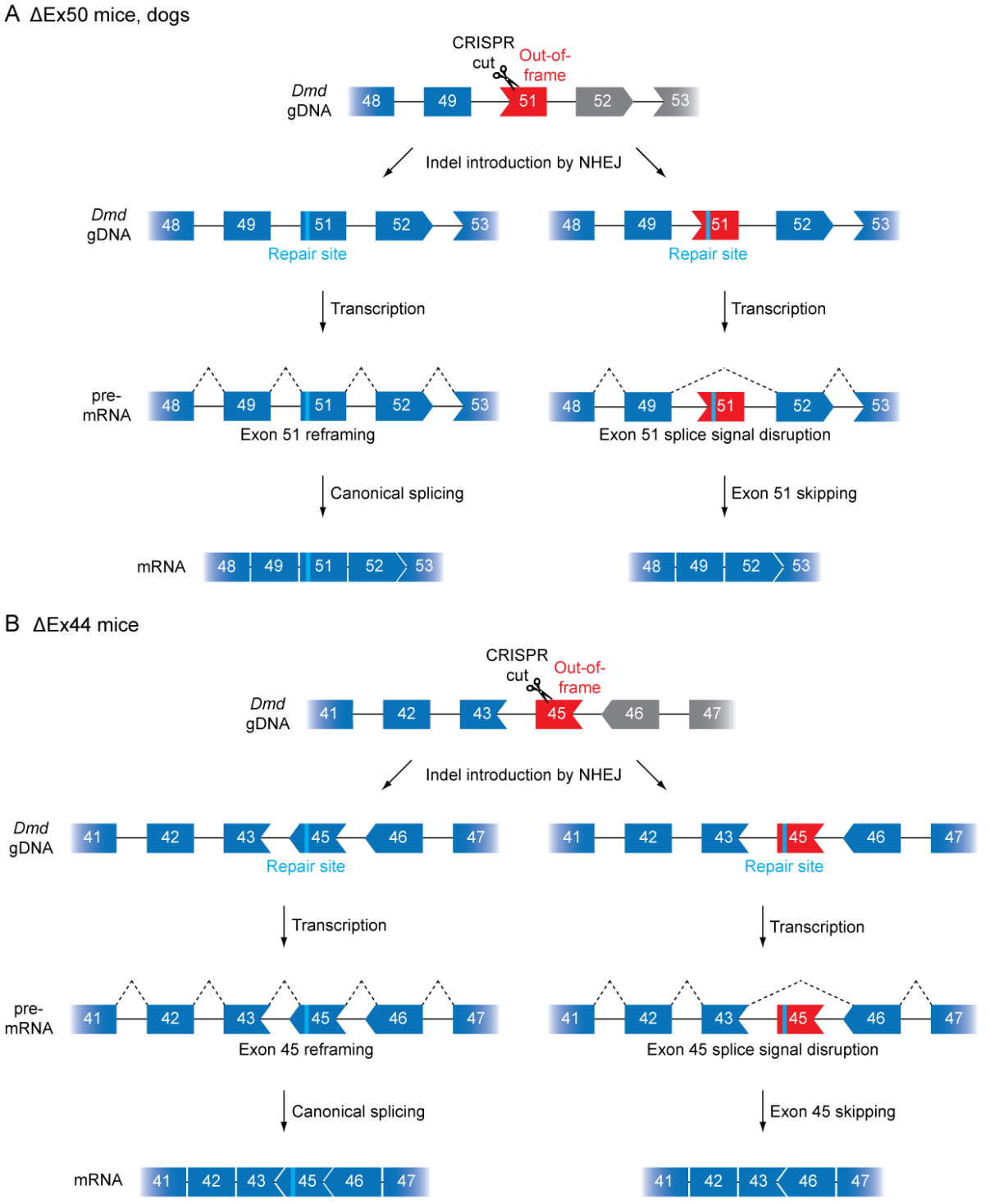

A ΔEx50 mice, dogs
CRISPR cut
Out-of-frame
Dmd gDNA
48 49 51 52 53
Indel introduction by NHEJ
Repair site
Transcription
pre-mRNA
Exon 51 reframing
Exon 51 splice signal disruption
Canonical splicing
Exon 51 skipping
mRNA
B ΔEx44 mice
Exon 45 reframing
Exon 45 splice signal disruption
Exon 45 skipping

Figure 1.4 CRISPR/Cas9 single-cut-mediated reframing or splice signal disruption for *Dmd* reading frame restoration.

CRISPR/Cas9 cleavage at the 5′ end of an exon introduces indels that either reframe the exon, or disrupt splice signal sequences (such as the splice acceptor (SA) or exonic splice enhancer (ESE)). Every transcript derived from the reframed or splice signal edited locus will retain the target exon, and the reading frame of the mature messenger RNA (mRNA) is restored by the reframing indels or exon skipping respectively. The single-cut approach to exon reframing and/or splice signal sequence disruption has been applied to (**A**) *Dmd* exon 51 in ΔExon50 mice [196] and dogs [197], and (**B**) *Dmd* exon 45 in ΔExon44 mice [198,199]. Out-of-frame exons are indicated in grey, DMD-causing exons in red, and repair indels in light blue.

1.4.3.3 Homology directed repair

An alternative DMD therapeutic approach involves CRISPR/Cas9-mediated targeting adjacent to an exonic point mutation, enabling seamless correction via the HDR pathway [195,200,201] (**Figure 1.5**). HDR requires that a single-stranded oligodeoxynucleotide (ssODN) repair template, with ~90 nt of sequence homology on either side of the cut site, be provided together with the Cas9 and sgRNA components. It is possible to supply the ODN template within an AAV genome as is typical for delivery of CRISPR components *in vivo*. HDR offers the potential to make precise, templated corrections to small genetic mutations. However, this repair pathway is only active during the late S and G2 phases of the cell cycle, whereas NHEJ occurs in all cells, regardless of cell cycle status [202–204]. Given that the relevant cell populations for therapeutic targeting in skeletal muscle are myofibres (which are post-mitotic) and satellite cells (resident muscle stem cells responsible for regeneration upon myofibre damage, which remain quiescent until activated), the utility of HDR approaches for treating DMD patients is likely to be very limited.

HDR has been tested for correction of the nonsense mutation within exon 53 in *mdx4cv* mice, although this approach was designed in such a way that alternative repair pathways could also restore dystrophin expression should HDR be unsuccessful [195] (**Figure 1.5A**). Dual sgRNAs flanking the nonsense mutation, and in close proximity to one another, enabled a partial in-frame deletion of the intervening exonic region. Targeted deep sequencing revealed that a total of 2.3% of target sites analysed were edited, and of these only 0.18% were seamlessly corrected via HDR [195]. As such, the vast majority of edits were a result of NHEJ repair, with most being non-productive indels. The overall editing efficiency was up to 4-fold lower than that achieved by the full exon 52 and 53 dual sgRNA excision approach (discussed above) [195]. This highlights the limited utility of HDR-mediated editing in post-mitotic cells. Zhang and colleagues sought to circumvent the limitation of the minimal HDR activity in post-

mitotic tissue by treating *mdx* zygotes [200]. The authors employed an alternative CRISPR system called Cpf1 (CRISPR from *Prevotella* and *Francisella*) derived from *Lachnospiraceae bacterium* (LbCpf1) [205–208] [121–124] to assess a single-cut-mediated HDR strategy, with the added potential for exon reframing (in the context of the *Dmd* exon 23 mutation, **Figure 1.5B**) [200]. LbCpf1 exhibits robust DNA cleavage activity in mammalian cells and in mice, and recognises a T-rich PAM sequence (5′- TTTV, where V is A, C or G). This is an attractive choice of CRISPR system as SA regions typically exist within a pyrimidine-rich sequence context. Furthermore, LbCpf1 is smaller than SpCas9 (i.e. 3.7 kb), and is thus more suitable for packaging in AAV vectors. Unlike SpCas9 and SaCas9 which generate blunt incisions, LbCpf1 generates a staggered cut which presents new possibilities for gene editing through the activation of alternative DNA repair pathways to NHEJ and HDR. Five of the twenty-four resulting offspring derived from the zygotes treated with the LbCpf1 CRISPR components carried corrected alleles, with levels of dystrophin restoration ranging from 2–100%. This mosaicism suggests a high level of variability in CRISPR editing efficiency, even when CRISPR-mediated gene editing was performed at the zygote stage [200].

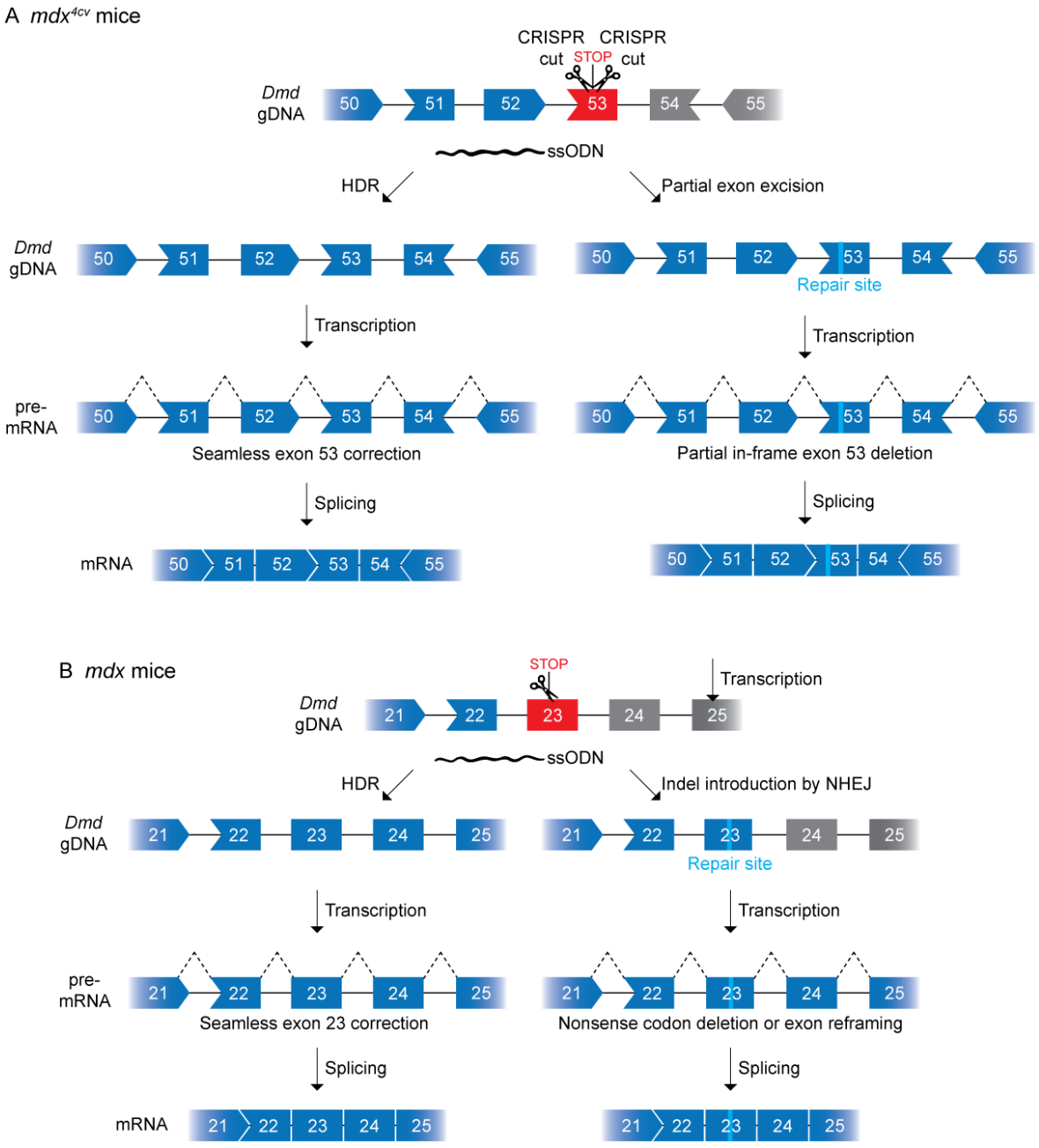

A mdx4cv mice
CRISPR cut
STOP
CRISPR cut
Dmd gDNA
50
51
52
53
54
55
ssODN
HDR
Partial exon excision
Repair site
Transcription
pre-mRNA
Seamless exon 53 correction
Partial in-frame exon 53 deletion
Splicing
mRNA
B mdx mice
21
22
23
24
25
Indel introduction by NHEJ
Seamless exon 23 correction
Nonsense codon deletion or exon reframing

Figure 1.5 CRISPR/Cas9-mediated homology directed repair for *Dmd* reading frame restoration.

(**A**) CRISPR-mediated targeting by dual sgRNAs flanking a nonsense mutation within *Dmd* exon 53, and the provision of a single-stranded oligodeoxynucleotide (ssODN) repair template consisting of the desired edited sequence, facilitates seamless homology directed repair (HDR) in *mdx4cv* mice [195]. (**B**) CRISPR-mediated targeting by a single gRNA adjacent to the nonsense mutation within *Dmd* exon 23, and the provision of a ssODN repair template, facilitates seamless HDR in *mdx* mice [200,201]. The HDR-corrected loci will encode a full-length wild-type *Dmd* transcript, thereby restoring the dystrophin reading frame. Should HDR be unsuccessful, an alternative productive editing outcome is possible whereby indels introduced at the CRISPR cut site by NHEJ DNA repair can either delete the nonsense codon, or reframe the exon, resulting in the production of a full-length transcript containing an in-frame edit within the target exon. Out-of-frame exons are indicated in grey, and DMD-causing exons in red, and repair indels in light blue.

1.4.3.4 Genomic DNA base editing

Nonsense mutations have been predicted to account for 11% of all mutations that cause human inherited disease [209], and account for a relatively large proportion of DMD-causing lesions (estimates range from ~10-30%) [140,210–213]. Base editing at both the genomic DNA and mRNA levels is a more recent addition to the CRISPR toolbox, providing a means for genetic correction of DMD SNPs, such as nonsense mutations, in a manner that is highly specific to the desired genomic target. The CRISPR DNA base editor complex facilitates nucleotide conversion at a desired genomic DNA target through the fusion of a catalytically impaired Cas9 variant (a 'nickase' Cas9, nCas9) and a deaminase enzyme. The first generation of DNA base editors to be developed employed a cytidine deaminase (initially derived from rat APOBEC1, although there are now various other versions) which facilitates the targeted conversion of specific cytidine bases to uridine within a defined target window (i.e. ~14-17 nt upstream of the PAM site) [214,215]. The uridine lesion can then be replaced with thymidine by the DNA replication machinery, or repaired via the base excision or mismatch repair pathways, in either case resulting in a C:G to T:A substitution. Subsequently, a second base editing system was developed: ABE (adenosine deaminase base editor), an RNA-programmable deaminase (derived from *Escherichia coli* TadA) fused to nCas9 [216]. The adenosine on the opposite strand to the 'nick' is converted to inosine which is either read as guanosine by the transcription machinery, or is replaced with guanosine by the base excision repair pathway, enabling an A:T to G:C base change. Several further advantages of base editing systems over the more conventional DSB-inducing CRISPR systems described above, are that they do not require a donor template, and only a single stranded 'nick' is introduced into the target DNA. Genomic 'nicks' are repaired via the high-fidelity base excision repair pathway [217]. This is advantageous over CRISPR systems that introduce DSBs where cleavage at unwanted off-target sequences

has the potential for indel formation by NHEJ, in addition to the possibility of major chromosomal rearrangements [193,218].

Ryu and colleagues employed the current generation of the ABE system (ABE7.10) to convert a nonsense mutation (from TAG to CAG) in exon 20 of a murine model of DMD [219] (**Figure 1.6**). In order to package the ABE and cognate sgRNA sequences (which are >6 kb in total) for viral delivery, the authors utilised a trans-splicing AAV (tsAAV) vector strategy that enables delivery of the ABE across two separate AAV vectors (the ABE is split into N- and C-terminal fragments and the targeting complex is reconstituted following transduction) [220,221]. The authors observed base editing activity of the target cytidine with an efficiency of ~3.3%, which resulted in ~17% dystrophin restoration as determined by immunohistochemistry (IHC) staining 8-weeks post treatment. This level of correction is less than that achieved by the other NHEJ-based strategies described above, which could be a result of either the low efficiency of the tsAAV trans-splicing system and/or the low activity of the ABE itself. A further limitation of this approach is that, in some cases, there may be no appropriate PAM sites within the optimal base editing window. Ryu and colleagues tested longer sgRNAs (~20-23 nt) which were found to expand the editing window by 3-4 nt (i.e. to 12-19 nt upstream of the PAM site). However, this expanded base editing window also increases the risk of unintended bystander editing events in close proximity to the desired target.

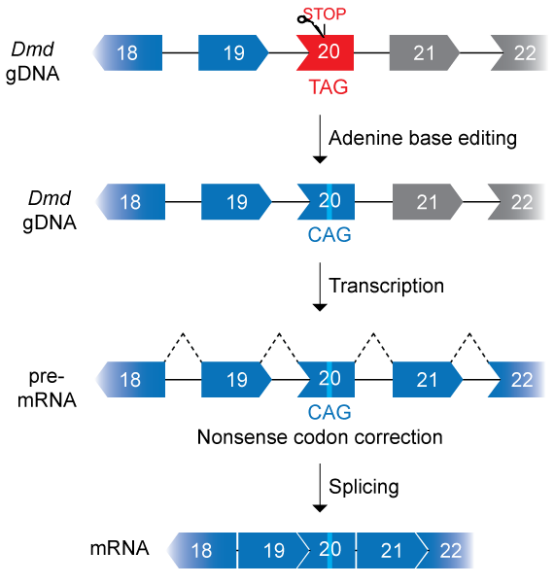


Figure 1.6 CRISPR-mediated base editing for *Dmd* reading frame restoration.

A CRISPR base editing system comprised of a catalytically impaired Cas9 that 'nicks' one strand of the target DNA (nCas9), fused to an adenosine deaminase enzyme, is directed to introduce a single-stranded break (SSB) adjacent to a nonsense mutation in *Dmd* exon 20 in a murine model of dystrophy [219]. The adenosine on the opposite strand to the 'nick' is converted to inosine by the deaminase component, and is either read as guanosine by the transcription machinery, or replaced with guanosine by the base excision repair pathway. This results in conversion of the TAG nonsense codon to a CAG missense codon, leading to dystrophin reading frame restoration. Out-of-frame exons are indicated in grey, and DMD-causing exons in red, and repair indels in light blue.

1.4.3.5 RNA editing

Transcriptome engineering using base editors that target mRNA species has several advantages over genomic DNA targeting systems. These include the transient nature of editing such that any unwanted bystander edits created will not permanently disrupt protein expression at the target site, and that the system is effectively 'PAM-less' [222]. Two orthologues of ADAR are expressed naturally in human cells (i.e. ADAR1 and ADAR2) and function to carry out endogenous RNA editing by the hydrolytic deamination of adenosine to inosine (A>I), where inosine is read as guanosine by both the splicing as well as translational machinery [223,224].

CRISPR-based RNA editors make use of a two-part fusion protein (i.e. a Cas protein component that recognises a double stranded RNA target, and an adenosine deaminase acting on RNA (ADAR) domain) in complex with a CRISPR RNA (crRNA) guide sequence to achieve A>I conversion [222]. The targeting protein component that has been engineered for use in mammalian cells is Cas13b, derived from *Prevotella sp* (i.e. PspCas13b) [222]. Cas13 proteins have two HEPN (higher eukaryotes and prokaryotes nucleotide-binding) endoRNAse domains that function in the natural bacterial systems to cleave RNA targets that are adjacent to a protospacer flanking site (PFS) in a highly precise manner [225,226]. The PFS is analogous to the PAM required for Cas9 activity. A catalytically inactive ('dead') version of this protein (i.e. dCas13b) has been developed for use in the RNA editing system so that the target RNA is not cleaved, and because the target recognition and engagement step is independent of PFS recognition [222]. ADAR consists of an N-terminal double stranded RNA binding domain, and a C-terminal deaminase domain. The crRNA hybridises to the mRNA transcript thereby generating a dsRNA ADAR substrate [227,228]. The crRNA is typically comprised of ~30-50 nt of targeting RNA transcriptionally fused to a Cas13 orthologue-specific scaffold sequence [222]. dCas13b in complex with its cognate crRNA directs ADAR to the target adenosine for RNA editing. Crucially, A>I editing at the desired nucleotide is facilitated by the inclusion of a

mismatched cytosine with the target adenosine [229]. Engineering of the CRISPR-based RNA editing system has resulted in the development of PspCas13b fused with a hyper-functional ADAR mutant domain (i.e. ADAR2-E488Q). This RNA editing approach has been applied to correct G>A SNPs (i.e. these SNPs introduce premature stop codons by altering the tryptophan TGG codon to either a TGA or TAG stop codon) in the dystrophin gene in mammalian cell culture [222] and *mdx* mice [230].

1.5 The skeletal muscle secretome

Muscle homeostasis is maintained by a number of factors that control satellite cell activation and regeneration, including a family of miRNAs [231,232] that are selectively released during postnatal development, after physical exercise, and as a result of muscle injury [233,234]. miRNAs are typically 22 nt in length, and regulate gene expression in a post-transcriptional manner by the direct interaction with mRNA transcripts [235]. These molecules are therefore important endogenous effectors of the cellular RNA interference (RNAi) pathway. RNAi can also be achieved by exogenously produced molecules known as small interfering RNA (siRNA) and short hairpin RNA (shRNA) which mimic the structure of miRNAs at different stages in their biogenesis pathway. The muscle-enriched miRNAs are referred to as myomiRs, and include miR-1, miR-133a and miR-206 [236,237]. myomiRs have been found to be elevated in plasma from DMD patients, as well as dystrophin-deficient animal models including *mdx* [233,234,238,239], *mdx*^{4cv} [240], and dKO mice [241], and the CXMD$_J$ [238] and GRMD [242] dystrophic dog models. Evidence for selectivity of myomiR release exists as small RNA sequencing revealed that there is an incongruence between the circulating levels of highly and lowly abundant cellular miRNAs [243]. Levels of serum myomiRs have also been shown to vary over time in cell culture [234] and *in vivo* [233,234]. An increase in serum myomiR abundance occurs in early postnatal mice suggesting involvement in normal muscle growth, both in WT and *mdx* mice [234]. The selective release of myomiRs from skeletal muscle tissue into circulation, independent of passive

leakage from perturbations to the sarcolemma in dystrophic disease [234,244,245], is suggestive of a mechanism involving the functional transfer of these molecules to recipient cells in a feed-forward, paracrine manner [234].

Exon skipping-mediated dystrophin restoration by treatment with PPMO [239], U1-snRNAs [246] or U7-snRNAs [241] was shown to restore the expression of extracellular myomiRs (ex-myomiRs) to WT levels in murine DMD models. BMD patients have been found to express intermediate levels of ex-myomiRs, relative to DMD patients and healthy individuals and, as such, these biomolecules reflect the severity and progression of dystrophinopathy [233,234,246–248]. This family of circulating myomiRs, also referred to as the dystromiRs [246,247], have thus been identified as important potential pharmacodynamic biomarkers of DMD which could enable non-invasive determination of therapeutic efficacy and disease progression [234,244,246]. Elevated levels of certain dystromiRs is not believed to be pathogenic, and instead a response to the lack of dystrophin, functioning to preserve the muscle architecture in an anti-fibrotic and pro-myogenic manner [249,250]. Specifically, miR-206 has been shown to delay the progression of DMD in dystrophic mice by inhibiting various negative regulators of myogenesis, thereby promoting skeletal muscle regeneration [249]. Intriguingly, circulating myomiRs are remarkably stable, even within the harsh conditions of the extracellular environment (i.e. the presence of proteases/nucleases) [234,251]. Extracellular vesicles (EVs) are believed to protect ex-myomiRs from nuclease degradation and play a role in the functional transfer of these factors to recipient cells [234,251,252]. However, only a small proportion of the total ex-myomiRs are contained within EVs, and the remainder is thought to associate with non-vesicular lipoprotein [253,254] or protein complexes (e.g. miRNAs have been detected in complexes containing extracellular Argonaute-2, AGO2) [255,256].

The absence of dystrophin in DMD results in a characteristic secretory profile that exemplifies both active and passive mechanisms of molecular release from damaged and

regenerating muscle fibres [84,243]. Cell therapy is an important avenue for DMD therapy development that is independent of patient mutation type. The premise of this approach is that healthy heterologous or autologous myogenic progenitor cells are transplanted into dystrophic muscle tissue in an effort to reconstitute the pool of muscle stem cells with those that possess full regenerative potential (this approach has been reviewed extensively elsewhere [257]). The vast expanse of skeletal muscle tissue in the human body, as well as difficulty in accessing the heart and diaphragm muscles (i.e. two important therapeutic targets for DMD pathology alleviation) makes stem cell therapy a technically challenging therapeutic option. Instead, cell-free methods involving EVs for the delivery of an array of bioactive signalling molecules is also being investigated in the context of DMD therapy development.

1.6 The role of EVs in cellular communication

EVs are nano-sized membrane-bound particles that are naturally secreted by all cell types within the body, mediating cell-cell crosstalk through the transfer and delivery of an array of different types of biomolecular cargo [251,258,259]. EVs were first identified in 1967, but were referred to as 'platelet dust' for lack of a better understanding [260]. Today, EVs are known to be a highly diverse and heterogenous pool of molecules that are classified based on their biogenesis, secretory pathway, and size. Critically, although EVs are encased in a lipid bilayer membrane and are cell-like, they are non-replicative as they do not possess a nucleus. Classification of EVs into two major classes has recently been proposed. These include the microvesicles which originate from the blebbing or exocytosis of the plasma membrane, and exosomes which arise from multi-vesicular bodies (MVBs) [261]. The largest types of EVs are the microvesicles, which include apoptotic bodies and oncosomes [262]and range in size from ~50-5,000 nm [261,263,264] (**Figure 1.7**). Apoptotic bodies are otherwise known as shedding particles as the belief has been that these particles are passively released by dead and dying cells, however, recent evidence suggests that they may be involved in important cellular

processes of healthy cells [265]. Exosomes are a well-characterised group of actively released nano-sized particles that range in size from ~30-150 nm [263,264]. The term exosome was coined as the mechanism of release is akin to 'reverse endocytosis' [266]. While these two major classes of EVs differ significantly with respect to biogenesis, secretion, and size, it is technically challenging to separate them as they possess overlapping biophysical properties [267]. The use of specific vesicle classification in the literature has been discouraged by the guidelines released by the International Society for Extracellular Vesicles (ISEV) as many multiple mechanisms are involved in EV biogenesis and release, much of which still remains relatively unknown [267]. ISEV instead encourages the use of the inclusive term 'extracellular vesicles' where the exact mechanism of biogenesis and secretion is unknown, to avoid misconceptions and contradictory definitions. Furthermore, EV-TRACK is a platform that has recently been launched to standardise EV-based research by the consolidation of findings relating to EV biology and methodology [268]. While EV research guidelines help to standardise the research field in general, new discoveries are being constantly made revealing aspects of EVs that may not easily fall into the rigid classification system currently proposed. Nonetheless, the biogenesis and secretion of exosomes is the most well-characterised of the EV group of biomolecules, and will be described in more detail.

1.6.1 Exosome biogenesis and secretion

Exosome production is a process of the endo-lysosomal pathway which occurs in three major steps: biogenesis, transport, and release. Briefly, early endosomes are formed by the inward budding (i.e. endocytosis) of the plasma membrane and thereafter fuse to create late endosomes [269] (**Figure 1.7**). Invaginations of the late endosomal membrane create intraluminal vesicles (ILVs), and ultimately the MVB which is trafficked to the plasma membrane for ILV released into the extracellular milieu by fusion of the plasma and MVB membranes [270,271]. Conversely,

MVBs can be taken up by intracellular lysosomes or autophagosomes and the ILV contents degraded [261] (**Figure 1.7**).

ILV formation occurs by two independent pathways which are compensatory to a certain extent. The endosomal sorting complexes required for transport (ESCRT) machinery (ESCRT-0, -I, -II, -III, and -IV in complex with the accessory ATPase VPS4 complex) are believed to be the classical mode of exosome biogenesis [270] (**Figure 1.7**). ESCRT proteins are involved in a number of membrane shaping and remodelling processes in the cell including, but not limited to, ILV biogenesis, protein sorting into ILVs, and vesicle budding at the plasma membrane. The exact mechanism of loading of cellular cargo into EVs is still relatively poorly understood, although some proteins are known to play a pivotal role. These include components and accessory proteins of the ESCRT machinery (e.g. tumour susceptibility gene 101 (TSG101), and programmed cell death 6 interacting protein (PDCD6IP, also known as ALIX)) [270].

The second major ILV production pathway is dependent on the activity of lipid ceramide and the protein required for ceramide production, neutral sphingomyelinase 2 (SMPD2, otherwise known as NSMase2) [270] (**Figure 1.7**). The involvement of this pathway in EV biogenesis was first proposed based on findings that ceramide participates in ILV formation in liposomes *in vitro* [272]. This hypothesis was confirmed in a study in which EV release was inhibited following inhibition of NSMase2 in cancer cells (following treatment with GW4869, a small molecule inhibitor of NSMase2) [273,274], or increased following ceramide supplementation [275]. Other proteins involved in ESCRT-independent exosome biogenesis are members of the large family of tetraspanins (i.e. mainly CD9, CD63, and CD81) which are also highly enriched in secreted exosomes [261,274]. Tetraspanins are four-transmembrane-domain proteins believed to play a variety of roles in exosome biogenesis, including ILV budding by direct effects on late endosomal membrane curvature, exosome cargo loading through

interactions with accessory proteins, as well as in exosome targeting and uptake by recipient cells post secretion [276].

Trafficking and docking of MVBs to the membrane involves the co-ordinated action of the cytoskeletal machinery and small RAB GTPases [261,277] (**Figure 1.7**). This is a process which is believed to occur in a polarised manner to direct MVBs to particular secretory sites at the plasma membrane [278]. RAB11 was the first protein determined to be involved in EV trafficking and secretion [279]. Several other RAB GTPases have also been implicated in this process in RNAi knockdown studies (e.g. RAB2B, RAB5, RAB9A, RAB27A, and RAB27B were found to be important in a study carried out in HeLa cells [280], and RAB7 in MCF7 breast cancer cells [281]). RAB GTPases function downstream of the molecular decision between targeting the MVB to the lysosome for degradation, or the plasma membrane for vesicle release.

The final step of exosome secretion is through MVB fusion with the plasma membrane which involves various SNARE (soluble NSF attachment protein receptor) proteins [271,282] (**Figure 1.7**). Exosome release has been shown to be increased by the inhibition of MVB fusion with autophagosomes [283,284], however, the exact mechanisms that determine the balance between ILV cargo degradation by the lysosome or autophagosome, and release into the extracellular environment are still somewhat unknown.

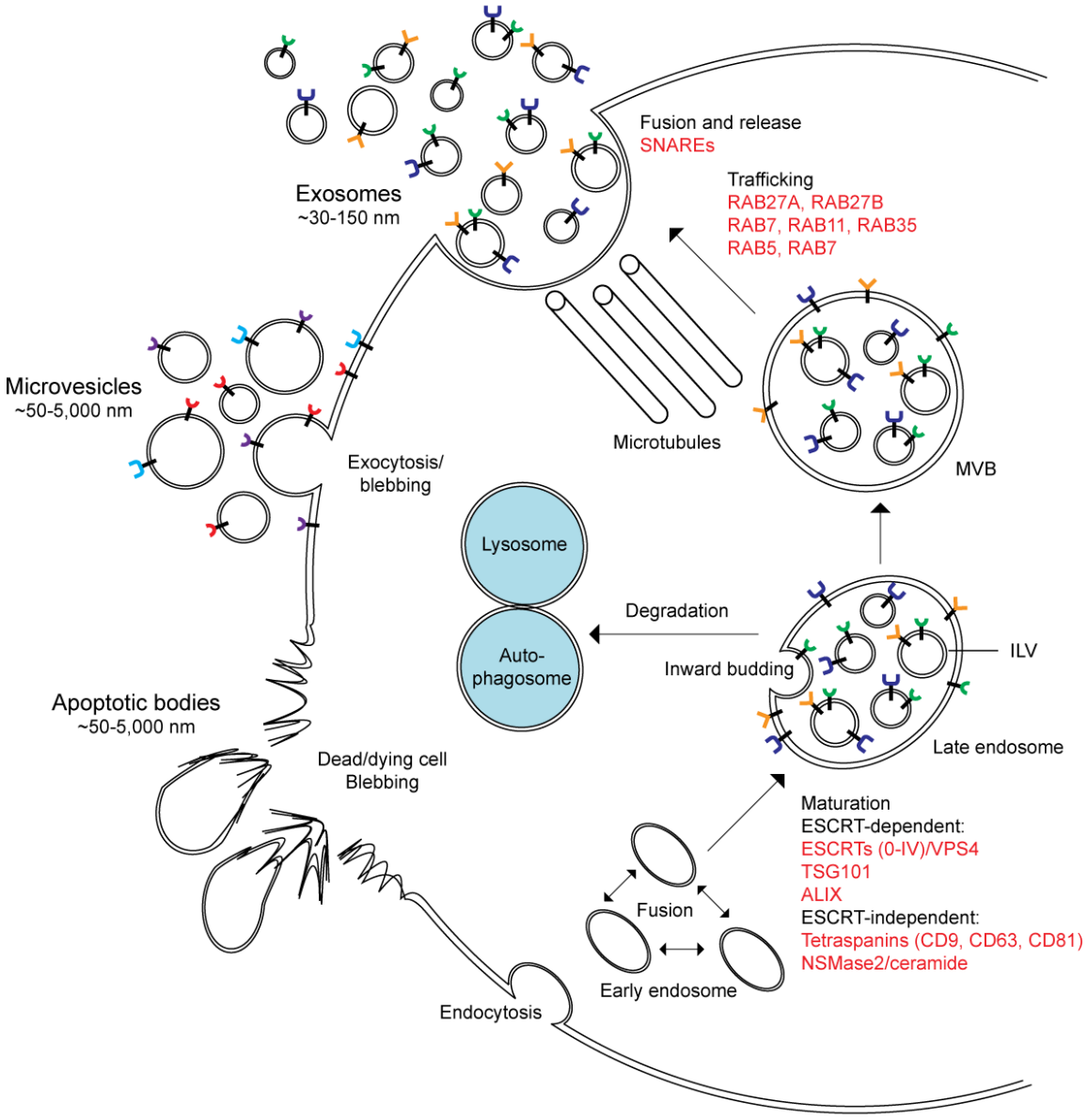

Exosomes
~30-150 nm
Fusion and release
SNAREs
Trafficking
RAB27A, RAB27B
RAB7, RAB11, RAB35
RAB5, RAB7
Microvesicles
~50-5,000 nm
Microtubules
MVB
Exocytosis/
blebbing
Lysosome
Auto-
phagosome
Degradation
Inward budding
ILV
Late endosome
Apoptotic bodies
~50-5,000 nm
Dead/dying cell
Blebbing
Maturation
ESCRT-dependent:
ESCRTs (0-IV)/VPS4
TSG101
ALIX
ESCRT-independent:
Tetraspanins (CD9, CD63, CD81)
NSMase2/ceramide
Fusion
Early endosome
Endocytosis

Figure 1.7 Extracellular vesicle biogenesis.

Mechanisms of extracellular vesicle (EV) release and the exosome biogenesis pathway. Apoptotic bodies and microvesicles are ~50-5,000 nm in size, and are both actively and passively released from cells by blebbing or exocytosis. Exosomes are ~30-150 nm in size and are actively released by healthy cells. Exosome biogenesis involves the inward budding of the plasma membrane to form early endosomes. These fuse to form the late endosome, and inward budding of the membrane generates the intraluminal vesicles (ILVs) which are actively and selectively loaded with cellular-derived cargo (i.e. lipids, proteins, nucleic acids) in a process that is either dependent on the ESCRT (endosomal sorting complexes required for transport) machinery, or the activity of neutral sphingomyelinase and ceramide. The late endosome matures into the multi-vesicular body (MVB) which is trafficked to the plasma membrane via the cellular microtubule network for fusion and release of the exosomes (requiring the activity of SNARE/ soluble NSF attachment protein receptor proteins). Proteins involved in the various stages of exosome biogenesis are indicated in red.

1.6.2 Therapeutic exploitation of EVs

EVs play a pivotal role in both physiological and pathological processes through the ability to transport and deliver lipids, proteins, and nucleic acids between cells and tissues [259]. Indeed, exosomes have been directly implicated in tumorigenesis and metastasis [273,285–287], as well as in neurodegenerative diseases including Alzheimer's [288–291], Parkinson's [292], and amyotrophic lateral sclerosis [293]. Furthermore, several studies have highlighted the role of exosomes in signalling within the muscle niche [245], and have demonstrated that this signalling process is highly dysregulated in muscular dystrophy [244,294]. Ongoing efforts are being made to improve methods of delivery of RNA-based genetic therapies, such as siRNAs and ASOs, particularly using synthetic biomolecules that avoid the effects of immunisation observed with viral vectors (which limits repeat administration). Synthetic lipid nanoparticle formulations have been an attractive alternative to viral delivery of nucleic acid genetic therapies for a number of years [295]. Alternatively, strategies that exploit the physical and biological properties of EVs have also emerged as highly attractive therapeutic delivery modalities that are better tolerated by the immune system than viral vector-based systems. EVs have the added benefits over lipid nanoparticles of improved biodistribution profiles, and the possibility to leverage cellular processes for cargo loading and surface receptor engineering [296].

EVs are being investigated for their use as therapeutic delivery vehicles with the potential to act as a decoy for key players of the immune system (such as interleukins), thereby modulating the inflammatory response [297]. Oncolytic viruses have also been 'hidden' in EVs to avoid detection and neutralisation of viral antigens by the immune system, and to facilitate improved targeting and uptake into cancer cells [298]. Additionally, active recruitment of Cas9 into extracellular vesicles has been demonstrated by a novel and innovative technological approach, coined NanoMEDIC, for CRISPR-mediated DMD correction in patient-derived iPSCs [299].

The way in which EVs carry out cell-cell communication makes them attractive candidates for the development of regenerative medicine therapeutic strategies. Particular cell types, such as mesenchymal stem/stromal cells, have been shown to express EVs with inherent anti-inflammatory and pro-regenerative properties [300–305]. Additionally, several studies have investigated the biomolecular composition of EVs isolated from myocytes in culture [245,306–309], or from murine skeletal muscle [307], and the role of EVs in transferring pro-myogenic signals *in vitro* [306–309], or *in vivo* [308,309] (**Table 1.2**). As such, extensive research efforts have been made over the past few decades to develop exosome-based tools for regenerative nanomedicine [310], although no FDA-approved therapies are available to date.

Table 1.2 Studies investigating the role of muscle cell EVs on myogenic differentiation and methods of EV isolation employed

First author	Findings	Methods	Ref
Duguez 2013	Intact dystrophic myofibres exhibit general dysregulation of EV production, demonstrating a link between dystrophin and the regulation of vesicular trafficking. Treatment with PMO for *Dmd* exon 23 skipping resulted in 30% dystrophin restoration and reversed the perturbed secretome profile.	Ultracentrifugation using a sucrose gradient to purify EVs from *mdx* and WT murine H-2K myoblasts differentiated into myotubes *in vitro*.	245
Forterre 2014	miRNAs are selectively incorporated into exosomes during myogenic differentiation.	Ultracentrifugation to purify C2C12 myoblast and myotube EVs.	311
Forterre 2014	EVs derived from C2C12 myotubes reduced cell proliferation (through downregulation of cyclin D1) and enhanced myogenic differentiation (through upregulation of myogenin), relative to myoblast-derived EVs.	Ultracentrifugation to purify EVs from C2C12 myoblasts and myotubes grown in media containing pre-spun serum. Treated C2C12 myoblasts with 0.25-2 µg myotube-derived EVs.	306
Nakamura 2015	Mesenchymal stem cell (MSC)-derived EVs enhanced proliferation and fusion index of C2C12 myotubes, and promoted migration of human umbilical-vein endothelial cells (HUVECs). MSC-derived EVs accelerated wound repair, enhanced angiogenesis, and reduced fibrosis following cardiotoxin treatment in the TA of mice. myomiRs were detected in MSC-derived EVs and the local injection of these accelerated wound repair in a rat model of skeletal muscle injury.	Ultracentrifugation to purify MSC EVs. Treated C2C12 myoblasts with MSC CM-containing or -depleted of EVs (no dose reported). Treated HUVECs with MSC-EVs in a trans-well experiment and assessed cell migration using a tube formation assay. Treated mice with cardiotoxin and then injected locally with EVs at day 1, 3, and 5 post-injury. Wound assessed 7 days post-injury.	309
Choi 2016	Human adipose-derived stem cells (HASCs) treated with EVs obtained from differentiating human skeletal myoblasts (HSkMs) resulted in increased proliferation with increasing doses of EVs (0-100 µg/mL) but this was inhibited at the highest dose (200 µg/mL), relative to a growth factor containing control medium. The fusion index of HASC myotubes was increased with 50 µg/mL treatment of HSkM-derived EVs. EVs injected locally into mouse model of skeletal muscle injury improved wound repair (collagen deposition reduction and increased centrally nucleated fibres at day 7 post treatment).	Polymer precipitation commercial kit to purify EVs from HSkMs during differentiation. Treated HASCs with EVs from HSkMs. Treated mouse model of skeletal muscle injury by local injection of HSkM-derived EVs. Treated rat model of skeletal muscle injury by injection of myomiRs (including miR-1, miR-133, and miR-206).	308

First author	Findings	Methods	Ref
Fry 2017	Myogenic progenitor cells (MPCs) interact with fibrogenic cells to regulate extracellular matrix remodelling through production of EVs containing miR-206. MPC-derived EVs containing miRNAs downregulated fibrogenic cell expression of collagen I, III, and VI, as well as fibronectin.	Polymer precipitation commercial kit to purify EVs from murine MPCs with and without miRNA depletion in host mice. Treated fibrogenic cells with EVs containing, or depleted of, miRNAs and measured collagen production.	307
Mytidou 2021	Observed different profiles of myomiR expression in EVs obtained from different hindlimb skeletal muscles, and levels were similar in the tissue of origin and the EVs produced. myomiR levels in EVs in serum were much lower than the muscle/muscle-derived EVs in culture. AntagomiR-133b reduced local and serum levels of miR-133b (and miR-133a) in EVs. Likely a local level of communication mediated by skeletal muscle EVs with proximal muscle tissue.	Polymer precipitation commercial kit to purify EVs from five whole hindlimb skeletal muscles from mice, cultured *in vitro* for 24h for EV production. Measured myomiR levels in murine skeletal muscle EVs (miR-1, miR-133a, miR-133b, and miR-206). Polymer precipitation commercial kit to purify EVs from serum of mice. AntagomiR-133b provided *in vitro* and *in vitro*.	312

1.7 Protein translation as a mechanism of gene regulation

Expression of protein-coding genes is tightly regulated at a number of stages, including transcription, post-transcriptional processing, nuclear export and localisation of pre-mRNA species, splicing and transcript stability, translation of mature mRNA, and post-translational modification. The cell's ability to control translation allows for an immediate response to cellular signalling or changes in environmental conditions.

1.7.1 Canonical eukaryotic translation

Ribosomes are the macromolecular machines that carry out the fundamental cellular process of translation, converting the genetic information encoded in single stranded messenger RNA (mRNA) transcripts into proteins. Translation is a broadly conserved process across eukaryotic, bacterial, and archaeal taxonomic domains, however, there are major differences in the ribosome subunits, as well as the structure of mRNA transcripts, and how each of the steps involved in translation is carried out. The scanning model of canonical eukaryotic translation states that the small ribosomal subunit binds to the 5′ end of a mRNA transcript and progresses unidirectionally until a start codon is encountered in an optimal sequence context [313,314]. The functional message for the sequence of amino acids that comprise a protein product is contained within an open reading frame (ORF). ORFs are defined by a series of trinucleotide units (i.e. codons), beginning with a preferred initiation codon (i.e. AUG, encoding methionine), and concluding with a termination codon (i.e. UAA, UAG, or UGA). The message within any given codon is read by the anti-codon sequence of a transfer RNA (tRNA) that is covalently bound to its cognate amino acid (i.e. an aminoacyl-tRNA). ORFs are typically flanked by 5′ and 3′ untranslated regions (UTRs), and the mRNA transcript is protected from nuclease degradation on either end by a 5′ 7-methylguanylate (m^7G) cap and 3′ poly(A) tail. Translation of a peptide chain involves the major steps of scanning [313,315], initiation, elongation, termination, and release of the peptide as well as recycling of ribosomal subunits from the mRNA transcript.

A 43S pre-initiation complex (PIC) is an assembly of ribonucleoproteins that scans a mRNA transcript from the m^7G capped 5′ end for the presence of an ORF initiation codon. The PIC consists of the 40S small ribosomal subunit, eukaryotic initiation factors (eIFs, namely eIF1, eIF1A, eIF5, and the multi-subunit eIF3), and the initiator tRNA carrying methionine ($tRNA_i^{Met}$) which is bound to phosphorylated eIF2. A rate-limiting step in translation is the phosphorylation of eIF2 by one or more eIF2α kinases (e.g. PERK, GCN2, HRI, or PKR) and this process is sensitive to changes in the cellular microenvironment as a natural means to controlling rates of translation (for instance in response to viral infection, UV irradiation, temperature) [316]. The mRNA transcript is activated for recognition by the PIC through the direct interaction between eIF4G, the m^7G cap, the poly(A) binding protein (PABP), and other eIFs (i.e. eIF4A and eIFB). Interaction between eIF4G and PABP generates a circular mRNA structure, and a protein bridge for recruitment of the PIC to the m^7G cap. eIF4A is a helicase that functions to unwind complex secondary structures where they occur in mRNA sequences, thereby enabling scanning of the PIC for the first AUG codon within a suitable sequence context (i.e. the Kozak consensus) [317,318]. The optimal Kozak context has been defined to be 5′-GCCGCC(**A/G**)CCAUG**G**-3′ (where the A of the AUG start codon is defined as the +1 position), where a purine in the -3 position is the most important determinant of a strong Kozak identity, and a G in the +4 position further strengthens the sequence context [318]. Scanning is halted, and translation is initiated upon recognition of a suitably located start codon through perfect complementarity base pairing between the anti-codon of $tRNA_i^{Met}$ and the AUG within the peptidyl-tRNA (P) site of the ribosome. Irreversible hydrolysis of guanosine triphosphate (GTP) to guanosine diphosphate (GDP) bound to eIF2 by a GTP-ase activating protein (i.e. eIF5A) primes the 48S complex for recruitment of the 60S large ribosomal subunit. The resulting 80S mRNP complex performs the elongation step of translation.

Amino acids carried by their cognate tRNAs are sequentially incorporated into the growing nascent peptide chain by decoding at the aminoacyl-tRNA (A) site. The nascent peptide chain is transferred to the aminoacyl-tRNA within the ribosomal A-site to enable peptide bond formation through the activity of the large ribosomal subunit peptidyl transferase centre. Once the bond is formed, the ribosome progresses and the unbound tRNA is translocated to the exit (E) site, the elongating peptide chain into the P-site, and the A-site is free to accept the next incoming aminoacyl-tRNA. Elongation is an energy dependent process facilitated by the co-ordinated action of the ribosome and eukaryotic elongation factors (eEFs) that undergo GTP to GDP hydrolysis. The processive activity of the ribosome depends on the amino acid sequence, kinetics of the peptidyl transferase reaction, and substrate availability. Ribosome rescue is a mechanism that functions to recycle a stalled ribosome that is unable to progress, for example as a result of stable secondary structures or the absence of an appropriate aminoacyl-tRNA, and is mediated by the activity of the ATP-binding cassette protein (ABCE1) and eRF1-like protein Pelota (PELO) [319].

The presence of a termination codon in the A-site is recognised by eukaryotic peptide chain release factors (eRFs), and the nascent peptide is released from the P-site tRNA [320,321]. ABCE1 mediates the dissociation of the 80S ribosomal complex, and the large and small ribosomal subunits are thereby recycled for subsequent rounds of translation [322].

1.7.2 *Trans* mechanisms of translational control

It was previously believed that mRNA length influences transcript stability and therefore protein expression levels [323], however, it has since been shown that there is no correlation between the length of eukaryotic mRNA or its poly(A), and the number of associated ribosomes or protein expression [324]. Indeed, often the levels of cellular mRNA transcripts correlate poorly with protein expression [325,326]. Translational control can be carried out on a transcriptome-wide,

or transcript-specific scale by various *trans* and *cis* regulators. A major way in which cells exert control over protein expression in *trans* is through the RNAi pathway, particularly by miRNA binding to cognate sequences which are typically contained within the 3′ UTR of mRNA transcripts. Fully complementary base pairing between the miRNA and its target sequence directs the mRNA transcript for cleavage and degradation, while partial complementarity results in slicing-independent mechanisms of repression (i.e. decapping, and deadenylation) [235,327]. Both of these fates of miRNA-mediated targeting result in downregulation of protein expression.

ADAR-mediated mRNA editing is a natural cellular process for the conversion of adenosine to inosine by deamination. A-to-I editing is thought to occur in 1.4% of human transcripts, and is mainly targeted to dsRNA formed by Alu repeats which are located within intronic and non-coding regions of the genome [328–331]. The functional consequence of this is believed to be the introduction, or destruction, of splice signals that result in splice modulation. Alternative-splicing is a similar means to achieving control of gene expression, often in a cell- and/or tissue-specific manner. Interestingly, human ADAR1 not only facilitates A-to-I editing of pre-mRNA species, but also functions in miRNA processing through direct interactions with Dicer, thereby promoting miRNA maturation and loading into the RNA-induced silencing complex (RISC) [332,333]. ADAR therefore plays an important role in translational control that, until recently, was relatively unknown. Ribosomal frameshifting, stop codon readthrough, and the use of non-canonical amino acids are other cellular mechanisms employed for translational control [334].

1.7.3 The discovery and annotation of small open reading frames

Major strides have been made to complete sequencing of the entire human genome [335], however, annotation of the full protein-coding potential lags significantly further behind [336].

Previous efforts to bioinformatically predict and annotate these regions of the human proteome have been biased by the assumption that functional ORFs should be greater than 100 codons in size for the production of proteins that fold into structures that can confer bioactivity [337]. Other limiting constraints, such as amino acid conservation, homology to functional protein domains, and the requirement for an AUG start codon, were also typically applied in such searches. Over recent decades, this simplified approach towards elucidating the protein-coding potential of the genome has been challenged. Advancements in proteomics methods, deep sequencing and genome engineering technologies has allowed for the unprecedented detection of peptides translated from small ORFs [338–352] (smORFs, typically 10-100 codons in length and otherwise referred to as micropeptides [338,353]), including those with non-canonical start codons [339]. A number of micropeptides have been characterised, playing important roles in cellular functioning. For example, the 60 amino acid *Drosophila* antennapedia homeodomain peptide [354] or the 86 amino acid human immunodeficiency virus (HIV) TAT trans-activator [355] from which therapeutically active cell-penetrating peptides (CPPs) are derived. The neutrally-charged character of PMO molecules means that they can be covalently coupled to non-natural cationic CPPs in order to generate peptide-PMO (PPMO) conjugates which have been developed by our group for the improved delivery of dystrophin splice-corrective therapies to skeletal and cardiac muscle [157]. The leading PPMO drug is SRP-5051 (developed by Sarepta and targeting *DMD* exon 51) which is currently in Phase II (NCT04004065) and for which preliminary data is suggestive of enhanced efficacy over eteplirsen [356].

The generality of rules defining non-coding RNA is constantly being revised as evidence for the production and function of smORFs encoded within these regions of the transcriptome is uncovered. For instance, long non-coding RNAs (lncRNAs) are believed to be similar to evolutionarily young protein coding sequences [357], and some evidence exists to suggest that ORFs within these sequences express functional micropeptides [338–340,344–346,358].

Examples with relevance to muscle biology include DWORF (34 amino acids) [340], myoregulin (46 amino acids) [338], mitoregulin [344]/MOXI [345] (56 amino acids), and SPAR (90 amino acids) [346]. The 5′ UTRs of mRNA transcripts are also known to harbour upstream AUG start sites and open reading frames (i.e. uAUGs and uORFs). The introduction of an out-of-frame uAUG into the leader sequence of preproinsulin was found to reduce the expression of proinsulin, consistent with the scanning model [313,314]. Furthermore, the functional relevance of upstream open reading frames (uORFs) was suggested by the discovery that uAUGs are the most conserved sites within the 5′ UTR of canonical mRNA transcripts [359], and fewer exist than would be expected by chance [359–361]. uORFs are the most abundant class of bioinformatically predicted smORFs within the transcriptome [336], and have been identified in approximately half of all human transcripts [359,362]. A recent study has identified >200 uORF-encoded functional micropeptides in induced pluripotent stem cells (iPSCs) by combining a CRISPR knockout screen with ribosome profiling (i.e. deep sequencing of ribosome-protected RNA) [363]. The *trans* activity of these uORFs was demonstrated by changes to growth phenotypes upon knockout which could be rescued by ectopic expression of the uORF peptide alone. The predominant mechanism of uORF activity appears to be the *cis* regulation of pORF translation, although the discovery of uORF-derived functional micropeptides also suggests that some uORFs can also function in *trans*.

1.7.4 Upstream open reading frame (uORF)-mediated *cis* translational control

uORFs are defined by the presence of a start codon within the 5′ UTR of mRNA transcripts with an in-frame stop codon which can be contained within the leader sequence, or overlap with the downstream primary ORF (pORF) [362,364–366]. Expression of protein from primary coding sequences (i.e. from the primary ORF, pORF) of transcripts containing uORFs was found to be lower than those without, and the half-life of associated transcripts was also reduced [324]. uORFs act as *cis*-regulatory elements that constitutively reduce translation at the pORF by

preventing the scanning ribosome from reaching the pORF [324,359,365,367–370], or by the induction of nonsense-mediated decay (NMD) of the associated mRNA transcript as a result of the presence of a premature termination codon [362]. Alternatively, uORFs have the potential to affect isoform usage [371,372], in some instances through co-ordinated action with a *trans*-acting protein component [373]. For example, CUGBP1 has been shown to interact with CUG repeats within the 5′ UTR of C/EBPβ to regulate expression of two different protein isoforms from neighbouring AUGs on a single mRNA transcript [373]. Interestingly, CUGBP1 activity is affected by the CUG repeat expansions in the mRNA of DMPK which are causative of myotonic dystrophy (DM1) [373,374]. Greater repression has also been observed with multiple uORFs per mRNA transcript, longer uORF peptide sequences, greater distance from the m^7G cap, proximity to the pORF start, or if the surrounding Kozak context was stronger [359,362,375–378].

The evolutionary reason for the existence of uORFs is relatively poorly understood and research into this is ongoing. A likely hypothesis is that uORFs exist to prevent wasteful or harmful protein production through restraining pORF expression. Furthermore, the incongruence between mRNA and protein levels in cells could be explained in part by the activity of uORFs, enabling cells to facilitate rapid changes in gene expression without necessitating the lag between transcription and translation.

In this thesis, uORFs are considered only as static regulators. The canonical *cis*-regulatory function of uORFs in repression of downstream translation is believed to occur via two main mechanisms, including recycling of ribosome components post termination, or ribosomal stalling during elongation or at termination (**Figure 1.8**). Canonical scanning of the PIC carrying the initiator tRNA ($tRNA_i^{Met}$) occurs in the search for an AUG codon within an optimal sequence context [315]. In the case of mRNA transcripts that lack a uORF within the 5′ UTR sequence, initiation of translation can logically only occur at the pORF start site (i.e. of

the coding DNA sequence, CDS) (**Figure 1.8A**). Similarly, the presence of a uORF that is not in an optimal sequence context may result in the uAUG not being recognised by the ribosome. Under such circumstances, 'leaky' scanning ensues and the probability of translation initiation at the pORF is increased [379,380] (**Figure 1.8B**). In some cases where the uORF micropeptide is successfully translated, the 40S ribosomal subunit may remain bound to the mRNA transcript and acquire the appropriate initiation factors to enable reinitiation of translation at the pORF start [376,377,380] (**Figure 1.8B**). Conversely, a non-productive pORF output could result from recycling of ribosomal subunits post uORF translation, or by ribosome stalling during elongation or at the termination stage (**Figure 1.8B**).

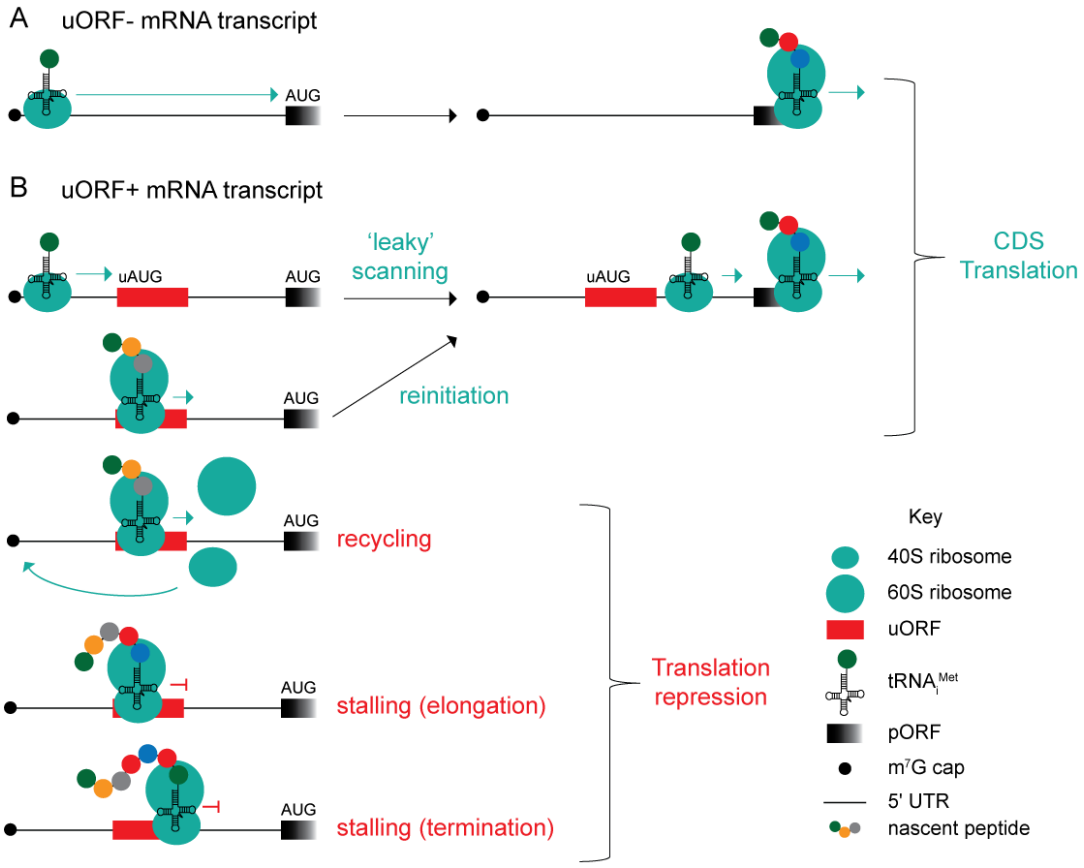


Figure 1.8 Ribosomal fates during scanning of uORF-containing mRNA transcripts.
The small subunit of the ribosome (40S) carrying the initiation transfer RNA with Methionine ($tRNA_i^{Met}$) scans the 5′ UTR from the 7-methylguanylate (m^7G) cap in the direction of the primary open reading frame (pORF). (**A**) Where no uORF is present in the leader sequence (uORF-), translation will initiate at the AUG start codon of the pORF (i.e. the coding DNA sequence, CDS) by association with the 60S subunit of the ribosome. (**B**) Where a uORF is present (uORF+), a number of ribosomal fates are possible. The CDS may be translated where the uAUG is not recognised by the scanning ribosome ('leaky' scanning), or where the uAUG is recognised and the uORF peptide is translated but the 40S subunit remains bound and reinitiates at the pORF start. CDS translation is repressed where the ribosome initiates translation of the uORF peptide and is recycled following termination, or is stalled either during the elongation or termination steps in the translation pathway.

1.7.5 Medical relevance of uORFs

Naturally occurring polymorphisms have been identified in 15.9% of all human genes that affect (i.e. either introduce or disrupt) uORF initiation or termination codons, or alter the strength of uORF Kozak consensus sequences [381]. Aberrant expression of proto-oncogenes is a key event in the process of tumorigenesis. uORFs are predicted to exist in a large proportion of proto-oncogenes [382,383] with loss-of-function uORF mutations identified in several human malignancies [384], and uORFs are abundant in genes associated with cellular differentiation, cell cycling, and stress response [382]. A number of genes with mRNAs containing uORFs that have been associated with particular human diseases are described in **Table 1.3**. Other genes with uORFs that have relevance to Alzheimer's disease (i.e. *BACE1*) [379], breast cancer (i.e. *ERBB2*) [385], general proto-oncogenes (i.e. *MYC*, *BCL2*, *VHL*, *CDKN1B*) [386], asthma [386], type 2 diabetes (*KCNJ11*, *ARAP1*) [386], and Parkinson's disease (*PARK7*) [386] have been identified in genome-wide studies. Furthermore, variation in the levels of human clotting factor XII (*FXII*) causing predisposition to thrombosis is thought to result from a SNP (rs1801020) present in 20% of Caucasian, and 70% of Asian populations whereby an out-of-frame uORF proximal to the *FXII* start site is disrupted [359,387–391].

Methods to directly modulate uORF activity are therapeutically attractive, and offer a novel means to achieving endogenous gene activation. The current gold standard approach to enhancing protein expression is through the use of exogenously overexpressed transgenes which may result in supraphysiological levels of transgene expression, or cellular treatment with small molecules that result in global epigenetic modifications and could have pleiotropic effects. In cases where a gene is under the control of a functional uORF, directly targeting the start codon of the uORF offers an alternative solution to overcoming these challenges. ASOs have been employed to mask the *LRPPRC* gene uORF start site [392], and other inhibitory motifs within the 5′ UTR of *ACP1* [393], resulting in upregulation of downstream protein expression *in*

vivo. Alternatively, several studies have employed single guide CRISPR targeting strategies for NHEJ-mediated knockout of uORF start sites in human cells [394] and in plants [395,396]. The CRISPR toolbox offers the potential for genomic DNA [171,172,397,398] and mRNA [222,225] targeting, and is promising both for therapeutic development, as well as for elucidating the role that uORFs play in their native genetic context without the need for synthetic reporter systems.

Table 1.3 Disease-causing uORF mutations

Mutation type	Gene	Notes	Associated disease	Ref
Disruption of uORF	Thrombopoietin (*THPO)*	Introduction of a non-canonical splice site, shorter 5' UTR sequence (deletion of one or more uORF features)	Hereditary thrombocythemia	399
	HR lysine demethylase and nuclear receptor corepressor (HR)	uORF start (uATG > uGTG) 34 aa; *321; †302	Marie Unna hereditary hair loss	400
	EPH receptor B1 (*EPHB1*)	uORF start (uATG > uGTG) Overlapping, out-of-frame, >70 aa; *211; †161	Cancer (breast and colon)	384
	Mitogen-activated protein kinase 6 (*MAP2K6*)	uORF start (uATG > uACG) 2 aa 'minimal' uORF (Met-STOP);*246; †42	Cancer (colon adenocarcinoma)	384
	Kinase insert domain receptor (*KDR*)	uORF start (uATG > uGTG) 68 aa; *271; †31	Cancer (lung, pancreatic)	384,401,402
Introduction of uORF	Cyclin-dependent kinase inhibitor 2A (*CDKN2A*)	uORF start (G > T) 4 aa; *75; †63)	Familial predisposition to melanoma development	403
	Sex determining region Y (*SRY*)	uORF start (downstream of existing uORF, G > A) 4 aa; *75; †63	Gonadal dysgenesis	359,404
	Interferon regulatory factor 6 (*IRF6*)	uORF stop (A > T) No uATGs, likely non-canonical uORF start used	van der Woude and popliteal pterygium syndromes	359,405
	GTP cyclohydrolase 1 (*GCH1*)	uORF start (C > T) Overlapping, out-of-frame; >6 aa (four Ala repeats); *23; †140	DOPA-responsive dystonia	359,406
	Hepcidin antimicrobial peptide (*HAMP*)	uORF start (G > A) Overlapping, out-of-frame; >8 aa; (Arg-Arg motif); *25; †25	Juvenile hemochromatosis	407
	Potassium inwardly rectifying channel subfamily J member 11 (*KCNJ11*)	uORF start (C > T) Overlapping, out-of-frame; >18 aa; *54; †513	Hyperinsulinemic hypoglycemia, 2	408
	Low density lipoprotein receptor (*LDLR*)	uORF start (del-C) Overlapping, out-of-frame; >7 aa; *22; †63	Familial hypercholesterolemia	409

Mutation type	Gene	Notes	Associated disease	Ref
Introduction of uORF	Peroxisomal biogenesis factor 7 (*PEX7*)	uORF start (C > T) Overlapping, out-of-frame; >15 aa; (Ala/Gly rich); *45; †28	Rhizomelic chondrodysplasia punctata	410
	Pro-opiomelanocortin (*POMC*)	uORF start (C > A) Overlapping, out-of-frame: >3 aa; *11; †92	Proopiomelanocortin deficiency	411
	Protein kinase cAMP-dependent type I regulatory subunit alpha (*PRKAR1A*)	uORF start (G > A) Overlapping; out-of-frame; >32 aa; *97; †28	Carney complex type 1	412
	Serine peptidase inhibitor Kazal type 1 (*SPINK1*)	uORF start (C > T) 2 aa; *54; †40	Hereditary pancreatitis	413
	Haemaglobin subunit beta (*HBB*)	uORF start (G > A) Overlapping, out-of-frame; >9 aa; *29; †21	Thalassaemia β	414
Introduction or strengthening of uORF stop codon	Phosphomevalonate kinase (*PMVK*)	uORF stop strengthening (UAG > UAA)	Ophthalmic, neurological, renal manifestations	415
	VPS53 subunit of GARP complex (*VPS53*)	uORF stop strengthening (UGA > UAA)	Anxiety disorder	
	Sodium leak channel, non-selective (*NALCN*)	uORF stop introducing (CAA > UAA)	Amyloidosis	
	BCL2 like 13 (*BCL2L13*)	uORF stop strengthening (UGA > UAA)	Benign mammary dysplasias, malignant neoplasm of testes, other male genital cancers	
	Serine hydroxymethyltransferase 2 (*SHMT2*)	uORF stop strengthening (UGA > UAA)	Salivary gland disease	
	Modulator of apoptosis 1 (*MOAP1*)	uORF stop introducing (UAC > UAA)	Abnormal movement	
Interactions with RNA binding proteins and uORF	Erb-B2 receptor tyrosine kinase 2 (*ERBB2*)	Interaction of a translational derepression element within the 3' UTR with the uORF through a 38 kDa RNA binding protein, upregulation of ERBB2	HER2 breast cancer	416

* Distance of uORF start to pORF start (nucleotides); † distance of transcription start to uORF start.

1.8 Thesis hypotheses and aims

This thesis describes the development of three innovative molecular medicine approaches that could be applied to neuromuscular disease through the following aims:

1. Applying CRISPR/Cas9 for the genetic correction of the dystrophin gene in a severely dystrophic mouse model.

2. Understanding the role that muscle-derived extracellular vesicles play in pro-myogenic signalling.

3. Elucidating the functional role of upstream open reading frames in translational repression and the potential for gene activation through the disruption of these elements.

2 Materials and methods

2.1 Cell culture

All cell lines were maintained in a humidified incubator at 37°C with 5% CO_2 and were regularly checked for *Mycoplasma* contamination by carrying out polymerase chain reaction (PCR) on supernatant from confluent cultures every two weeks. Cells were detached for passaging and seeding using 1× TrypLE Express Enzyme (Thermo Fisher Scientific, MA, USA). Cells were pelleted prior to nucleic acid or protein extraction by centrifugation at 300 *g* for 5 mins.

2.2 Golden Gate Assembly

2.2.1 Phosphorylation and annealing of ssODNs

The plus and minus strands of complementary single stranded oligodeoxynucleotides (ssODNs, listed in **Table 2.1**) were 5′-phosphorylated and annealed in a reaction mix containing 5 µM of each ssODN, 5 units (U) of T4 Polynucleotide Kinase (PNK; New England Biolabs, MA, USA), 1× T4 PNK Reaction Buffer (70 mM Tris-HCl, 10 mM $MgCl_2$ and 5 mM DTT at pH 7.6) and 1 mM adenosine triphosphate (ATP; Thermo Fisher Scientific) in a final volume of 20 µl of nuclease-free water. Phosphorylation and annealing of ssODNs was carried out in a thermal cycler with the following conditions: 1 hr at 37°C, 10 mins at 75°C, and incremental cooling to 25°C at a rate of -0.2°C/sec. dsODNs were diluted 1:25 before ligation with the linearised backbone.

2.2.2 Backbone digestion

The appropriate plasmid backbone for ssODN cloning was linearised by digestion with 10 U of *BbsI* per µg of DNA, in a reaction mix containing 1× NEB Buffer 2.1 (New England Biolabs) and nuclease-free water to a final volume of 20 µl. Samples were incubated at 37°C for 1 hr

and *BbsI* inactivated at 65°C for 20 mins. Successful digestion was confirmed by agarose gel electrophoresis, and linearised backbones purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific) (**Section 2.4.1**).

2.2.3 Ligation and bacterial transformation

Diluted dsODNs and 50 ng of linearised plasmid backbone were ligated, transformed into competent bacteria, and vector DNA isolated as described below.

2.3 Ligation, transformation, and plating

Ligation of dsDNA was carried out at 22°C for 30 mins in a reaction mix containing 5 U of T4 DNA Ligase, 1× T4 DNA Ligase Reaction Buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP and 10mM DTT at pH 7.5; New England Biolabs) and nuclease-free water to a final volume of 20 µl. The T4 DNA Ligase was inactivated at 65°C for 10 mins. Subcloning Efficiency DH5α Competent Cells (New England Biolabs) were transformed with ligated plasmids on ice for 30 mins, heat shocked at 42°C for 90 secs, returned to ice for 2 mins, and plated on Luria Broth (LB, Thermo Fisher Scientific) agar plates containing 1 µg/mL Ampicillin sodium salt (Sigma-Aldrich, MO, USA) grown at 37°C overnight. Colonies were picked and grown in LB containing 1 µg/mL Ampicillin at 37°C overnight on an orbital shaker. Cultures were pelleted at 1,690 *g* for 10 mins at 4°C, and plasmid DNA extracted as described in **section 2.4.2**.

2.4 DNA purification

2.4.1 PCR products and restriction enzyme-digested plasmid backbones

PCR products and restriction enzyme-digested plasmids were purified using the GeneJET PCR Purification kit (Thermo Fisher Scientific). An equal volume of Binding Buffer was added to each sample volume in water, and the mix was supplemented with 10 µl of 3M sodium acetate (pH 7.0; Thermo Fisher Scientific) to obtain the correct pH for optimal DNA isolation. The

DNA was bound to the purification column, and the protocol was thereafter carried out according to the manufacturer's instructions. DNA was eluted in 20-50 µl of nuclease-free water.

2.4.2 Plasmid vector DNA

Plasmid DNA was extracted from bacterial cultures (which were pelleted at 1700 ×*g* for 5-10 mins) using a QIAGEN Mini, Midi or Maxiprep Kit (Qiagen, Hilden, Germany) as required and according to the manufacturer's instructions. Plasmid DNA was eluted in 50 µl of nuclease-free water.

2.4.3 Genomic DNA obtained from mouse tissue

The DNeasy Blood & Tissue Kit (Qiagen) was used to obtain genomic DNA from mouse tissue samples, as according to the manufacturer's instructions. Tissue sections were collected by cryosectioning (~100 sections of 8 µm width) and kept on dry ice until DNA extraction was carried out. Purified genomic DNA was eluted from the column in 50 µl of nuclease-free water.

2.5 RNA purification

2.5.1 Maxwell RSC simplyRNA Tissue Assay

RNA extraction from cells or tissue was performed using the Maxwell RSC simplyRNA Tissue Kit (Promega, WI, USA). Briefly, elution tubes containing 50 µl of nuclease-free water were placed into the deck tray, and the cartridges prepared according to the manufacturer's instructions. Pelleted cells were then suspended in 200 µl of chilled Homogenisation solution (containing 1-Thioglycerol) by pipetting. Cells were lysed by the addition of 200 µl of Lysis solution, and mixed thoroughly by vortexing for 15 secs. The lysate was transferred to well #1 of the cartridge, and 5 µl of DNAse-I solution added to well #4. RNA was extracted by running the automated simplyRNA Cells protocol on the Maxwell Instrument. Samples were kept on ice throughout.

2.6 Reverse transcription quantitative PCR (RT-qPCR)

2.6.1 cDNA synthesis

RNA concentrations were quantified by absorbance at 260 nm using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), and converted to cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, MA, USA). A 2X master mix was prepared containing 2X RT Buffer, 2X RT Random Primers, 2X dNTP Mix (4mM), and 50U of MultiScribe Reverse Transcriptase, made up to a final volume of 10 µl per reaction with nuclease-free 400 ng of RNA was reverse transcribed in a 20 µl reaction in a thermal cycler with the following conditions: 10 mins at 25°C, 30 mins at 37°C, and 5 mins at 85°C.

2.6.2 qPCR using SYBR Green

cDNA templates were amplified on a StepOne Plus real-time PCR Thermocycler (Life Technologies, CA, USA) using Power SYBR Green Master Mix (Thermo Fisher Scientific) and gene-specific primers. Duplicate technical replicates were performed per biological sample. Each reaction consisted of 2 µl of undiluted cDNA, 10 µl of 2× SYBR Green Master Mix, 1 µl of a 20× primer mix, and nuclease-free water to a final volume of 20 µl. Cycling conditions were as follows: initial denaturation for 10 mins at 95°C, followed by 40 cycles of 15 secs at 95°C and 1 min at 60°C. Melt curve analysis was performed after completion of the cycling protocol (temperature range: 60 – 95°C, at a ramp rate of 0.3°C per sec). A single peak confirmed an absence of primer-dimer products. A no template control (where water was substituted for the cDNA) was included for each assay run. Transcript levels were analysed using the Pfaffl method [417].

2.7 Western blotting for proteins obtained from cell culture

Protein isolation from cell or EV lysates and sample concentration measurements are described in more detail in the relevant methods sections of each results chapter. Protein was added to 1×

NuPAGE LDS Sample Buffer and 1× NuPAGE Reducing Agent (Thermo Fisher Scientific) in water and incubated at 75°C for 10 mins for protein denaturation. Samples were loaded into a NuPAGE 4 to 12% Bis-Tris gel, 1.0 mm, Mini Protein Gel (Thermo Fisher Scientific) alongside a Precision Plus Protein Dual Color Standards ladder (Bio-Rad Laboratories) for size determination, and the gel run at 100 V for 1 h and 15 mins in NuPAGE MOPS SDS Running Buffer (Thermo Fisher Scientific). The proteins were transferred onto an Immobilon-fl polyvinylidene difluoride (PVDF) membrane (Merck, NJ, USA) at 100V for 1 hr on ice in NuPAGE Transfer Buffer (Thermo Fisher Scientific). All antibodies used in this thesis are mentioned in the relevant methods sections, and described in detail in **Table 2.2** and **Table 2.3**.

Table 2.1 Plus and minus strand oligonucleotide sequences for sgRNAs and crRNAs

ID	Sequence (5' to 3')
***Dmd* sg1 +**	caccGTTATTTACCTTCTTCTTGAT
***Dmd* sg1 -**	aaacATCAAGAAGAAGGTAAATAAC
***Dmd* sg2 +**	caccGCAAATATGCGTGTTAGTGTA
***Dmd* sg2 -**	aaacTACACTAACACGCATATTTGC
***Dmd* sg3 +**	caccGAGTCCTTCAAAGATATTGAT
***Dmd* sg3 -**	aaacATCAATATCTTTGAAGGACTC
***Dmd* sg4 +**	caccGCAAAAGCCAAATCTATTTCA
***Dmd* sg4 -**	aaacTGAAATAGATTTGGCTTTTGC
crRNA1 +	caccGCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTGAAG
crRNA1 -	caacCTTCAAAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGC
crRNA2 +	caccGCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTGAA
crRNA2 -	caacTTCAAAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGC
crRNA3 +	caccGCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTGAA
crRNA3 -	caacTTCAAAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGC
crRNA4 +	caccGTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTGA
crRNA4 -	caacTCAAAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAC
crRNA5 +	caccGCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTG
crRNA5 -	caacCAAAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAGC
crRNA6 +	caccGTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTG
crRNA6 -	caacCAAAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAC
crRNA7 +	caccGTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTT
crRNA7 -	caacAAAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAGAC
crRNA8 +	caccGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTT
crRNA8 -	caacAAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAGAGC
crRNA9 +	caccGGGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCT
crRNA9 -	caacAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAGAGCCC
crRNA10 +	caccGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCT
crRNA10 -	caacAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAGAGC
crRNA11 +	caccGCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCT
crRNA11 -	caacAGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAGC
crRNA12 +	caccGGGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTC
crRNA12 -	caacGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAGAGCCC
crRNA13 +	caccGGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTC
crRNA13 -	caacGAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAGAGCC
crRNA14 +	caccGGGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCT
crRNA14 -	caacAGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAGAGCCC
crRNA15 +	caccGTGGGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCC
crRNA15 -	caacGGCAGCTGCAGTGGAGAATCGTGTTAAGCTCGGCTACTGCGGAGAGCCCAC

Note: nucleic acids in lowercase are 5′ overhangs compatible with restriction digested backbone; G nucleotides in red are mismatched with the guide target site but are included to facilitate pol III promoter expression

Table 2.2 Primary antibody information

Application	Primary Antibody	Host	ID	Manufacturer	Dilution	Expected Size (kDa)
WB	Anti-Dystrophin	Mouse	NCL-DYS1	Leica Biosystems	1:100	427
WB	Anti-Vinculin	Mouse	V9131	Sigma-Aldrich	1:200	116
IHC	Anti-Dystrophin	Rabbit	ab15277	Abcam	1:1,000	N/A
IHC	Anti-Laminin-2 α-chain	Rat	L0663	Sigma-Aldrich	1:1,000	N/A
IF	MYH1E	Mouse	MF 20	DSHB	1:20	N/A
WB	Anti-ALIX	Mouse	ab117600	Abcam	1:1,000	96
WB	Anti-Calnexin	Rabbit	ab22595	Abcam	1:1,000	75
WB	Anti-CD9	Rabbit	ab92726	Abcam	1:2,000	25

WB, Western blot. IHC, immunohistochemistry. IF, immunofluorescence.

Table 2.3 Secondary antibody information

Application	Secondary Antibody	Host	ID	Manufacturer	Dilution
WB	Anti-mouse IgG, HRP-linked	Horse	7076	Cell Signaling Technology	1:5,000
WB	Anti-rabbit IgG, HRP-linked	Goat	7074	Cell Signaling Technology	1:5,000
IHC	Anti-rabbit IgG Alexa Fluor 594	Goat	A-11012	Thermo Fisher Scientific	1:500
IHC	Anti-rat IgG Alexa Fluor 488	Goat	A-21212	Thermo Fisher Scientific	1:500
IHC	Anti-mouse IgG Alexa Fluor 488	Goat	A-11001	Thermo Fisher Scientific	1:500

WB, Western blot. IHC, immunohistochemistry.

3 Results I - *In vivo* dystrophin rescue by genomic exon deletion using the *Staphylococcus aureus* CRISPR/Cas9 system

3.1 Introduction

The myriad disease-causing DMD mutations present a challenge for developing personalised corrective gene therapies [418,419]. The most advanced therapies include ASO-mediated targeted skipping of single exons, leading to restoration of the *DMD* reading frame. There are currently four FDA-approved drugs utilising this technology, however, these require lifelong repeat administration, constituting an enormous cost to healthcare systems and patients. The development of technologies that can directly manipulate genomic DNA sequences, such as CRISPR/Cas9, has revealed new possibilities for single-intervention treatments and/or cures for DMD [420–422]. One such approach, which is the focus of this chapter, is the dual-cut strategy for deletion of one or more exons at the genomic DNA level, thereby achieving a permanent 'exon skipped' state [193].

Efficient delivery of the CRISPR/Cas9 system components *in vivo* is of paramount importance for the success of the gene editing approach. AAVs are attractive genetic therapy vectors as multiple serotypes are available with a broad range of tissue tropisms (including for skeletal and cardiac muscle), they are able to transduce both dividing and non-dividing cells, and exhibit low immunogenicity in most people (i.e. for a first-time exposure, and with no pre-existing neutralising antibodies) [423,424]. The well-characterised AAV serotype 9 (AAV9) was chosen for use in this study owing to its skeletal and cardiac muscle tropism [425]. The most widely adopted CRISPR/Cas9 systems are those derived from *Streptococcus pyogenes* (SpCas9) and *Staphylococcus aureus* (SaCas9), as these exhibit the most efficient *in vivo* cleavage activity reported to date [426–428]. The SaCas9 orthologue was selected for use in this study as it has the advantage of occupying a smaller DNA footprint than SpCas9 (3.3 kb

compared with 4.1 kb), and is thus more easily packaged within the limited genomic capacity of the AAV genome (~4.7 kb) [429].

The choice of an appropriate animal model for preclinical research should be carefully considered. The most commonly-used murine model for DMD is the *mdx* mouse, which harbours a nonsense mutation in *Dmd* exon 23, leading to premature termination of dystrophin translation [430]. Multiple studies have demonstrated the utility of the dual sgRNA exon excision strategy *in vivo* by targeting the introns flanking the *mdx* murine *Dmd* exon 23 [431–435]. *Mdx* mice appear phenotypically normal, but when forced to exercise [436,437] or placed under dobutamine stress [162,438–440], the cardiorespiratory pathology in these mice is exacerbated. The lifespan of the *mdx* mouse is only slightly reduced relative to wild-type (WT) controls [441]. In contrast, the lifespan of DMD patients is greatly reduced, with patient death usually occurring in the 2nd or 3rd decade of life [2]. The reason for the mild phenotype in the *mdx* mouse is partly a result of compensatory upregulation of utrophin (the developmental paralogue of dystrophin) [68,69]. A mouse model that more closely resembles the DMD phenotype was generated by crossing the *mdx* mouse with a line that harbours a null mutation in the utrophin gene (*Utrn*^{tm1Ked}), resulting in the utrophin/dystrophin double knockout (dKO) mouse [442,443]. Dystrophic pathology in these mice is much more severe and progressive than in *mdx* mice, with affected animals succumbing by the age of ~8-10 weeks [444,445]. It is important to note that female mice homozygous for *Dmd*^{mdx} and *Utrn*^{tm1Ked} mutations present with similar pathologies to male mice that are homozygous for *Utrn*^{tm1Ked}, and hemizygous for *Dmd*^{mdx}. While this is unlike human disease where females are almost always heterozygous for DMD (and so non-symptomatic carriers), the dKO model better represents the physical debilitation of human DMD disease, while also experiencing premature death. The pathological severity of the dKO mouse therefore makes it an attractive model for testing the success of genetic therapies such

as that employed in this study. Importantly, the dKO model is not applicable to all therapeutic strategies including, for instance, a wide-range of utrophin upregulation approaches.

An emerging theme in the treatment of neuromuscular disorders is that early intervention is a major determinant of therapeutic success. This has been demonstrated in patients with spinal muscular atrophy (SMA) where, if patients are treated at birth with the ASO drug nusinersen, the onset and severity of the disease is significantly mitigated [446,447]. This is also likely to be true for DMD where in the later stages of disease progression, fibro/fatty degeneration of muscle may impair delivery and distribution of genetic therapies [448]. Furthermore, Nelson and colleagues have demonstrated that adverse immune responses to AAV and Cas9 antigens are avoided when *mdx* mice are treated at the neonatal stage [433]. For these reasons, dKO mice used in this study were treated at the earliest possible postnatal timepoint with a single administration of CRISPR-AAVs for the targeted excision of *Dmd* exon 23. Notably, early post-natal treatment in mice, where muscle pathology does not occur until 2-3 weeks of age, is not equivalent to neonatal humans who exhibit muscle pathology at, or very shortly after, birth.

3.2 Hypothesis and aims

I hypothesise that treatment of dKO mice with an AAV9-delivered dual sgRNA CRISPR/SaCas9 system for excision of *Dmd* exon 23 will:

- Result in dystrophin rescue in the skeletal and cardiac muscles of dKO mice.
- Result in lifespan extension in the dKO mice.

This hypothesis was tested with the following aims:

- Screen candidate SaCas9 sgRNAs, and select the most efficient pair for *Dmd* exon 23 excision in murine 10T1/2 fibroblast cells.
- Assess the success of gene editing and dystrophin restoration in dKO mice treated with CRISPR-AAVs by intraperitoneal (IP) and intravenous (IV) injection at the earliest possible postnatal timepoint.

3.3 Materials and methods

3.3.1 Cell culture

10T1/2 *Mus musculus* fibroblast cells were a kind gift from Professor Peter K. Vogt. These cells were grown in Dulbecco's Modified Eagle Medium containing glucose (DMEM, Thermo Fisher Scientific, MA, USA) supplemented with 10% foetal bovine serum (FBS, Thermo Fisher Scientific) and 1% Antibiotic-Antimycotic (PSA: Penicillin, Streptomycin and Amphotericin B; Thermo Fisher Scientific).

3.3.2 Animals

Mice were housed in the University of Oxford Biomedical Services Building under 12 hr light/12 hr dark conditions, with *ad libitum* access to food and water. All experimental procedures were authorised by the UK home office under the project licence 30/2907 in accordance with the Animals (Scientific Procedures) Act of 1986. Mice were housed in individually ventilated cages (IVCs) and handled in sterile laminar flow biosafety cabinets.

Utrophin/dystrophin double knockout (dKO) mice were obtained in the F1 generation by breeding with *Utrn*^{tm1Ked}*Dmd*^{mdx}/J mice (JAX stock #019014; Jackson Laboratory, Bar Harbour, ME, USA) which are heterozygous for the utrophin knockout mutation and homozygous (females)/hemizygous (males) for the *mdx* dystrophin knockout mutation [89]. The *Dmd*^{mdx} mutation is a premature stop codon within *Dmd* exon 23 that aborts full-length dystrophin protein expression [77,430]. The *Utrn*^{tm1Ked} mutation was generated by Deconinck and colleagues via the introduction of a neomycin cassette within exon 7 of the utrophin gene [89,449]. dKO mice in the study reported in this thesis were born at a frequency of ~0.25. The mother's diet was supplemented with wet-mash from the time of littering, and this was also provided to the pups once weaned. dKO mice typically exhibit pronounced kyphosis, joint contractures and skeletal muscle wasting towards the end of their lifespan, and were closely monitored

throughout the study. Tail tip biopsies were obtained on the day of birth (Postnatal day 0, P0) for genotyping. Mice were tattooed at P1 by superficial injection of dye on the underside of the foot for identification purposes. Mice were treated during normal (daytime) working hours, weighed three times a week, and sacrificed at the humane end-point (which was determined by rapid laboured breathing, and reduced mobility leading to difficulty accessing food and water) by increasing CO_2 concentration.

C57BL/10 and *mdx* mice were used as WT and negative controls of dystrophin expression relatively. Notably, the dKO mouse has a mixed genetic background that includes C57BL/10 from the *mdx* line, and C57BL/6 from the utrophin null line. As the WT control in this study was used purely as positive control for dystrophin protein expression, potential differences in physiology between the two genetic backgrounds was not considered to be important. Control mice were culled by increasing CO_2 concentration at 8-12 weeks of age, as this is a typical lifespan of the dKO mouse. Exact age-matching with the experimental dKO mice was difficult as the lifespan of these mice is unpredictable, and the mice were culled at their humane end-point rather than at a fixed timepoint for the lifespan extension study.

3.3.2.1 Genotyping dKO mice

Genomic DNA was extracted from mouse tail tip biopsies using DNA Extract All Lysis Reagents (Applied Biosystems, CA, USA) by lysing the tissue with 50 µl of Lysis Solution, and incubating at 95°C for 3 mins in a heating block. The DNA extract was stabilised by the addition of 50 µl of DNA Stabilising Solution. A TaqMan probe-based qPCR assay was used to genotype the F1 litter of the $Utrn^{tm1Ked}Dmd^{mdx}$/J mice to identify dKO individuals for treatment, and $Utrn^{+/-}/Dmd^{mdx}$ individuals for breeding and colony maintenance. qPCR was carried out using TaqMan Gene Expression Master Mix (Applied Biosystems) and primers to detect the WT utrophin gene (*Utrn*-WT-Fwd and Common-Rev, and a HEX-labelled probe

obtained from Sigma-Aldrich, MO, USA) or the mutant utrophin gene (*Utrn*-mut-Fwd and Common-Rev, and a FAM-labelled probe obtained from Sigma-Aldrich). Fluorescence signals of the probes were quenched using the Black Hole Quencher (BHQ) technology (Sigma-Aldrich). The fluorescence signal is detected during PCR elongation as the probe is displaced and degraded by the 5′ to 3′ exonuclease activity of the Taq polymerase. The following cycling conditions were used: 50°C for 2 mins; followed by 10 cycles of 95°C for 15 secs, 65°C-60.5°C (decreasing at 0.5°C per cycle) for 30 secs, and 68°C for 30 secs; followed by 34 cycles of 95°C for 15 secs, 60°C for 30 secs, and 68°C for 30 secs, with a final hold of 60°C for 30 secs. Primer and probe sequences are listed in **Table 3.1**.

Table 3.1 TaqMan primer and probe sequences for dKO mouse genotyping

ID	Primer sequence (5' to 3')	Probe sequence (5' to 3')
***Utrn*-WT-Fwd**	CGTTATAAGCAGGTGTGGGTC	/5HEX/TCTGCACAGGTGAAGGATGTCATGAAA/BHQ1/
***Utrn*-mut-Fwd**	CTTTACGGTATCGCCGCTC	/5FAM/CGCATCGCCTTCTATCGCCTTCT/BHQ1/
Common-Rev	AGGATCTTCTCGCTGTTTGTC	-

3.3.3 CRISPR/Cas9 plasmid design

The pAAV-CMV-SaCas9(opt) plasmid employed for expression of the *Staphylococcus aureus* Cas9 protein (pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA) was a gift from Feng Zhang (Addgene plasmid #61591; http://n2t.net/addgene:61591; RRID:Addgene_61591) [178]. This plasmid encodes a human codon optimised SaCas9 gene driven by a cytomegalovirus (CMV) promoter/enhancer, flanked by a C-terminal nucleoplasmin nuclear localisation sequence (NLS), N-terminal simian vacuolating virus 40 (SV40) NLS, and 3× human influenza hemagglutinin (HA) tags, with a bGH poly(A) transcription termination signal. The pAAV-U6-sgRNA(SaCas9) plasmid was employed for expression of sgRNAs (targeting sequences fused to a modified SaCas9-specific scaffold sequence for enhanced cleavage activity) [181]. The CRISPR genetic components in each of these

plasmids were contained within AAV2 inverted terminal repeat (ITR) sequences to enable packaging into AAV particles.

The plus (+) and minus (-) strands of single stranded oligodeoxynucleotides (ssODNs) encoding the *Dmd* exon 23 sgRNA target sequences with appropriate overhangs for cloning into the *BbsI* restriction cloning site within pTZ-U6sgRNA(SaCas9) were synthesised by Integrated DNA Technologies (IDT, IA, USA), and cloned using Golden Gate Assembly [450] (**Section 2.2**). The guide sequences designed for targeting within *Dmd* intron 22 (*Dmd* sg1 and sg2) and intron 23 (*Dmd* sg3 and sg4) are listed in **Table 3.2**, and ssODN sequences are listed in **Table 2.1**.

Table 3.2 Target sequences of SaCas9 sgRNAs for targeted removal of *Dmd* exon 23

ID	Target Sequence (5' to 3')
***Dmd* sg1**	TTATTTACCTTCTTCTTGAT
***Dmd* sg2**	CAAATATGCGTGTTAGTGTA
***Dmd* sg3**	AGTCCTTCAAAGATATTGAT
***Dmd* sg4**	CAAAAGCCAAATCTATTTCA

3.3.4 Dual sgRNA screening in murine 10T1/2 fibroblast cells

3.3.4.1 Transfections

10T1/2 mouse fibroblast cells were seeded at a density of 1×10^5 cells in 0.5 mL of DMEM (10% FBS and 1% PSA) per well in a 24 well multiplate. Cells were transfected with 800 ng of plasmid DNA (400 ng of SaCas9 expression plasmid, and 200 ng of each of the dual sgRNA plasmids) using Lipofectamine 2000 (Thermo Fisher Scientific) as according to manufacturer's instructions at a weight/volume ratio of 1:1 of DNA:Lipofectamine 2000. Cells were treated with combinations of sgRNAs flanking *Dmd* exon 23 for excision of the intervening genomic DNA sequence. Cells were harvested 48 hrs later to assess CRISPR-mediated *Dmd* exon 23 excision **Section 2.1**.

3.3.4.2 Genomic DNA extraction, PCR, and Sanger sequencing

Genomic DNA was extracted from CRISPR-treated cells using the KAPA Express Extract kit (Roche, Basel, Switzerland), according to the manufacturer's instructions. PCR was carried out on 2 µl of the DNA extract solution using KAPA2G Robust HotStart ReadyMix (Roche) and primers that amplified a sequence spanning the targeted deletion site. PCR cycling conditions were as follows: initial denaturation at 95°C for 3 mins, and then 35 cycles of denaturing at 95°C for 15 secs, annealing at 53°C for 15 secs, and extension at 72°C for 30 secs, with a final extension step at 72°C for 2 mins. Primer sequences are listed in **Table 3.3**.

Table 3.3 Primer sequences for genotyping of *Dmd* exon 23 excision

ID	Sequence (5' to 3')
***Dmd*-check-Fwd**	GCCTAAATGTCTTAATAATGTTTCAC
***Dmd*-check-Rev**	GCTGTGAGCTAAATCATATCTACA

PCR products were separated on a 3% agarose gel (UltraPure agarose, Life Technologies, CA, USA) and visualised by ethidium bromide staining. PCR products of ~833 bp were expected for full-length (unedited) target sites, and of ~446 bp for ΔExon23 *Dmd* (edited) gDNA. The frequency of exon 23 excision was measured by densitometry of the smaller ΔExon23 *Dmd* product relative to the unedited *Dmd* PCR product using ImageJ software. Band-stab PCR [451] was carried out on the ΔExon23 PCR product. The resulting PCR product was purified using the GeneJET PCR Purification kit (Thermo Fisher Scientific) (**Section 2.4.1**), and analysed by Sanger sequencing (Source BioScience, Cambridge, UK) for genotyping of the CRISPR-edited target site.

TOPO TA cloning for sequencing of individual CRISPR target sites was carried out on bulk PCR products using the TOPO TA Cloning Kit for Sequencing (Thermo Fisher Scientific), as according to the manufacturer's instructions. Ligation, transformation, and

plating was carried out as described in **Section 2.3**, and Sanger sequencing was carried out by Source BioScience.

3.3.5 dKO treatment with CRISPR-AAVs

The *Dmd* dual sgRNA- and SaCas9-expressing plasmid backbones which contained AAV2 ITR sequences (**Figure 3.2A**) were sent to Atlantic Gene Therapies (Institut de Recherche Thérapeutique, University of Nantes, France) for production of two separate AAV vectors of serotype 9 (AAV9). Both vectors will be referred to collectively as CRISPR-AAVs, or individually as SaCas9-AAV and dual sgRNA-AAV, from hereafter.

3.3.5.1 Injections and tissue harvesting

dKO mice were treated at P2 by intraperitoneal (IP, $n = 3$) or intravenous (IV via the facial vein, $n = 4$) injection with a mixture of 10 µl of SaCas9-AAV (1×10^{11} vector genomes, vg) and 10 µl of *Dmd* dual sgRNA-AAV (5×10^{11} vg) using a BD Micro-Fine 29G 0.5 mL Insulin Syringe (BD, NJ, USA). This corresponds with a 1:5 vector dose of SaCas9:sgRNA. Control dKO mice were treated with a 0.9% saline solution in sterile nuclease-free water. Mice were culled at the humane end-point and the heart, diaphragm, tibialis anterior (TA), and triceps muscles were harvested immediately thereafter by macrodissection. Tissues were mounted onto cork discs using O.C.T. Compound Mounting Media for Cryotomy (VWR, PA, USA) and snap-frozen in 2-methylbutane (isopentane, Merck, Darmstadt, Germany) pre-cooled on dry ice. Tissue samples were stored at -80°C prior to downstream processing.

3.3.5.2 Genomic DNA extraction, PCR, and Sanger sequencing

Genomic DNA was isolated from the heart, diaphragm, TA, and triceps muscles (~100 8 µm sections of the heart, TA, and triceps, and 1/6 of the entire diaphragm) using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) (**Section 2.4.3**). PCR was carried out using KAPA2G Robust HotStart ReadyMix (Kapa Biosystems), with primers spanning the dual sgRNA target

sites flanking *Dmd* exon 23 (*Dmd*-check-Fwd and -Rev, **Table 3.3**). Cycling conditions are described above (**Section 3.3.4.2**). PCR products were separated on a 2% agarose gel (UltraPure agarose, Life Technologies) and visualised by ethidium bromide staining. PCR products of ~833 bp were expected for full-length (unedited) target sites, and of ~446 bp for ΔExon23 *Dmd* (edited) gDNA. Band-stab PCR [451] and Sanger sequencing was carried out on the ΔExon23 product according the method described in **Section 3.3.4.2**.

3.3.5.3 RNA extraction and RT-qPCR

Total RNA was isolated from the heart, diaphragm, TA, and triceps muscles (~100 8 µm sections of the heart, TA, and triceps, and 1/6 of the entire diaphragm) using the Maxwell RSC simplyRNA Tissue Kit (Promega, WI, USA) (**Section 2.5.1**). Real-time quantitative PCR (RT-qPCR) was conducted following the MIQE (minimum information for publication of quantitative real-time PCR experiments) guidelines [134]. Complementary DNA (cDNA) was reverse-transcribed using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) as according to the manufacturer's instructions (**Section 2.6.1**). Reverse Transcriptase-quantitative PCR (RT-qPCR) was carried out using a TaqMan probe-based assay to determine the level of *Dmd* exon 23 skipped (using qEx22-24-Fwd and -Rev primers, and a FAM-labelled probe synthesised by IDT) relative to unskipped (with qEx23-24-Fwd and -Rev primers, and a HEX-labelled probe obtained from IDT), transcripts respectively. The probes contained modifications whereby a 5′ FAM or HEX fluorescent dye is double-quenched by internal ZEN, and 3′ Iowa Black (IABkFQ), fluorescence quenchers. As with the BHQ technology described above, Taq polymerase activity displaces and degrades the probe thereby enabling detection of the fluorescent signal during polymerase procession. The double-quenching technology was designed to reduce background fluorescence. qPCR reactions were run in 96 well multiplates using TaqMan Universal PCR Master Mix (Thermo Fisher Scientific) and the StepOnePlus Real-Time PCR System (Applied Biosystems) with the

following cycling conditions: a hold of 50 °C for 2 mins and 95 °C 10 mins, followed by 40 cycles of 95 °C for 15 secs and 60 °C for 1 min. Absolute quantification was carried out using a four-point 10-fold dilution of 'skipped' (965,300 copies per reaction at the maximum concentration) and 'unskipped' (797,100 copies per reaction at the maximum concentration) standards comprised of gBlocks Gene Fragments (IDT). Standards were run in triplicate and unknown samples in duplicate. The following calculation was performed to determine the percentage of transcripts with successful exon 23 excision: ((no. of unskipped molecules)/(no. of skipped molecules + no. of unskipped molecules)) × 100. Primer and probe sequences are listed in **Table 3.4**.

Table 3.4 TaqMan primer and probe sequences for *Dmd* exon 23 exon skipping RT-qPCR

ID	Primer sequence (5' to 3')	Probe sequence (5' to 3')
qExon22-24-Fwd	CTGAATATGAAATAATGGAGGAGAGACTCG	/5FAM/ATGTGATTC/ZEN/TGTAATTTCC/3IABkFQ/
qExon22-24-Rev	CTTCAGCCATCCATTTCTGTAAGGT	
qExon23-24-Fwd	CAGGCCATTCCTCTTTCAGG	/5HEX/TCAACTTCA/ZEN/GCCATCCATTTCTGTAAGGT/3IABkFQ/
qExon23-24-Rev	GAAACTTTCCTCCCAGTTGGT	

3.3.5.4 Protein isolation and Western blotting

Total protein was isolated from the heart, diaphragm, TA, and triceps muscles (~100 8 µm sections of the heart, TA, and triceps, and 1/6 of the entire diaphragm) by lysis in a modified RIPA buffer (10% sodium dodecyl sulphate/SDS, 50 mM tris (pH 8), 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate) containing 1× cOmplete protease inhibitor cocktail (Merck), and homogenised in Precellys homogenisation tubes (Bertin Technologies, Paris, France). These are reinforced tubes containing metal ball bearings which disrupt the tissue when shaken vigorously using the Precellys 24 Tissue Homogeniser (4 cycles at 5,000 rpm for 30 secs). Samples were heated to 100°C for 3 mins, and centrifuged at room temperature for 10 mins at 15,800 *g*. The concentration of each protein lysate was measured using the DC protein assay

(Bio-Rad, CA, USA) according to the manufacturer's instructions, and 20 μg of total protein was loaded onto a NuPAGE 3-8% Tris-Acetate 1 mm 10-well gel (Thermo Fisher Scientific). A standard composition was prepared by combining different ratios of protein isolated from WT (C57BL/10) and *mdx* dystrophin negative control mouse tissues. Gels were run at 150 V for 1 hr and 30 mins in 1× NuPAGE Tris-Acetate SDS Running Buffer (Thermo Fisher Scientific). Protein was transferred onto a PVDF membrane (Merck) at 30 V for 1 hr, switched to 100 V for 1 hr (kept at 4°C) in 1× NuPAGE Transfer Buffer (Thermo Fisher Scientific) supplemented with 0.1 g/L of SDS (Sigma-Aldrich) and 10% methanol. Membranes were blocked in Intercept (PBS) Blocking Buffer (Sigma-Aldrich) for 30 mins at room temperature and incubated in Blocking Buffer supplemented with 0.1% Tween-20 (Sigma-Aldrich) with mouse monoclonal dystrophin (Leica Biosystems, Newcastle upon Tyne, UK) and vinculin (Sigma-Aldrich) primary antibodies at 4°C overnight. Blots were washed with PBS supplemented with 0.1% Tween-20 (PBS-T), and incubated in Horse Anti-Mouse IgG, HRP-linked secondary antibody (Thermo Fisher Scientific) for 1 hr at room temperature. Antibody information is described in **Table 2.2** and **Table 2.3**. Blots were developed using Clarity Western ECL Substrate (Bio-Rad) and imaged using a LI-COR Biosciences Odyssey Fc instrument (LI-COR, NE, USA).

3.3.5.5 Immunohistochemistry staining

Tissues from untreated dKO, saline-treated dKO, and CRISPR-AAV treated dKO mice, as well as WT (C57BL/10) mice, were prepared for cryosectioning as follows:

- The heart was bisected along the cross-sectional axis, and the apex mounted onto the cork disc.
- The diaphragm was laid flat, bisected, and one half rolled and mounted upright on the cork disc.

- One TA muscle per mouse was mounted upright, and the other lengthways, on cork discs for transverse- and longitudinal-sectioning respectively.

- The triceps muscle was mounted upright on the cork disc for transverse-sectioning.

Cryosections with a width of 8 µm were obtained from the heart, diaphragm, TA, and triceps muscles and placed onto positively charged Superfrost Microscope Slides (Thermo Fisher Scientific). Mounted sections were air dried prior to storing at -80°C. Slides were thawed at room temperature and a water repellent barrier was drawn around each section using a Super PAP Pen Liquid Blocker (Daido Sangyo, Tokyo, Japan). Sections were soaked in PBS for 10 mins to remove residual OCT medium, before a blocking solution containing 20% foetal calf serum (FCS, Thermo Fisher Scientific) and 20% normal goat serum (NGS, Thermo Fisher Scientific) in PBS was applied for 1 hr at room temperature. The blocking solution was removed, and sections were stained with dystrophin (Abcam, Cambridge, UK) and laminin (Sigma-Aldrich) primary antibodies in 20% NGS in PBS for 2 hrs at room temperature. Sections were washed three times with PBS and then co-incubated with Goat Anti-Rabbit IgG H&L Alexa Flour 594 and Goat Anti-Rat IgG H&L Alexa Flour 488 secondary antibodies (Thermo Fisher Scientific) in PBS for 1 hr at room temperature, protected from light, and then washed three times with PBS. Antibody information is described in **Table 2.2** and **Table 2.3**. Vectashield Hard Set mounting medium with DAPI H-500 (2BScientific, Oxfordshire, UK) was applied to each section, and a coverslip placed on top. Slides were stored overnight at 4°C and imaged using a Leica DMIRB Inverted Modulation Contrast Microscope (Leica Biosystems) with MetaMorph imaging software. Images were processed using ImageJ software.

3.4 Results

3.4.1 Selection of dual sgRNAs for *Dmd* exon 23 excision in murine fibroblasts

In order to correct the genetic defect in the dKO mouse, a multiplexed CRISPR genome editing strategy intended to excise *Dmd* exon 23 was utilised. Dual guide RNAs were designed to target within intron 22 and 23 in order to introduce DNA DSBs on either side of the target exon, leading to the fusion of flanking introns and excision of exon 23 (ΔExon23) in some cases by NHEJ DNA repair (**Figure 3.1A**). SaCas9 PAM sites (i.e. NNGRRT, where R = A or G) [452] were identified in the intronic sequence surrounding *Dmd* exon 23. Four potential target sites were selected (two each on either side of the exon) and cloned into sgRNA expression cassettes (**Figure 3.1B**). To identify the optimal dual guide combination, mouse fibroblast cells (10T1/2) were transfected with plasmids encoding SaCas9 and all the possible combinations of the four sgRNAs (**Figure 3.1C**). Genome editing was assessed by PCR using primers which span the edit site and produce an 833 bp amplicon in the absence of editing, whereas the presence of a smaller band of ~446 bp indicates successful exon excision. The size of the edited product may vary slightly as a consequence of the introduction of small indels by the error-prone NHEJ repair pathway upon fusion of the flanking intronic regions. Dual sgRNA combination #4 (*Dmd* sg1+4) was found to be the most efficient at inducing the desired genome editing effect with an estimated efficiency of ~4.7% (**Figure 3.1C**). The resulting indels were identified by direct Sanger sequencing of the lower molecular weight PCR product (**Figure 3.1D**). The sequence chromatogram depicts seamless repair of the up- and downstream introns at the expected sgRNA cut sites. An increase in background signal in the chromatogram around the repair site indicates that a low level of indels were introduced by the NHEJ pathway. The dual *Dmd* sg1 and sg4 pair was selected for further investigation.

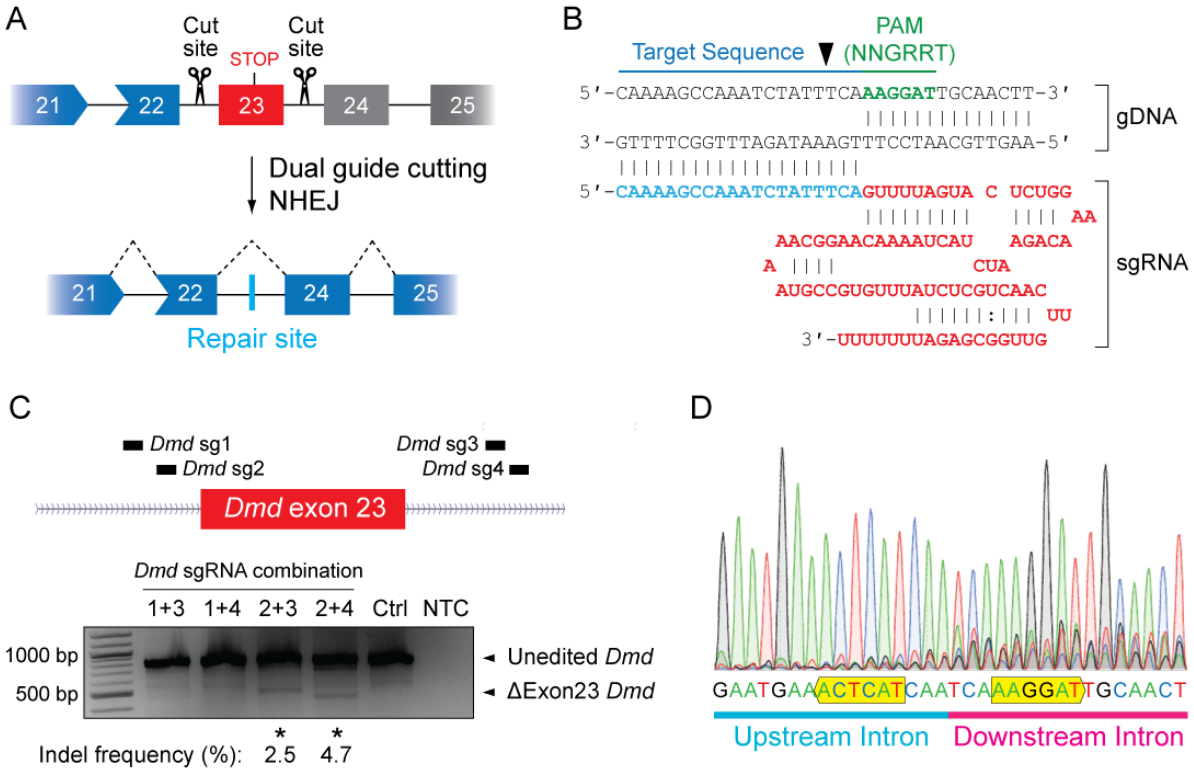


Figure 3.1 Design and validation of a dual sgRNA CRISPR/SaCas9 system for *Dmd* exon 23 excision in cell culture.

(**A**) Schematic of dual sgRNA CRISPR/SaCas9-mediated excision of murine *Dmd* disease-causing exon 23 (shown in red), resulting in NHEJ and a repair indel within the intron while restoring the *Dmd* reading frame. Out-of-frame exons are shown in grey. (**B**) Schematic of the structure of an SaCas9 sgRNA in complex with its cognate gDNA target site. The SaCas9-specific protospacer adjacent motif (PAM) sequence (NNGRRT, where R is A or G) is shown in green and is located on the non-targeted DNA strand. (**C**) Screening for the activity of four combinations of sgRNAs for *Dmd* exon 23 excision in murine 10T1/2 fibroblast cells (*Dmd* sg1 and sg2 targeting in intron 22, and sg3 and sg4 targeting in intron 23). The genomic target site was amplified by PCR and resolved by agarose gel electrophoresis. Successful editing is indicated by a shorter, ΔExon23 *Dmd* product. (**D**) Sanger sequencing was performed by band-stab PCR of the lower molecular weight product (from SaCas9 and *Dmd* sg2+4), demonstrating successful exon 23 excision. The PAM sequences from the original target sites are highlighted in yellow.

3.4.2 Dystrophin restoration in dKO mice by IP administration of CRISPR-AAVs

To evaluate the potential for CRISPR/Cas9-mediated *Dmd* editing *in vivo*, two different AAV vectors were designed. The first was an AAV containing an SaCas9 transgene where compact regulatory sequences were selected (the CMV promoter and SV40 minimal poly(A) signal) to drive expression of SaCas9 in order to minimise the size of the AAV genome. The total size of the resulting AAV genome (including ITRs) was 4,806 bp, which is within the limit of AAV packaging [453]. For the second AAV vector, two different RNA polymerase III promoters (the human H1 and U6 snRNA promoters) were selected to drive expression of the dual sgRNAs in order to minimise the repetitive sequence within the AAV genome (**Figure 3.2A**).

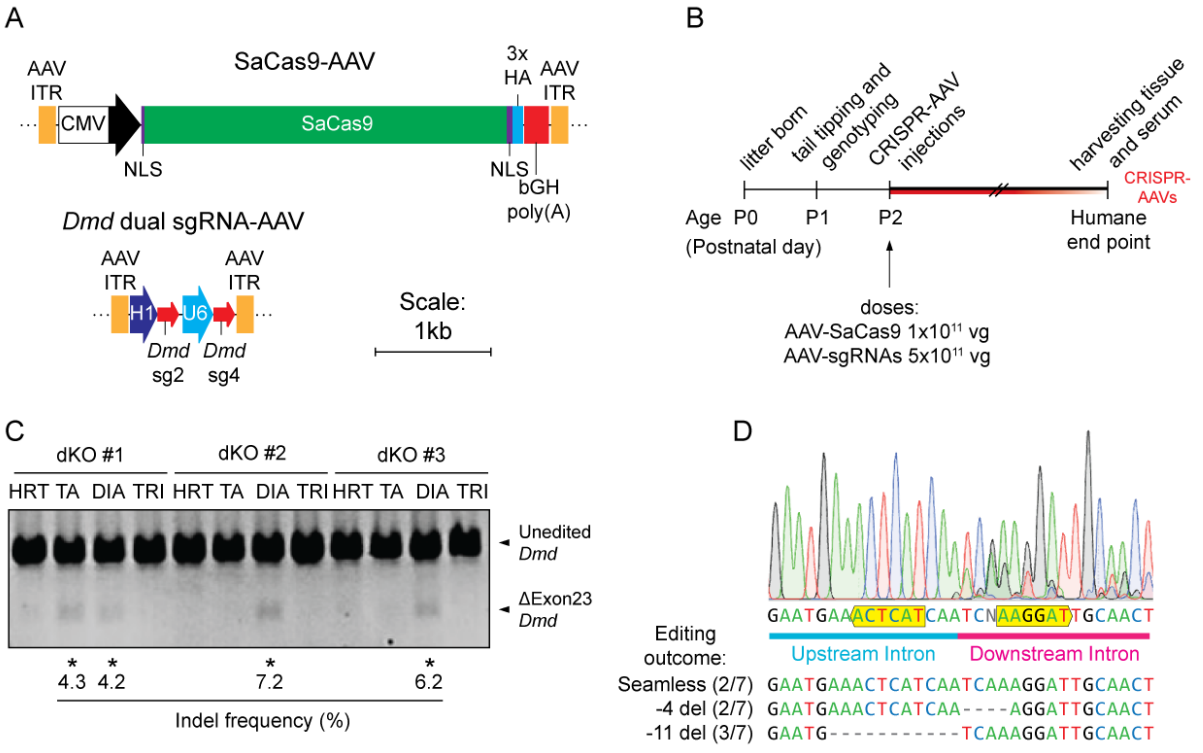

A
SaCas9-AAV
AAV ITR
CMV
SaCas9
3x HA
AAV ITR
NLS
NLS
bGH poly(A)
Dmd dual sgRNA-AAV
AAV ITR
H1
U6
AAV ITR
Dmd sg2
Dmd sg4
Scale: 1kb
B
litter born
tail tipping and genotyping
CRISPR-AAV injections
harvesting tissue and serum
CRISPR-AAVs
Age P0 P1 P2
(Postnatal day)
Humane end point
doses:
AAV-SaCas9 1x10^11 vg
AAV-sgRNAs 5x10^11 vg
C
dKO #1
dKO #2
dKO #3
HRT TA DIA TRI HRT TA DIA TRI HRT TA DIA TRI
Unedited Dmd
ΔExon23 Dmd
4.3 4.2 7.2 6.2
Indel frequency (%)
D
GAATGAAACTCATCAATCNAAGGATTGCAACT
Editing outcome:
Upstream Intron
Downstream Intron
Seamless (2/7) GAATGAAACTCATCAATCAAAGGATTGCAACT
-4 del (2/7) GAATGAAACTCATCAA----AGGATTGCAACT
-11 del (3/7) GAATG----------TCAAAGGATTGCAACT
E
ΔExon23 Dmd transcripts (%)
14
12
10
8
6
4
2
0
HRT DIA TA TRI
F
Heart
Diaphragm
Dystrophin (% of WT)
50 25 5 3 1 AAV-IP
Dystrophin
Metavinculin
Vinculin
Tibialis anterior
Triceps
G
Dystrophin protein (% of WT)
14
12
10
8
6
4
2
0
HRT DIA TA TRI

Figure 3.2 Dystrophin restoration in dKO mice following neonatal IP injection of CRISPR-AAVs.

(**A**) Schematic of the AAV-SaCas9 and AAV-sgRNA genomes within AAV2 inverted terminal repeats (ITRs). The SaCas9 gene was driven by a CMV promoter and enhancer, flanked by an SV40 nuclear localisation signal (NLS) on the N-terminus and a nucleoplasmin NLS on the C-terminus, tagged with three HA tags and terminated with a bGH poly(A) signal. *Dmd* sg2 and sg4 were expressed using human H1 and U6 snRNP promoters respectively. Scale bar represents 1 kb of DNA. (**B**) Neonatal utrophin/dystrophin double knockout (dKO) mice were genotyped at Postnatal day 1 (P1). dKO mice were treated by intraperitoneal (IP) injection (*n* = 3) at P2 with 1×10^{11} vector genomes (vg) of AAV-SaCas9 and 5×10^{11} vg of AAV-sgRNA. Mice were sacrificed at the humane end-point, and DNA, RNA and protein were extracted from the heart (HRT), diaphragm (DIA), tibialis anterior (TA), and triceps (TRI) muscles. (**C**) The genomic target site was amplified by PCR and resolved by agarose gel electrophoresis where successful editing is indicated by a shorter, ΔExon23 *Dmd* product (shown with asterisks and indel frequency calculated as a percentage of the edited band relative to the WT PCR product). (**D**) Sanger sequencing was performed on the lower molecular weight band from the DIA of dKO #2, demonstrating successful exon 23 excision in bulk (chromatogram shown) and following TOPO TA cloning and sequencing of 7 individual target sites (sequences shown beloe). (**E**) Levels of successful exon excision were detected in *Dmd* mRNA transcripts by RT-qPCR, and quantified relative to the number of full-length (unedited) sequences. (**F**) Dystrophin protein expression was detected by Western blot, and (**G**) quantified relative to vinculin which was included as a loading control. Vales are means ± SEM.

Postnatal day 2 (P2) dKO mice were treated with 1×10^{11} vector genomes (vg) of SaCas9-AAV, and 5×10^{11} vg of dual sgRNA-AAV, by intraperitoneal (IP) injection (**Figure 3.2**B). This was the highest possible dose that enabled treatment with a 1:5 ratio of Cas9:sgRNA which was chosen as the sgRNA component has been shown to be limiting for CRISPR/Cas9 gene editing activity [195]. Mice were treated with the CRISPR-AAVs at the earliest possible postnatal timepoint in order to assess the effect of pre-symptomatic intervention on dystrophin restoration and survival. Treated mice were monitored until they reached the humane end-point, at which time the heart, diaphragm, TA, and triceps were harvested in order to determine the efficacy of gene editing.

The ΔExon23 *Dmd* gene product, indicative of productive editing at the genomic target site, was detected sporadically across the tissues as assessed by PCR (**Figure 3.2C**), and successful excision and repair of the flanking intronic regions at the cut sites was confirmed by Sanger sequencing (**Figure 3.2D**). The chromatogram shows seamless fusion of the flanking intronic regions, however, the background signal is increased around the repair site which is indicative of low levels of indel formation. The bulk PCR product was cloned into a TOPO TA vector, and 7 clones sequenced revealing three major editing events, including 2/7 with seamless repair, 2/7 with a 4 bp deletion downstream of the fusion site, and 3/7 with an 11 bp deletion upstream of the fusion site (**Figure 3.2D**).

A TaqMan probe-based RT-qPCR assay was employed to determine the levels of ΔExon23 *Dmd* transcripts, with amplicons spanning exons 22-24 (exon 'skipped'), compared to those spanning exons 23-24 (exon 'unskipped'), when normalised to gene fragment standard curves. Transcript level exon skipping assays can be used here as the gene editing employed manifests as a permanent 'exon skipped' mRNA. Successful exon excision was detected in all tissues from the treated mice by RT-qPCR, and thus it is evident that this method of detecting and quantifying gene editing is more sensitive than end-point PCR of genomic DNA. ΔExon23

Dmd transcript levels were highest in the heart and diaphragm at ~12% and 8% respectively (**Figure 3.2E**). Editing in the TA and triceps was less successful, with only ~2% and 1% ΔExon23 *Dmd* transcripts detected respectively.

Dystrophin protein expression was observed in the heart, diaphragm, and TA muscles of the treated dKO mice, but at negligible levels in the triceps (**Figure 3.2F and G**). The highest level of restoration was detected in the heart and diaphragm at ~4% of WT levels (**Figure 3.2F and G**). The size of the major skeletal muscle isoform (Dp427) of dystrophin in WT mice is 427 kDa, while CRISPR-mediated excision of *Dmd* exon 23 removes only 71 amino acids, generating a 420 kDa internally truncated protein which is difficult to distinguish visually by Western blot. Notably, it is not expected that the dKO mice are completely devoid of dystrophin due to the possibility for the muscles to contain revertant fibres [89], however, this phenomenon is relatively rare and the Dp427 isoform was undetectable by Western blot in saline-treated control dKO animals (**Figure 3.3**).

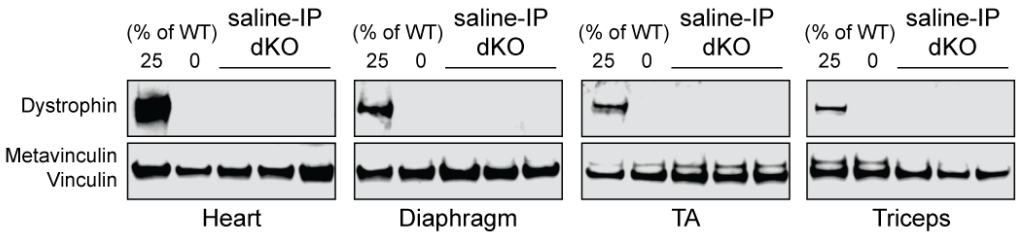


Figure 3.3 Dystrophin protein is undetectable by Western blot in saline-treated dKO mouse tissues.

Dystrophin Western blot in saline-IP treated dKO mice (*n*=3) heart, diaphragm, TA, and triceps muscles. Vinculin was included as a loading control. Mixtures of dystrophin positive and negative tissues (i.e. 25% and 0% of WT) are included as controls.

Immunohistochemistry (IHC) staining for dystrophin protein was performed for visualisation of the pattern of successful gene editing within CRISPR-AAV-treated muscle tissue sections. This method is more qualitative than Western Blot, and exhibits a smaller dynamic range [454]. Transverse sections obtained from each of the four muscle tissues were stained with anti-dystrophin fluorescent antibodies to assess the distribution of restored fibres throughout the selected tissues. Sections were co-stained for laminin, a basal lamina protein that delineates the boundaries of individual myofibres (**Figure 3.4**). IHC of untreated WT mice was included as a positive control for dystrophin organisation within the myofibres of the four tissues analysed. Untreated and saline-IP-treated dKO mice were included as negative controls for dystrophin staining.

Correct dystrophin localisation at the sarcolemma membrane was confirmed in all four tissues analysed following CRISPR-AAV treatment by IP injection. In the heart and diaphragm, where the highest levels of editing were observed by RT-qPCR and Western blot (**Figure 3.2**E-G), the restored fibres were distributed throughout the tissue section (**Figure 3.4**). In the two limb muscles, however, the corrected fibres formed a more clustered pattern, with patches of dystrophin-positive and -negative fibres. This between-fibre 'patchiness' may be due to mosaicism of successful gene editing early in the muscle development of these mice, or as a result of differences in vascular permeability throughout the muscle tissue. Regions of dystrophin-negative staining within CRISPR-corrected dystrophin-positive myofibres were also observed (indicated by arrowheads, **Figure 3.4**). This within-fibre patchiness was most apparent in the TA and triceps muscles and was not observed with laminin staining, indicating that there are regions of individual CRISPR-targeted myofibre sarcolemma that lack dystrophin expression.

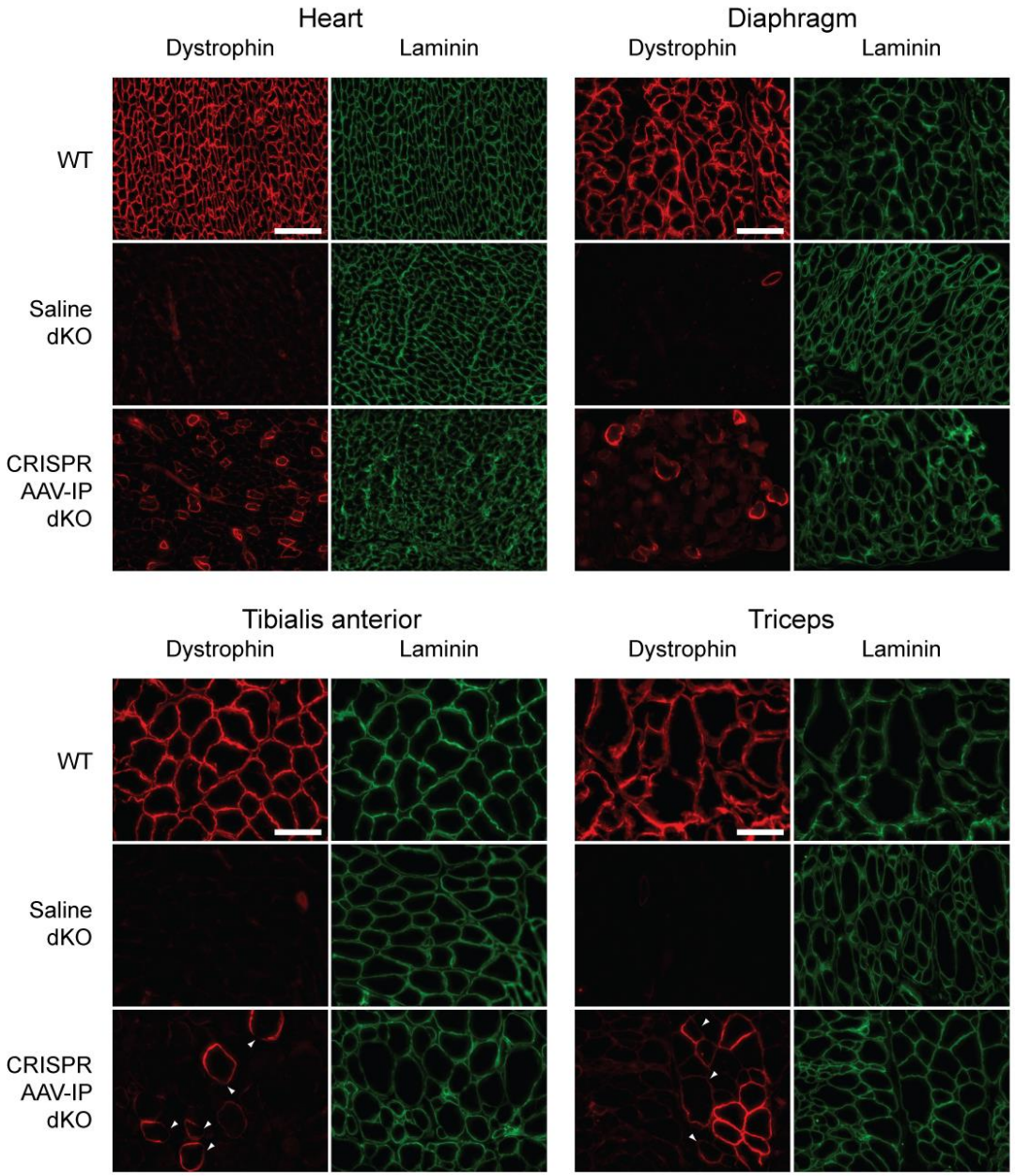


Figure 3.4 Dystrophin restoration in tissue sections from IP-treated dKO muscle.
Immunohistochemistry (IHC) staining of dystrophin and laminin in transverse tissue sections obtained from heart, diaphragm, TA, and triceps muscles of dKO mice treated with CRISPR-AAVs by IP injection at P2. Sections obtained from WT (C57BL/10) and saline-IP-treated dKO mice were included as controls. Images were taken at 20× magnification, and scale bars represent 100 μm. Arrowheads indicate dystrophin-negative regions within dystrophin-positive myofibres (i.e. within-fibre patchiness).

The higher level of gene editing efficacy in the heart and diaphragm relative to the peripheral TA and triceps skeletal muscles, as indicated by qPCR of ΔExon23 *Dmd* mRNA transcripts and Western blotting for dystrophin protein (**Figure 3.4**), may have resulted from their relative anatomical proximity to the injection site within the intraperitoneal space. This phenomenon has been reported by another group [455]. Importantly, the method of IP administration used here is unlikely to be feasible for genetic therapy delivery in humans due to the injected volume that would be required to achieve comparable doses to those used in mice, and because of major differences in human and mouse anatomical structures.

3.4.3 Dystrophin restoration in dKO mice by IV administration of CRISPR-AAVs

The utility of systemic (intravenous, IV) administration of the CRISPR-AAVs was assessed in the dKO mouse model to determine whether broader targeting of muscle tissues could be achieved, and because this is a more widely used method for drug therapy delivery in patients. Several other studies have employed the IV route of administration for dystrophin restoration genetic treatment in neonatal mice, demonstrating the utility of early intervention CRISPR-AAV treatment [181,190,191,194]. An equivalent dose to that used for IP administration of CRISPR-AAVs (i.e. 1×10^{11} vg of AAV-SaCas9 and 5×10^{11} vg of AAV-sgRNA) was delivered to P2 dKO mice via the facial vein (see schematic in **Figure 3.2**B). Successful excision of *Dmd* exon 23 was detected by PCR in ~70% (i.e. 11/16) of the tissues analysed (**Figure 3.5A**), and successful gene editing was confirmed by Sanger sequencing of the lower molecular weight PCR product (**Figure 3.5B**). The precise indels introduced were assessed in 5 clones following TOPO TA cloning of the bulk PCR product (**Figure 3.5B**). While seamless fusion was one of the major editing outcomes, the other products also contained a variety of indel types, as is typical of repair following NHEJ.

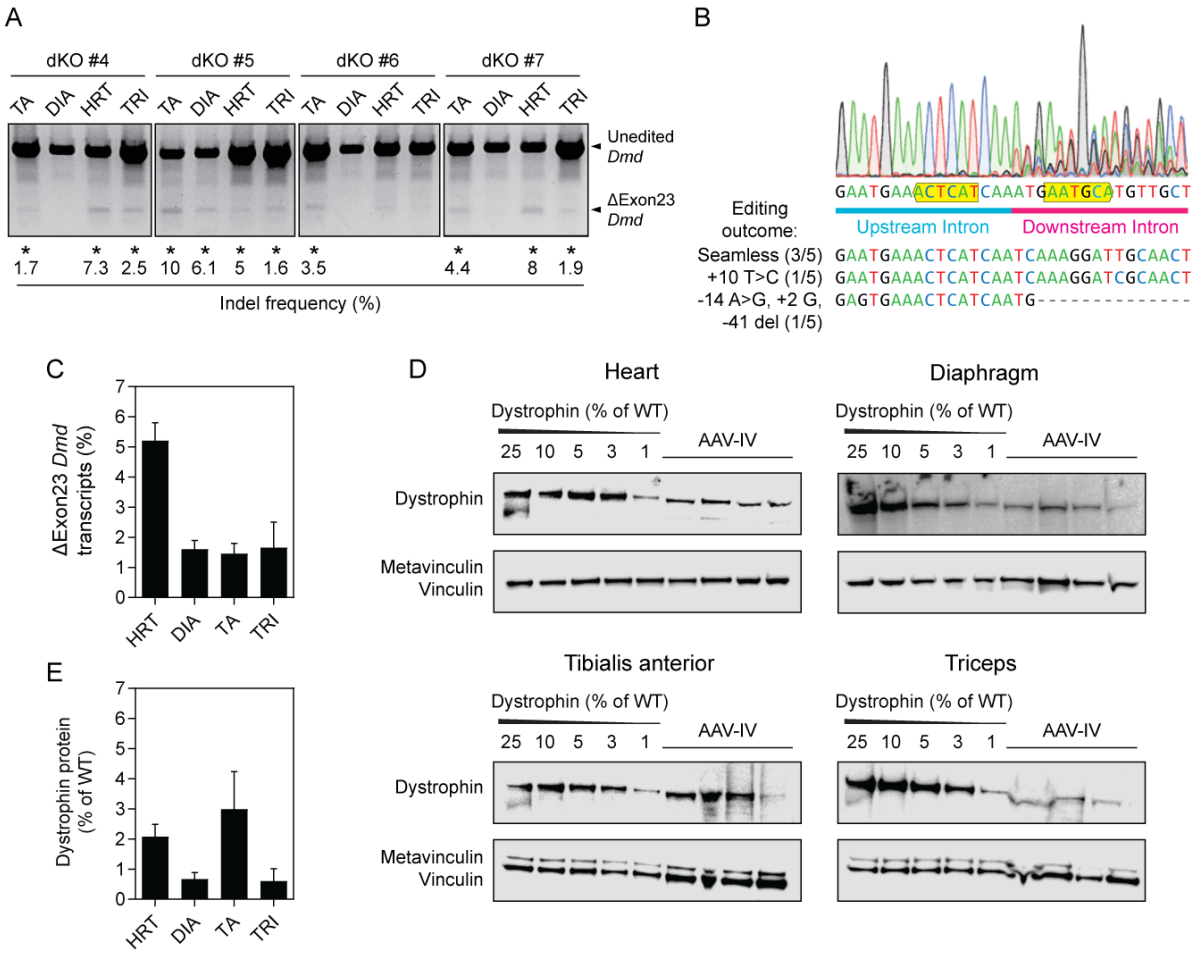

A
dKO #4
dKO #5
dKO #6
dKO #7
TA DIA HRT TRI
Unedited Dmd
ΔExon23 Dmd
1.7 7.3 2.5 10 6.1 5 1.6 3.5 4.4 8 1.9
Indel frequency (%)
B
Editing outcome:
Seamless (3/5)
+10 T>C (1/5)
-14 A>G, +2 G, -41 del (1/5)
GAATGAAACTCATCAAATGAATGCATGTTGCT
Upstream Intron
Downstream Intron
GAATGAAACTCATCAATCAAAGGATTGCAACT
GAATGAAACTCATCAATCAAAGGATCGCAACT
GAGTGAAACTCATCAATG
C
ΔExon23 Dmd transcripts (%)
HRT DIA TA TRI
D
Heart
Diaphragm
Tibialis anterior
Triceps
Dystrophin (% of WT)
25 10 5 3 1
AAV-IV
Dystrophin
Metavinculin
Vinculin
E
Dystrophin protein (% of WT)
HRT DIA TA TRI

Figure 3.5 Dystrophin restoration in dKO mice following neonatal IV injection of CRISPR-AAVs.

dKO mice were treated by intravenous (IV) injection via the facial vein (n = 4) at P2 with 1×10^{11} vg of AAV-SaCas9 and 5×10^{11} vg of AAV-sgRNA in a total volume of 20 µl. Mice were sacrificed at the humane end-point and DNA, RNA and protein were extracted from the heart (HRT), diaphragm (DIA), tibialis anterior (TA), and triceps (TRI) muscles. (**A**) The genomic target site was amplified by PCR and resolved by agarose gel electrophoresis where successful editing is indicated by a shorter, ΔExon23 *Dmd* product (shown with asterisks and indel frequency calculated as a percentage of the edited band relative to the WT PCR product). (**B**) Sanger sequencing was performed by band-stab PCR of the lower molecular weight band from the HRT of dKO #4, demonstrating successful exon 23 excision in bulk (chromatogram shown) and following TOPO TA cloning and sequencing of 5 individual target sites (sequences shown below). (**C**) Levels of successful exon excision were detected in *Dmd* mRNA transcripts by RT-qPCR, and quantified relative to the number of full-length (unedited) sequences. (**D**) Dystrophin protein expression was detected by Western blot, and (**E**) quantified relative to vinculin which was included as a loading control. Vales are means ± SEM.

The proportion of ΔExon23 *Dmd* mRNA transcripts was highest in the heart tissue at ~5%, following IV administration of CRISPR-AAVs (**Figure 3.5C**). Transcripts lacking *Dmd* exon 23 in the diaphragm, TA, and triceps were detected at a much lower abundance of ~1.5%. Dystrophin protein, as measured by Western blot, was partially restored in all four tissues (**Figure 3.5D** and **E**), with the highest expression in the heart and TA muscles, at ~2%, and lowest in the diaphragm and triceps at <1% of WT levels. IHC staining for dystrophin and laminin confirmed that dystrophin was partially restored at the sarcolemma in all of the muscles analysed (**Figure 3.6**). The pattern of dystrophin restoration in the heart, and to a lesser extent also in the diaphragm, was interspersed throughout the muscle, while in the TA and triceps there was distinct between-fibre patchiness. Regions of dystrophin-negative staining in the CRISPR-corrected muscle fibres of the limb skeletal muscles can be seen (indicated by arrowheads), where laminin co-staining was clearly present (**Figure 3.6**). Both between- and within-fibre patchiness were also observed previously with the IP route of CRISPR-AAV administration (**Figure 3.4**), suggesting that dystrophin restoration does not occur uniformly throughout all CRISPR-corrected myofibres.

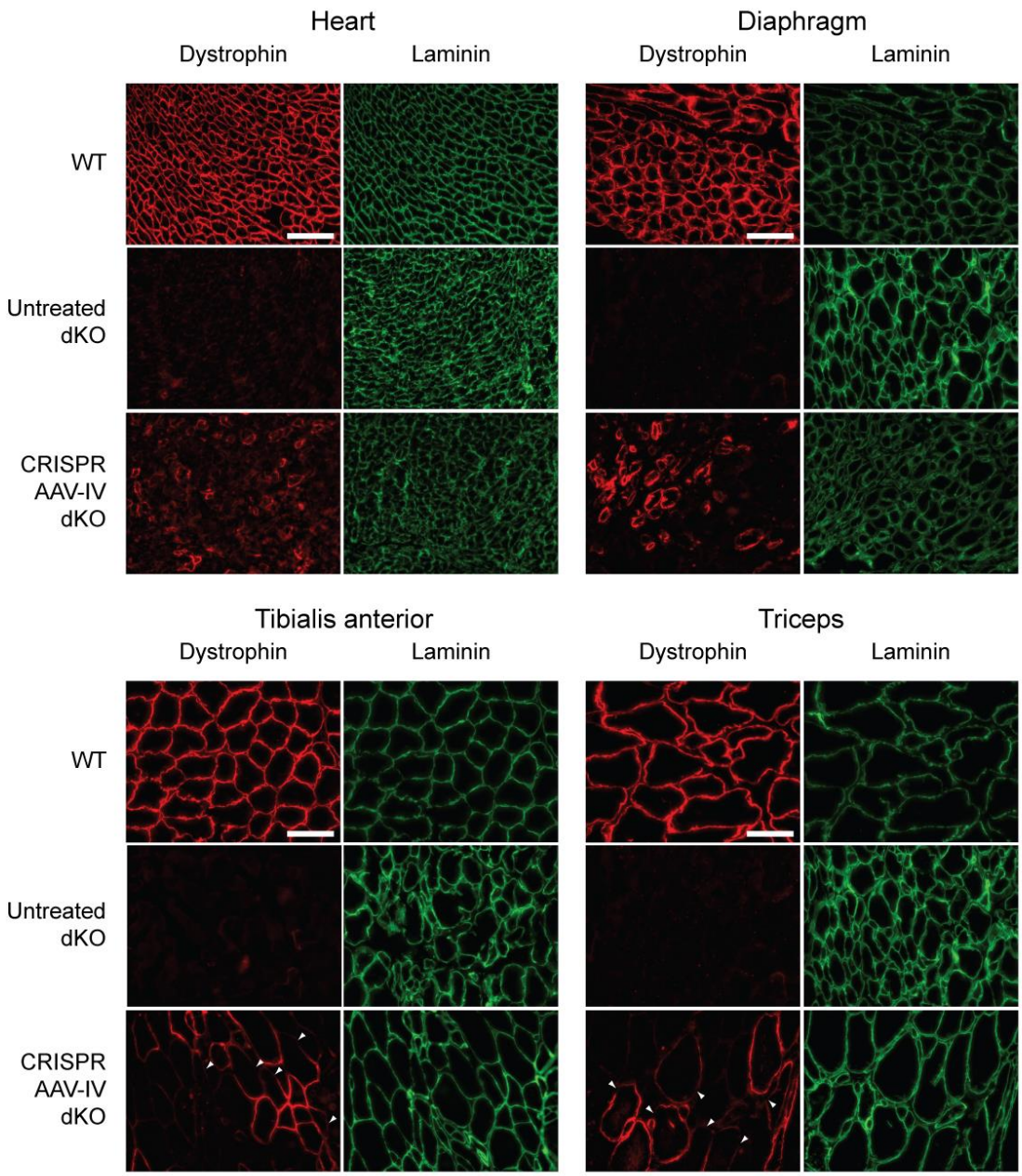


Figure 3.6 Dystrophin restoration in tissue sections from IV-treated dKO muscle.
Immunohistochemistry (IHC) staining of dystrophin and laminin in transverse tissue sections obtained from heart, diaphragm, TA, and triceps muscles of dKO mice treated with CRISPR-AAVs by IV injection at P2. Sections obtained from WT (C57BL/10) and untreated dKO control mice were included as controls. Laminin delineates the boundaries of the myofibres. Images were taken at 20× magnification, and scale bars represent 100 μm. Arrowheads indicate dystrophin-negative regions within dystrophin-positive myofibres (i.e. within-fibre patchiness).

3.4.4 Dystrophin restoration by CRISPR-AAV treatment was insufficient to extend the lifespan of the dKO mice

The dKO mouse model experiences a severe pathology that results in premature death (~8-12 weeks) as well as low body weight, and other overt dystrophic phenotypes such as kyphosis, limb contractures, and laboured breathing towards the end of life. This is in contrast to the commonly used mdx model which experiences an almost normal lifespan (~2 years), body size, and appears physically similar to WT controls. The dKO mouse was selected for use in this study as a means to monitor the clinical benefit of CRISPR-AAV-mediated dystrophin restoration on the lifespan as well as the visual dystrophic phenotype (i.e. the low weight, kyphosis, and limb contractures) of these severely affected mice. The dKO mice treated by IP and IV routes of administration were physically indistinguishable from the untreated or saline-treated controls (data not shown). Furthermore, no significant difference in the weight gain (**Figure 3.7A**) or the average lifespan (~7- and ~8-weeks for the AAV-IP and -IV treatment groups respectively) of the CRISPR-AAV-treated dKO mice was observed, relative to the control group (~8-weeks, $P = 0.7857$ and > 0.9999 for the AAV-IP and AAV-IV treatment groups relative to the saline control respectively) (**Figure 3.7 B**).

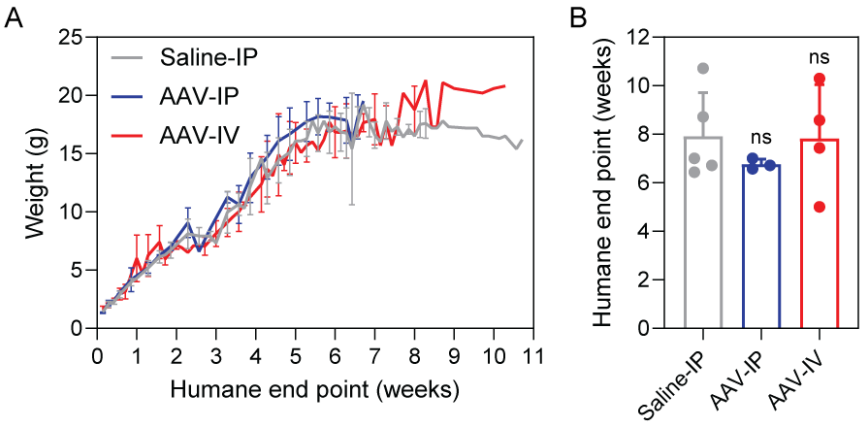


Figure 3.7 Treatment of dKO mice with CRISPR-AAVs by IP or IV injection has no effect on lifespan.

dKO mice were treated with CRISPR-AAVs by IP (n = 3) or IV (n = 4) injection, and weighed three times weekly until they were sacrificed at the humane end-point. Control dKO mice were treated with saline by IP injection (n = 5). (**A**) Average body weight over time for all mice from the three treatment groups. (**B**) Histogram showing age at end-point (in weeks) for each treatment group. Values are means ± SEM. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test. ns, not significant.

3.4.5 CRISPR-AAV treated skeletal muscle exhibits within-fibre patchiness of dystrophin expression

Previous work from our group has identified the importance of uniformity of dystrophin expression at the sarcolemma along the length of individual myofibres for stabilising fibre turnover using the *mdx-Xist*$^{\Delta hs}$ mouse model [456]. These mice express varying levels of dystrophin in a within-fibre patchy manner as a consequence of skewed X-chromosome inactivation. We sought to assess the pattern of dystrophin expression in longitudinal sections of the TA muscle from dKO mice that received CRISPR-AAV treatment by IP and IV injection as within-fibre patchiness is more apparent in this orientation. Such a patchy dystrophin distribution was observed for both treatment regimens, whereby adjacent regions of dystrophin-positive and dystrophin-negative sarcolemma were observed in the same myofibre (**Figure 3.8**). Importantly, this within-fibre patchiness was evident in both male and female skeletal muscle tissue sections (data not shown), indicating that the effect is CRISPR-specific, and not a function of X-chromosome inactivation [456].

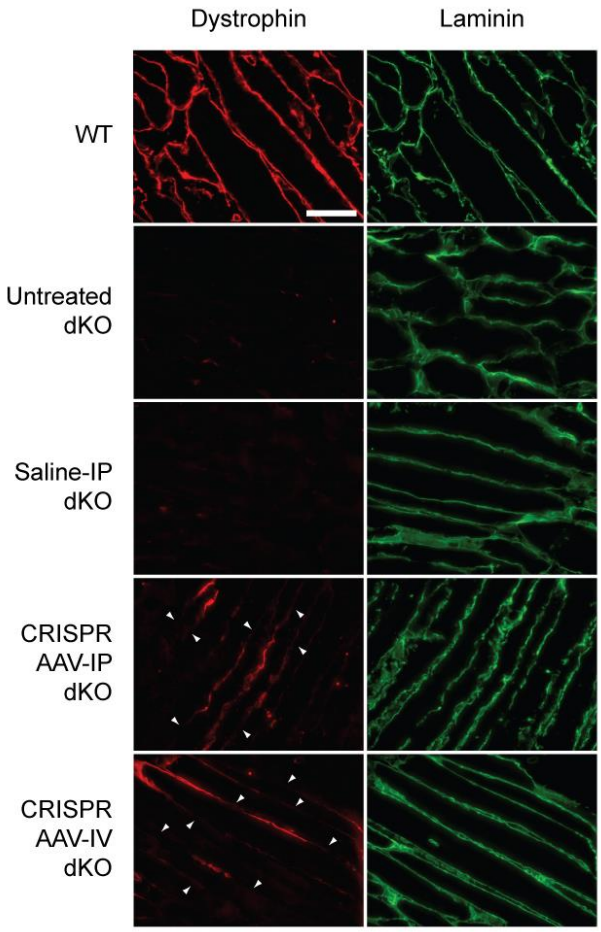


Figure 3.8 CRISPR-mediated correction of dystrophin results in within-fibre patchiness in longitudinal sections of the TA muscle.

Immunohistochemistry (IHC) staining of dystrophin and laminin in longitudinal tissue sections obtained from TA muscles of dKO mice treated with CRISPR-AAVs for dystrophin restoration by IP and IV injection at P2. Laminin delineates the boundaries of the myofibres. WT mice, and untreated or saline-IP treated dKO mice were included as controls. Images were taken at 20× magnification, and scale bars represent 100 μm. Arrowheads indicate dystrophin-negative regions within dystrophin-positive myofibres (i.e. within-fibre patchiness).

3.5 Discussion

This study is the first to demonstrate CRISPR-mediated rescue of dystrophin expression in the severely affected dKO mouse model of DMD. The benefit of this mouse over other commonly used models is the severe dystrophic phenotype that leads to impaired mobility and premature death. The AAV9 serotype was employed in this study to deliver SaCas9 and dual sgRNAs for *Dmd* exon 23 excision in neonatal dKO mice. This enabled assessment of the effect of early intervention genetic therapy on dystrophin correction and survival in two treatment groups, one receiving CRISPR-AAVs by IP administration and the other by IV injection, the latter of which is more clinically relevant.

CRISPR-mediated excision of *Dmd* exon 23 in the dKO mouse resulted in dystrophin restoration in the heart, diaphragm, TA, and triceps muscles. This was highest, at ~4%, in the heart and diaphragm of the dKO mice treated intraperitoneally. Similarly, the IV route of administration resulted in the highest expression of dystrophin in cardiac tissue, although at only ~2%, relative to WT levels. The IP route of administration of the CRISPR-AAVs resulted in higher levels of restoration in the diaphragm, whereas the limb skeletal muscles were more readily targeted following IV delivery of the therapeutic. Favourable correction in the heart independent of the route of administration is promising as cardiomyopathy is a major determinant of death in DMD patients. Differences in the efficiency of delivery to other muscle tissues was likely impacted by the anatomical distribution resulting from the site of injection and delivery of the CRISPR-AAVs to the bloodstream or peritoneal space. Further experiments are required to determine whether a greater number of AAV genomes transduced the heart and diaphragm, or if CRISPR/Cas9-mediated gene editing was more successful in these tissues. Dystrophin analysis of longitudinal sections of the TA muscles obtained from CRISPR-treated mice exhibited distinct within-fibre patchiness of dystrophin expression in both male and female mouse tissues, an as yet unreported consequence of CRISPR-mediated gene editing in

muscle tissue. Despite successful gene editing and dystrophin protein expression, survival of the treated dKO mice and overt dystrophic phenotypes were not improved relative to the control animals.

3.5.1 Application of the dual-cut CRISPR strategy *in vivo*

Several studies have employed the dual sgRNA CRISPR exon excision strategy in dystrophic murine models [431–435,457,458]. A range of AAV serotypes were used in these studies, specifically AAV6 [457], AAV8 [433,434], AAV9 [432,434,435] (which was also employed in our study), and AAVrh74 [458], all of which exhibit tropism for muscle tissue, and favour transduction in the heart [425,459,460]. A number of these studies closely mirror the approach we have taken to assess the utility of CRISPR-mediated *Dmd* exon 23 excision in neonatal mice, but instead employed the *mdx* mouse model which has a mild phenotype and almost normal lifespan [432–435]. Comparable levels of dystrophin restoration to the 4% cardiac expression we have observed in our study were reported in the hearts of treated *mdx* mice assessed at 3-weeks [435], and 8-weeks [432–434] post treatment. This is within the time-frame of the limited lifespan of the dKO mouse, however, sveral of the aforementioned *mdx* studies have reported progressive increases in dystrophin protein levels at 6-months [434], 12-months [433], and 18-months [431] post treatment, with up to ~20% restoration in the heart at the latest timepoint [431]. Similarly low levels of dystrophin protein restoration to ours have been reported in the *mdx* mouse at 8-weeks post CRISPR-AAV treatment, and the progressive accumulation of dystrophin observed at later timepoints has been attributed to the long half-life (>100 days) of the dystrophin protein [461,462]. It has also been suggested that corrected myofibres experience a selection advantage over dystrophin-negative, unprotected fibres that remain sensitive to contractile damage and myonecrosis [201].

The experimental approach taken by Nelson and colleagues closely resembles that of our study, whereby neonatal (P2) *mdx* mice were treated by IV administration via the facial

vein with CRISPR-AAVs for *Dmd* exon 23 excision [433]. By using the *mdx* mouse the authors were able to assess dystrophin expression over time, reporting a progressive increase from 8-weeks to one-year post treatment [433]. Importantly, the levels of dystrophin at the earlier timepoints in these long-term studies were comparable to what we have observed here, suggesting that the severity of the dKO model did not allow sufficient time for dystrophin protein to accumulate in the edited myofibres to reach the threshold required for phenotypic correction and lifespan extension in our study. The importance of reinforced mechanical support by dystrophin positive myofibres in the heart of *mdx* carrier mice was highlighted in a study conducted by Bostick and colleagues, where mosaic expression of dystrophin alleviated cardiopathology [463]. An important note here is that utrophin was also found to be upregulated through compensatory mechanisms in the neighbouring dystrophin negative fibres, which is not possible in the dKO mouse model, but which could be therapeutically relevant to DMD patients treated with sub-optimal dystrophin corrective therapies that result in between-fibre patchiness of dystrophin expression. While it is possible that the dKO mouse model is inappropriate for testing of low dose or low efficacy genetic therapy regimens, the comparatively longer time scale of disease progression in DMD patients might allow for gradual accumulation and restoration of dystrophin through low-dose treatment, which could be clinically achievable.

Other mouse models have been developed with slight modifications to the dKO phenotype that exhibit a more severe phenotype than the *mdx* mouse, but with longer lifespans which may better recapitulate the human phenotype. The Fiona/dKO mouse expresses a full-length human utrophin transgene restricted to skeletal muscle only, through the use of a skeletal muscle-specific promoter/enhancer (i.e. *Acta1*) [443,464,465]. As such, the therapeutic efficacy of a particular treatment in this mouse can be measured with regards to alleviation of the cardiomyopathic phenotype. For example, micro-dystrophin gene therapy was shown to

improve heart failure in the Fiona/dKO mouse in a recent study [464]. Alternatively, an *mdx* mouse line that is heterozygous for utrophin knockout [466], similar to the parental mice in our dKO study, have a relatively normal lifespan but haploinsufficiency of the utrophin gene significantly enhances the skeletal muscle dystrophic pathology [467].

In vivo studies provide a benchmark for the success of neonatal treatment with CRISPR-AAVs for exon excision in dystrophic mice. However, the disease pathology in the *mdx* models described above are significantly different to that of the dKO mouse [431–435,457,458]. The levels of dystrophin protein restoration in our study fell short of the widely accepted threshold of ~15-30% believed to be required for phenotypic correction in *mdx* mice [468,469]. It is thus important to try to determine if this threshold holds true for the dKO model, and to compare the efficacy of our therapy with that of other splice-corrective strategies carried out in these severely dystrophic mice. Notably, while the level of dystrophin restoration observed in this study is minimal, it is still greater than that observed with the FDA-approved exon skipping drug eteplirsen (i.e. treated patients expressed 0.93% dystrophin relative to healthy controls) [153].

3.5.2 The effect of constitutive low levels of dystrophin expression in dKO mice

Two mouse models have been generated to assess the effect of low levels of dystrophin expression on lifespan of severely dystrophic mice [470–472]. van Putten and colleagues created a modified dKO model expressing varying levels of full-length dystrophin as a result of skewed X chromosome inactivation (*mdx*/utrn$^{-/-}$/*Xist*$^{\Delta hs}$) [472]. The authors reported that mice in the group expressing less than the median level of dystrophin (<4% of WT) outlived the dKO controls, with 62% of the mice alive at the end of the 6-month test period [472]. Interestingly, in the survival study cohort, one mouse expressing only 3% of WT dystrophin lived for 17 months, and even marginal increases in dystrophin protein above 4% resulted in significant improvements to the severely dystrophic muscle pathology exhibited by control dKOs [472]. Crucially, these mice

experienced the restorative benefits of full-length dystrophin protein, while the method of genetic correction used in our study results in expression of a BMD-like, partially internally truncated, form of dystrophin, which has been shown to have slightly reduced functionality relative to the full-length protein [473].

A second murine model that naturally expresses low levels of near full-length dystrophin protein harbours a utrophin knockout mutation on an *mdx*^{3cv} background (uko/*mdx*^{3cv}) and consequently express 5% dystrophin [470,471]. One study reports that uko/*mdx*^{3cv} mice survived between 18 and 22 months [471], while another found that these mice succumb by the age of 8-10 weeks [470]. The reason for the disparity between the impact of 5% dystrophin in these two uko/*mdx*^{3cv} lines is still somewhat unclear, however, a major difference between the mice in these two studies was the way in which the utrophin gene (expressing a 3,430aa protein) was knocked out, with one mutation being closer to the N-terminus (at the codon corresponding to amino acid position 230) [443] and the other at the C-terminal end (targeting the codons for amino acids 2851-2875) [442]. The specific utrophin mutation harboured by the dKO mice in our study was based on the uko/*mdx*^{3cv} line where no lifespan extension was observed, and our survival data support this phenotype [470]. Regardless of the disparities in the literature reporting the correlation between dystrophin correction and survival in uko/*mdx*^{3cv} and *mdx*/utrn$^{-/-}$/*Xist*$^{\Delta hs}$ mice, both of these murine models express dystrophin constitutively from birth. In studies such as ours where genetic intervention is carried out postnatally, other factors could have influenced the therapeutic success of genetic therapy in the severe dKO mouse.

3.5.3 Other splice correction strategies assessed in the dKO mouse

As no other study to date has assessed CRISPR-mediated gene editing in these mice, a benchmark for lifespan extension can be taken from studies employing other splice-corrective approaches. Two independent groups reported that PPMO-mediated *Dmd* exon 23 skipping

significantly improved the survival of dKO mice from ~10-weeks to >6-months [445,474]. The PPMO chemistries used in these studies were of previous generations (i.e. $(RXR)_4XB$ and $(RXRRBR)_2XB$ [161], the efficacy of which has since been iteratively improved) and did not result in efficient dystrophin restoration in the heart, however, high levels of correction in the diaphragm are believed to have compensated for this and contributed towards lifespan extension [475]. Whether these findings will be recapitulated in human patients remains to be determined. Novel ASO chemistries, such as tricyclo-DNA, are in the early stages of development to improve cardiac uptake, and have already proved promising in restoring dystrophin expression (up to ~50%) in the hearts of dKO mice [476]. Even so, a major drawback of ASO therapy relative to CRISPR/Cas9 is the requirement for repeat administration.

An alternative splice-corrective therapy to PPMO-mediated exon skipping is the U7 snRNA expression system, which can be delivered using AAV vectors. The advantage of viral-mediated delivery is that a single intervention approach can be utilised, and several serotypes exist that exhibit favourable tropism for cardiac muscle tissue [425,459,460]. Goyenvalle and colleagues applied this strategy for scAAV-mediated delivery of U7 snRNAs to effect *Dmd* exon 23 skipping in dKO mice [444]. Efficient dystrophin restoration was reported in multiple muscle tissues, including the heart and diaphragm, which resulted in a substantial extension of the lifespan of these mice from 10-weeks to one-year [444]. Dystrophin expression ranged from ~50-90% in the muscle tissues analysed, assessed 9-weeks post treatment [444]. While this is a comparable time-fame to the lifespan of the dKO mouse, the levels of dystrophin expression achieved with sustained exon skipping are significantly higher than that for gene editing. AAV-encoded U7/U1 snRNA-mediated exon skipping systems are highly attractive therapeutic strategies as the corrective effects are rapid, requiring only expression of an RNA sequence. Furthermore, transgene expression is accelerated by the use of scAAVs which form dsDNA hairpin structures, thereby bypassing the rate-limiting second strand synthesis step [477,478]. In

contrast, the effects of AAV-mediated CRISPR gene transfer are dependent on second strand DNA synthesis of the viral genome, meaning that there is a ~7-10 day delay between injection of the virus and the manifestation of therapeutic effects [479,480]. Importantly, the turnover of muscle fibres associated with dystrophic pathology means that viral genomes are progressively lost over time, including during this second-strand synthesis lag period [481,482]. Indeed, even with scAAV-expressed U7-snRNA for exon skipping, the effects on dystrophin expression were diminished over time, likely due to muscle turnover and loss of the episomally-maintained vector genomes with each cycle of myonecrosis and repair [444].

3.5.4 Uniformity of dystrophin expression

Analysis of longitudinal sections through CRISPR-treated dKO mouse skeletal muscle tissue revealed an important finding of within-fibre patchiness of dystrophin restoration following CRISPR treatment. This is in accordance with previous hypotheses generated by our group [456,483]. The observed patchiness is consistent with the myonuclear domain hypothesis, whereby each myonucleus within the myofibre syncytium serves its proximal cytoplasmic territory [484]. Data from this thesis are indicative of limited dystrophin protein distribution from CRISPR-corrected myonuclei in the sarcolemma, and mRNA transcripts in the sarcoplasm. This phenomenon is similar to that reported previously in our lab using a genetic model (the *mdx-Xist*$^{\Delta hs}$ mouse) which express varying levels of dystrophin as a result of skewed X-chromosome inactivation [456]. In both cases, myofibres are heterokaryons that contain both dystrophin-expressing, and non-dystrophin-expressing myonuclei. Crucially, in the *mdx-Xist*$^{\Delta hs}$ mouse model, serum miRNA biomarker levels and regenerative pathology was not restored to WT levels, whereas PPMO treatment that resulted in similar expression levels of total dystrophin provided phenotypic correction [456]. This is important as it suggests that dystrophin uniformity is a critical factor in the success of dystrophin restoration therapies, a facet that has been

overlooked in the majority of genetic therapy studies which have focused primarily on total dystrophin levels.

The functionality of muscle relies on co-ordinated contraction and force dissipation along the length of the myofibre, and, as such, incomplete restoration by gene editing could result in non-functional correction. Even with high levels of total dystrophin protein, molecular pathological events could still initiate at unprotected dystrophin-negative regions within the myofibre. The findings from this study may present a major obstacle for the successful implementation of CRISPR therapies for DMD, and for other muscle-wasting disorders. Indeed a threshold of myonuclei may be required to observe functional correction by gene editing therapies in muscle, and whether micro-dystrophin and cell therapy approaches are able to circumvent within-fibre patchiness of dystrophin restoration remains to be determined.

3.5.5 Non-productive gene editing outcomes

The generation of myofibre heterokaryons is inevitable while there is a possibility for non-productive repair events to occur following CRISPR-mediated cleavage of the gDNA (**Figure 3.9A**). Where delivery of either (or both) of the CRISPR components to a given myonucleus fails, the target site remains unedited (**Figure 3.9B**). Conversely, temporally synchronised target engagement of the dual sgRNA CRISPR system potentially enables successful exon excision, and fusion of the flanking intronic regions (**Figure 3.9B**). However, the possibility for several other non-productive editing outcomes also exists. Synchronous cleavage, required for exon removal, has been shown to result in inversion of the excised genomic fragment, or integration of AAV genomic DNA in other studies [433,485,486] (**Figure 3.9B**). Asynchronous cleavage and indel introduction at either (or both) of the sgRNA cut sites renders the target locus refractive to future recognition and cleavage, and the DSBs at individual sgRNA target

sites also provides further opportunities for the incorporation of AAV genomic DNA (**Figure 3.9C and D**).

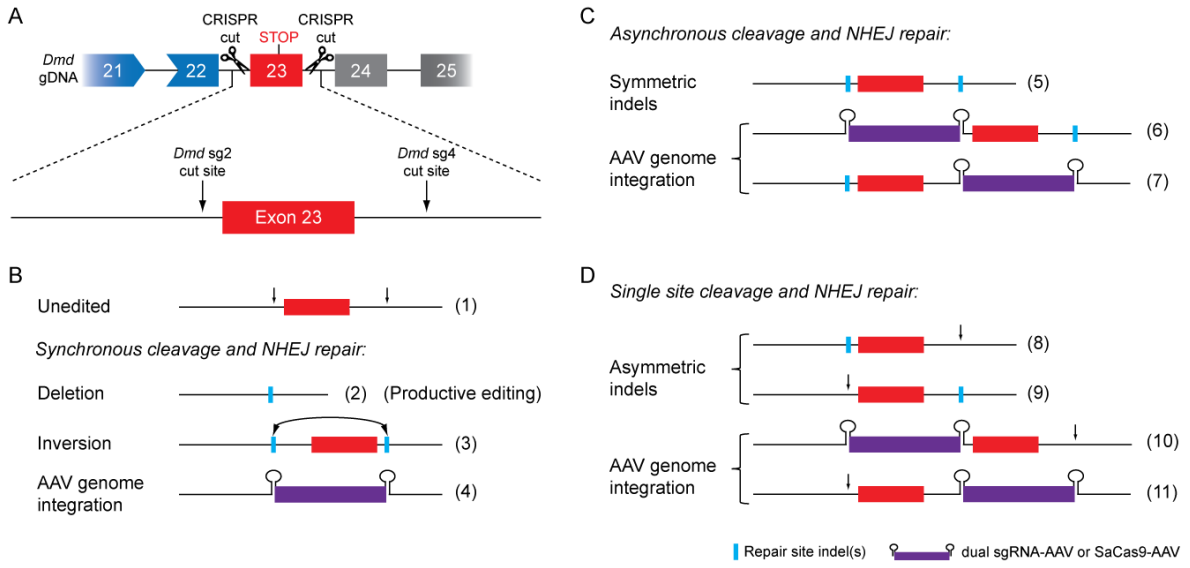


Figure 3.9 Possible editing outcomes resulting from dual sgRNA CRISPR-mediated exon excision.

(**A**) An overview of the dual sgRNA CRISPR approach employed in this study for *Dmd* exon 23 excision, indicating the target sites of *Dmd* sg-2 and -4 within introns 22 and 23 respectively. (**B**) A representation of the unedited target site (1) is shown, as well as editing outcomes resulting from synchronous cleavage and NHEJ repair including the successful removal of exon 23 (productive editing, 2), inversion of the excised genomic fragment (3), and integration of AAV genomic DNA into the dual-cut site (4). Other possible non-productive editing outcomes include (**C**) those resulting from asynchronous cleavage with symmetric indels at the target sites (5), AAV genome integration at either or both of the target sites (6 and 7), as well as (**D**) single site cleavage and NHEJ repair resulting in asymmetric indel introduction (8 and 9), or AAV genomic DNA integration at one of the target sites (10 and 11).

An alternative to entire exon excision is the use of a single-cut CRISPR strategy to disrupt splice accepter sites and putative exonic splicing enhancer motifs in the gDNA. As a result of the indels introduced, the spliceosome is no longer able to recognise these splicing signals, thereby achieving permanent 'exon skipping'. This approach has been utilised in *mdx* mice [198,457,487,488], as well as a canine model of muscular dystrophy [489] with high levels of gene editing success. Importantly, at 8-weeks post treatment of neonatal *mdx* mice, widespread dystrophin expression was observed in the heart, diaphragm and peripheral skeletal muscles [487]. This suggests that the single-cut CRISPR strategy may be superior to the dual-cut approach owing to the simplicity of design and propensity for a greater number of productive editing outcomes. Several limitations to this single-cut approach exist, however, including the possibility for non-productive editing outcomes (i.e. where indels do not disrupt the reading frame, or where the mutation to the splice acceptor site does not impair spliceosome recognition), and AAV genome integration, as well as the inability or difficulty to target for multi-exon 'skipping'.

An extension of the work described in this chapter will include deep sequencing across the target region in the CRISPR-treated dKO mouse tissues to gain a deeper insight into the occurrence of non-productive editing outcomes, relative to the overall editing efficiency of the dual sgRNA CRISPR system. This information would contribute towards understanding the basis for CRISPR-mediated patchy restoration of dystrophin in edited myofibres. El Refaey and colleagues reported few asymmetric editing events by deep sequencing of the individual target sites. However, the chance of co-ordinated cleavage by both sgRNAs is still likely to be lower than the frequency of individual cleavage [458].

Other gene editing tools that do not rely on DSB formation may help to overcome the hurdle of non-productive editing outcomes, particularly the potential for AAV genome integration into the cut site. These include base editing strategies which rely on nicking of a

single strand of the genomic DNA, followed by high fidelity base excision DNA repair [490–493], or CRISPR/Cas13 systems which carry out transient editing at the mRNA level [494]. An exciting new CRISPR tool has recently been developed by the Liu lab, called 'prime editing', which could offer an elegant alternative strategy for rewriting the genetic code at splicing motifs or premature stop codons within the DMD reading frame [398]. This approach enables the introduction of user-defined indels into an extended target region that is less constrained by PAM availability than for the conventional base and RNA editing technologies. Prime editing has recently been applied in cultured cardiomyocytes for *DMD* restoration [493]. Alternatively, non-viral genetic therapy delivery vehicles are currently being developed and have the potential to avoid the risk of viral genome integration into a CRISPR-induced DSB [495]. Such technologies include delivery of Cas9 protein:sgRNA ribonucleoprotein (RNP) complexes facilitated by lipid- or gold-nanoparticles [496], cationic lipids [497,498], or cell-penetrating peptides [499]. Active recruitment of Cas9 into extracellular vesicles has also been demonstrated by the novel and innovative NanoMEDIC approach which has been used for CRISPR-mediated DMD correction in patient-derived iPSCs [500].

3.5.6 Limitations and future directions

The dose of each AAV9 vector used in this study (1×10^{11} of SaCas9-AAV, and 5×10^{11} vg of dual sgRNA-AAV, a dose which is equivalent to ~5E13 and 2.5E14 vg/kg respectively) was comparable to those described in published *mdx* studies at the time that this study was initiated [432,434,435]. However, it has since emerged that both the dose of AAV as well as the ratio of sgRNA:Cas9 are important determinants of gene editing success. The effect of increasing the dose of vectors carrying the CRISPR machinery was demonstrated by Bengtsson and colleagues, where the authors assessed co-excision of 'hotspot' exons 52 and 53 in the *mdx*^{4cv} mouse (which harbours a nonsense mutation within exon 53) [457]. Dystrophin restoration was achieved in the heart at both low and high doses of CRISPR-AAV6 (low dose: 1×10^{12} vg of

each vector, or high dose: 1×10^{13} vg of SpCas9-AAV and 4×10^{12} of sgRNA-AAV), although widespread skeletal muscle restoration was only observed at the higher dose [457]. The implications of these findings are that the dose regimen used in the present study may have been sub-optimal. The dose selected was constrained by limitations of the concentration of the virus, as well as the maximum volume that can be injected into neonatal pups via the facial vein.

sgRNA levels have also been determined to be a limiting factor for long-term systemic DMD correction *in vivo* [431,485,501]. Hakim and colleagues attributed this limitation to the preferential depletion of sgRNA-AAV genomes relative to Cas9-AAV, and overcame this by increasing the ratio of sgRNA-AAV:Cas9-AAV from 1:1 to 1:3 [431]. A ratio of 1:5 of SaCas9:sgRNA was used in the study reported in this thesis, however, Zhang and colleagues have recently reported an optimal dose of 1:10 of the Cas9:sgRNA vector using the single-cut approach to gene editing of the dystrophin reading frame in *mdx* mice [501]. We performed a follow-up study using a 1:10 ratio of SaCas9:dual sgRNA vector (~80 µl of a mixture of 1×10^{11} vg of SaCas9-AAV, and 1×10^{12} vg of dual sgRNA-AAV) in P4 dKO pups, however, the lifespan of these mice was only ~6-weeks, and marginal levels of the ΔExon23 *Dmd* transcript was detected in the heart (<1%, data not shown). The reason for the ineffectual result from this study have not yet been explored in detail, however, a number of possibilities exist. These include the inability to deliver a comparable dose to that used for the 1:10 ratio in the Zhang study [199] (due to insufficient viral titres), as well as technical difficulties experienced with regards to ejection of the injected solution out of the IP space via the injection site, both of which could have resulted in sub-optimal dosing. Even so, we would expect to observe similar results to those of the previous 1:5 vector ratio experiments, thus a more likely explanation is that the higher dose of AAV could have resulted in additional unknown side-effects, particularly in the context of the severely affected dKO mouse. The feasibility of using high

dose AAV regimens in patients has been called into question as important safety concerns have recently resurfaced, with the deaths of children participants in a clinical trial for X-linked myotubular myopathy (NCT03199469) [502], and another child participant in a separate AAV trial for mucopolysaccharidosis type IIIA (NCT03612869) [503]. This limitation could be circumvented by the use of scAAVs which form dsDNA hairpin structures, thereby bypassing the rate-limiting second strand synthesis step and accelerating transgene expression [477,478]. Zhang and colleagues observed a 70-fold improvement in editing efficiency in *mdx* mice with sgRNAs expressed from scAAV relative to conventional AAV at the same dose when applying the 1:10 optimised AAV-Cas9:AAV-sgRNA ratio [501]. While the SaCas9 gene is too large, delivery of the smaller sgRNA component(s) using scAAV vectors could significantly improve the editing outcomes of CRISPR *in vivo*. This could assist with managing the delicate balance between optimal vector dose and safety for bringing these genetic therapies to the clinic.

There is a need for DMD genetic therapies not only to target the skeletal and cardiac muscle, but also the satellite cells which are the precursor nuclei contributing dystrophin to future regenerating fibres. Two studies have reported gene editing in satellite cells mediated by AAV9-encoded CRISPR/Cas9 [435,504], although it has been described elsewhere that these cells are refractory to AAV transduction [505]. Similar CRISPR applications for DMD therapy development have employed the muscle-specific CK8 promoter to circumvent expression of CRISPR machinery in non-target cells and avoid off-target mutagenesis [457]. This promoter, however, exhibits negligible activity in satellite cells [506,507]. Further studies are necessary to elucidate the potential for genetic correction in satellite cells. In the study described in this chapter, the AAV9 serotype was employed for delivery of CRISPR components driven by ubiquitously-expressing promoters (CMV, and H1 and U6 snRNA), however, as the levels of gene editing were found to be relatively low and the lifespan of the treated dKO mice was not improved, the targeting capacity for satellite cells was not investigated.

3.5.7 Conclusions

Important future developments are required to increase the overall editing efficiency of the CRISPR-AAV dual-cut strategy employed in this study, including optimisation of delivery to a broad range of muscle tissues, particularly the heart and diaphragm, and reducing the likelihood of non-productive editing outcomes. In order for this therapeutic strategy to be translated into clinical applications, testing the safety and efficacy in larger animal models will be of paramount importance. Furthermore, treatment with CRISPR-AAV results in a patchy distribution of dystrophin at the sarcolemma which may be insufficient to correct the disease pathology and may explain the lack of lifespan extension in treated dKO mice.

4 Results II - Investigation of the pro-myogenic potential of extracellular vesicles and secreted proteins derived from differentiating myotubes

4.1 Introduction

The pathology of DMD is highly complex and arises as a result of both genetic and cellular insults to muscle tissue. The major advantage of gene editing for the correction of DMD is that it is a single intervention strategy, however, if the success of this approach is sub-optimal, repeat administration is an unlikely possibility due to safety concerns both with regards to off target genome editing, as well as immune tolerance to AAV/Cas9-derived epitopes. An additional therapeutic hurdle for neuromuscular diseases, such as DMD, is the progressive fibro/fatty degeneration of muscle which may be an impediment to delivery and the even distribution of such genetic therapies [448]. Notably, efficacy in patients is likely to be lower than in animal models. Furthermore, incomplete restoration of the dystrophin gene in all myonuclei, and non-uniform expression along the length of myofibres (discussed in chapter 1), is a potential additional limitation to AAV-delivered CRISPR-mediated genetic correction. It is thus important to investigate other therapeutic avenues that focus on addressing the cellular nature of DMD disease, and which can be applied in combination with genetic therapies to bolster and sustain therapeutic efficacy throughout a patient's life.

Extracellular myomiRs have been shown to be released from regenerating myofibres in a selective manner [234,244,245], and can be detected in the serum of DMD patients, and dystrophic animal models [233,234,238,238–242]. The majority of these extracellular myomiRs are detectable in non-vesicular fractions, although it remains to be determined whether non-vesicular miRNAs can be actively taken up into recipient cells, at least the uptake of high density lipoprotein (HDL)-associated myomiRs in an *in vitro* model of cardiovascular disease was found to be

ineffective [254]. Several studies have, however, demonstrated the cell-to-cell functional transfer of pro-myogenic factors, which could include myomiRs, in extracellular vesicles (EVs) between donor and recipient C2C12 murine myoblast cells [306,508], C2C12 myotubes and neuronal cells (NSC-34) [509], activated myogenic progenitor cells (satellite cells) and myofibroblasts [307], human skeletal myoblasts and human adipose-derived stem cells [308], and human mesenchymal stem cells (MSCs) and mouse myofibres *in vivo* [309]. The transferred EVs resulted in enhanced myogenic differentiation in recipient cell cultures [306,308,510], or in improved wound repair *in vivo* [309]. However, these studies employed methods of EV isolation that are known to co-purify other non-vesicular proteins and soluble factors, such as ultracentrifugation (UC) [308,510] and commercial exosome isolation kits [306,309], which could have confounded the pro-myogenic effects observed. For instance, polymer precipitation methods of EV purification have been shown to co-purify vesicle-free miRNAs from rat plasma [511], and non-vesicular extracellular Argonaute-2 (AGO2) complexes from the secretome of MCF-7 breast cancer cell cultures [512]. Furthermore, it has been suggested that the use of EV precipitation techniques has led to the misattribution of pro-angiogenic and regenerative cell signalling effects as a result of the co-purification of non-vesicular soluble factors expressed by mesenchymal stem cells [513]. Ultracentrifugation results in inconsistent yields and insufficient purity as a result of co-sedimentation of some soluble proteins with EVs [514]. Conversely, methods that apply a density gradient during the ultracentrifugation step outperform polymer precipitation kits and conventional ultracentrifugation with regards to the purity of isolated EVs [512].

Our group has previously developed a scalable method for the isolation of highly pure EVs within the size range of exosomes that combines ultrafiltration with size exclusion liquid chromatography (UF-SEC) [515]. This method results in low recovery but high specificity (subtypes of EVs in the size range of exosomes and small microvesicles, with low amounts of non-EV components), as emphasised in the most recent minimal guidelines from ISEV

(MISEV) [267]. In contrast, polymer precipitation and ultracentrifugation are high recovery, low specificity methods (isolating mixed EVs and non-EV components). A number of measures can be used to determine the purity of purified EVs, including light scattering methods (e.g. by nanoparticle tracking analysis, NTA) to determine particle count and size range, and Western blot for known exosome markers (such as the transmembrane tetraspanin CD9, and a cytosolic accessory protein of the ESCRT pathway Alix/PDCD6IP). Furthermore, excluding contamination with other non-vesicle associated intracellular compartments, such as calnexin (CANX), is important. Even so, it is difficult to be certain of the unique biogenesis pathway resulting in the production of the purified vesicles (i.e. distinguishing between exosomes versus other small microvesicles) using these quality control measures. As such and in line with the MISEV2018 guidelines [267], the pool of nanoparticles in the size range of exosomes (<200 nm) used in experiments described in this thesis will be hereon referred to as EVs.

C2C12 cells are a robust *in vitro* model of myogenic differentiation. C2C12 myoblast to myotube differentiation is initiated by serum withdrawal at the time at which cells are confluent (or almost confluent). The secretome is greatly altered during differentiation [516,517], and factors released by differentiating C2C12 myotubes condition the culture media with a heterogenous pool of biomolecules. The overarching aim of this study was to determine whether EVs and/or non-vesicular soluble factors derived from differentiating myotubes possess pro-myogenic properties that can be transferred to recipient myoblast cultures.

The work described in this chapter contributes towards understanding the innate and transferrable pro-myogenic properties of myotube-derived EVs, as well as non-vesicular proteins and soluble factors. Recently, there has been intense research interest into developing methods for cell-free, non-viral delivery of therapeutics. The use of EVs could enable pro-myogenic paracrine signalling without the need for cellular engraftment, and thereby circumvent issues associated with immune reactivity and cell viability. EVs with intrinsic pro-

myogenic properties could be used for the purposes of regenerative medicine for muscle-wasting diseases. For example, EVs have been shown to contain miR-206 [233] which is capable of controlling the balance between myofibre regeneration and fibrotic degeneration [249]. The muscle architecture in DMD patients could be improved by the administration of EVs containing pro-myogenic factors such as miR-206. Trafficking of therapeutic molecules using exosomes also offers several other advantages over viral-mediated genetic therapy delivery. These include flexibility in cargo size and biotype (i.e. DNA, RNA, and protein), stability within the extracellular environment, and the potential for targeting and surface receptor engineering. In the first instance it is important to elucidate the role of the myotube-derived secretome on myogenic differentiation.

4.2 Hypothesis and aims

I hypothesise that differentiating C2C12 myotubes condition the extracellular milieu with pro-myogenic biomolecules which, when contained within EVs, function in a paracrine manner to promote myogenic differentiation in recipient myoblasts.

This was investigated with the following aims:

- To measure the effect of the following on C2C12 myogenic differentiation:
    - Inhibition of EV uptake and release using small molecules and RNA interference, hypothesising that these will result in a decrease in myogenic differentiation.
    - Transfer of vesicular and non-vesicular fractions obtained from myotube-conditioned-media, hypothesising that the vesicular fraction, and not non-vesicular biomolecules, will enhance myogenic differentiation.

4.3 Materials and methods

4.3.1 Cell culture

C2C12 *Mus musculus* myoblast cells (obtained from ATCC, cat no. CRL-1772) were maintained in Growth Medium (GM) which consists of DMEM (Thermo Fisher Scientific, MA, USA) supplemented with 15% FBS and 1% Antibiotic-Antimycotic (PSA: Penicillin, Streptomycin and Amphotericin B; Thermo Fisher Scientific). These cells originate from the thigh muscle of a C3H mouse following crush injury [518], and myoblast to myotube differentiation is initiated by switching to low serum Differentiation Medium (DM, referred to as DM day 0 or DM0), consisting of DMEM supplemented with 2% horse serum (Thermo Fisher Scientific) and 1% PSA (Thermo Fisher Scientific).

4.3.2 Immunofluorescence staining of myotubes

A biomarker of myofibre maturity is the expression of myosin heavy chain (MHC), which can be detected by immunofluorescence (IF) staining of fixed myotubes. MHC IF experiments were carried out in 24 well multiplates. Cell culture media was removed and cells were washed gently to prevent cell detachment with 0.5 mL of 1× PBS (Thermo Fisher Scientific). Cells were fixed by incubation in 0.2 mL of 4% paraformaldehyde in PBS (Santa Cruz, CA, USA) for 10 mins at room temperature. Cells were washed three times with 0.5 mL PBS, and permeabilised by incubation in 0.2 mL of 0.25% Triton X-100 (AppliChem, Darmstadt, Germany) diluted in PBS for 15 mins at room temperature. Cells were washed three times with 0.5 mL PBS, and non-specific antigens blocked with 0.2 mL of 5% Bovine Serum Albumin (BSA, Sigma-Aldrich, MO, USA) in PBS for 30 mins at room temperature. MHC was stained for using a 1:20 dilution of MYH1E (MF 20) primary antibody (the monoclonal antibody, developed by Fischman, D.A. at Weill Cornell Medical College, was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at

The University of Iowa, Department of Biology, Iowa City, IA 52242, USA) in 0.2 mL 5% BSA (in PBS) for 2 hrs at room temperature. Notably, this antibody does not distinguish between different MHC isoforms. Unbound primary antibody was removed by washing cells three times with 0.5 mL PBS. Cells were then incubated in a 1:500 dilution of 2 mg/mL Goat Anti-Mouse IgG H&L (Alexa Fluor 488) secondary antibody (Thermo Fisher Scientific) and 1:5,000 dilution of 10 mg/mL Hoechst 33258 Pentahydrate (Bis-Benzimide) nuclear stain (Sigma-Aldrich) in 0.2 mL of 5% BSA for 1 hr at room temperature. From this point onwards, plates were stored wrapped in aluminium foil to shield from light as the secondary antibody is conjugated to a fluorophore that is sensitive to photobleaching. Cells were washed three times with 0.5 mL PBS and then cells were left in 0.5 mL of PBS. Plates were stored at 4°C, and imaging was carried out using a fluorescence microscope. To minimise unconscious imaging bias, images from four fields of view distributed in a similar manner to points of the compass were taken, avoiding the well edges and centre where myogenic differentiation is typically enhanced. Antibody information is described in **Table 2.2** and **Table 2.3**.

Myogenic differentiation was quantified in this study using three different parameters. The first is the *MHC+ area* which measures the total MHC positive stained area relative to the area of the representative field of view, calculated by running a custom script in ImageJ. This measure is conveniently automated and therefore uninfluenced by user bias. Two other standard measures, the myogenic and fusion indices, were also utilised to further assess the extent of myogenic differentiation and the maturity of myofibres respectively.

The following are definitions of the measures used:

- *MHC+ area* is the total MHC positive stained area relative to the total image area.
- *Myogenic index* is the percentage of nuclei contained within all MHC+ cells relative to the total number of nuclei.
- *Fusion index* is the percentage of nuclei contained within myotubes that have three or more nuclei, relative to the total number of nuclei.

4.3.3 Small molecule inhibition of EV secretion and uptake

C2C12 cells were seeded at a density of 1×10^5 cells in 0.5 mL of GM per well in a 24 well multiplate. Cells were treated at the time of switching to DM with 10 µM of GW4869 (Sigma-Aldrich), an inhibitor of neutral sphingomyelinase 2 (NSMase2) in dimethyl sulfoxide (DMSO, Sigma-Aldrich) to impair EV biogenesis [519], or 10 µM of heparin sulphate in water (Sigma-Aldrich) to prevent EV uptake or re-uptake [520] within the confined cell culture system. Heparin is a negatively charged glycosaminoglycan that has been shown to inhibit EV uptake in HEK293T cells to near complete levels [520]. Cultures treated with DMSO and DM only were included as controls. Cells were fixed and immunofluorescence staining for MHC carried out 48 hrs post treatment to assess the effect of the small molecules on myogenic differentiation.

4.3.4 siRNA knock down of exosome biogenesis markers *Rab27a* and *Rab27b*

C2C12 cells were seeded at a density of 1×10^5 cells in 0.5 mL of GM per well in a 24 well multiplate. Cells were treated with 100 µM of ON-TARGETplus Mouse *Rab27a* (11891) or *Rab27b* (80718) siRNA SMARTpools (Dharmacon, CO, USA) using RNAiMAX (Thermo Fisher Scientific), as according to the manufacturer's instructions. Transfection complexes were prepared in DMEM containing no supplements (Thermo Fisher Scientific). Target regions of the ON-TARGETplus SMARTpool siRNAs are listed in **Table 4.1**.

Table 4.1 ON-TARGETplus SMARTpool siRNAS targeted to *Rab27a* and *Rab27b*

ID	Target sequence (5' to 3')
siRab27a	CGGAUGGAGAUUACGAUUA
	CAGGAGAGGUUUCGUAGCU
	GUACAGAGCCAAUGGGCCA
	GGGCAUUGAUUUCAGGGAA
siRab27b	GCAGAUUAGAAGCUAGUAU
	GGGAAUAGAUUUUCGGGAA
	GGAAGUCAAUGAACGGCAA
	GCAGAGUAGUCAUAGUGUU
Non-targeting siRNA	UGGUUUACAUGUCGACUAA
	UGGUUUACAUGUUGUGUGA
	UGGUUUACAUGUUUUCUGA
	UGGUUUACAUGUUUUCCUA

Treatment with 100 µM of ON-TARGETplus Non-targeting Pool (Dharmacon) was included as a negative control. The Dharmacon SMARTpools are a single reagent comprised of a mixture of 4 siRNA with proprietary target sequences based on technology that enhances on-target specificity and favours antisense strand uptake into the RISC complex. Levels of knock down of *Rab27a* and *Rab27b* mRNA were determined by RT-qPCR using Power SYBR Green MasterMix (Thermo Fisher Scientific) (refer to **Section 2.6**). *Rab27a* and *Rab27b* transcript levels were detected using gene-specific primers and normalised to the stable murine reference gene *Rplp0*, and data were scaled such that the mean of the control group was returned to a value of 1. Primer sequences are listed in **Table 4.2**.

Table 4.2 Primer sequences for RT-qPCR of *Rab27a*, *Rab27b*, and *Rplp0* reference gene

ID	Sequence (5' to 3')
qRab27a-Fwd	CGACCTGACAAATGAGCAAAG
qRab27a-Rev	CCTCTTTCACTGCCCTCTG
qRab27b-Fwd	CAGACCTGCCAGACCAAAG
qRab27b-Rev	AGCGTTTCCACTGACTTCTC
qRplp0-Fwd	AAGCAAAGGAAGAGTCGGAG
qRplp0-Rev	CCAGACCGGAGTTTTAAGAGAAG

Cycling conditions are described in **Section 2.6.2**.

4.3.5 Crude fractionation of myotube conditioned media

C2C12 cells were seeded at a density of 4.7×10^5 cells in 100 mm plates in 10mL of GM and switched to DM 48 hrs later. The DM was changed 72 hrs later and the conditioned media (CM) collected after a further 72 hrs from highly differentiated myotubes. The CM was centrifuged at 300 *g* for 5 mins at 4°C to remove extracellular debris, but without lysing any whole cells present in the extracellular matrix to avoid the release of intracellular vesicles and proteins. The supernatant was transferred to a fresh 50 mL tube, and centrifuged at 2,000 *g* for 10 mins at 4°C to pellet the remaining larger, non-vesicular particulate matter. This was then passed through a 0.22 µm filter. The first of four treatment fractions was obtained from the crude unfractionated CM, before separating the remaining mixture by spin filtration using sequential molecular weight (Mr) cut-off filters. Some crude CM was kept aside for CM only control treatments, and 6 mL of the remaining crude CM was passed through a 300 kDa Vivaspin concentrator (Sartorius AG, Göttingen, Germany) by centrifugation at 3,000 *g* for ~15 mins (or until the retentate was ~300 µl). The EV-containing fraction (CM-EVs) was obtained from the retentate and the volume of the flow-through (FT) was increased to 3 mL in DM before filtration through a 100 kDa Amicon Ultra-15 centrifugal unit (MilliporeSigma, MA, USA) by centrifugation at 3,000 *g* for ~10 mins. The EV-depleted fraction containing particles and soluble factors of 100-300 kDa in size (e.g. AGO2 bound to miRNAs [255,512]) was obtained from the retentate and the volume of the FT was increased to 3 mL in DM once again before filtration through a 10 kDa Amicon Ultra-15 centrifugal unit (MilliporeSigma) by centrifugation at 3,000 *g* for ~10 mins. The EV-depleted fraction containing particles and soluble factors of 10-100 kDa in size (e.g. growth factors and cytokines) was obtained from the retentate. In summary, treatment fractions obtained by spin filtration included: (i) crude unfractionated CM, (ii) CM containing EVs and other particles >300 kDa in size, and CM

depleted of EVs with particles that are (iii) 100-300 kDa, or (iv) 10-100 kDa, in size. Treatment of recipient C2C12 myoblasts was performed immediately following isolation of CM fractions.

4.3.6 Transfer of ultrafiltration (UF)-derived CM fractions to C2C12 myoblasts

C2C12 myoblasts were seeded at a density of 1×10^5 cells in 0.5 mL of GM per well in a 24 well multiplate, and cultured for 2 days. Two experiments were run in parallel to determine the effect of treatment with crude CM fractions on (i) myogenic differentiation and (ii) cell proliferation. A 50:50 mixture comprised of 0.5 mL of treatment fraction (described in **Section 4.3.5**) and 0.5 mL of fresh DM was prepared to replenish the nutrients depleted by the metabolic activity of the donor cell cultures, and filtration. GM was removed from the confluent myoblasts, cells washed gently with PBS, and 50:50 mixes of CM fractions and DM added to the cultures. An equal volume of DM (1 mL) was added to cultures to be included as a DM only control. Myogenic differentiation was assessed 72 hrs post treatment by IF staining for MHC (as described in detail in **Section 4.3.2**). Cell proliferation was assessed 24 hrs post treatment using the Click-iT EdU Alexa Fluor 555 Imaging Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The Click-IT EdU assay measures incorporation of EdU (5-ethynyl-2′-deoxyuridine), a modified thymidine analogue, into newly synthesised DNA. The cultures were pulse incubated with 10 µM EdU for 1 hr prior to detection. This assay utilises click chemistry (azide-alkyne cycloaddition) [521], whereby a copper-catalysed covalent reaction conjugates an Alexa Fluor 555-azide dye to the alkyne-containing EdU incorporated into newly synthesised genomic DNA during S-phase. The advantage of this assay over other commonly used proliferation assays, such as BrdU, is that it can be carried out under mild fixation and permeabilisation conditions, meaning it is compatible with the MHC IF staining protocol (described in **Section 4.3.2**). The p*roliferation index* was calculated as the percentage of nuclei positively stained by Alexa Fluor 555 using Click-iT EdU chemistry (i.e.

the nuclei that are undergoing, or recently underwent DNA synthesis during S-phase), relative to the total number of nuclei.

4.3.7 Pure isolation of extracellular vesicles by ultrafiltration size exclusion liquid chromatography (UF-SEC)

Ultrafiltration-Size Exclusion Liquid Chromatography (UF-SEC) was employed to isolate pure EVs from (i) C2C12 myoblast-conditioned cultures, and (ii) myotube-conditioned cultures (referred to hereon as MB-EVs and MT-EVs respectively). UF is a mechanical method of biomolecule separation, making use of filters with polymeric membranes containing well-defined pore sizes. Fluid pressure drives smaller molecules through the pores, and simultaneously retains larger molecules. Tangential flow filtration (TFF) is a fast and efficient means to achieve UF, while also concentrating the supernatant, and was utilised in this study prior to SEC, which was then carried out to further separate molecules based on a broader size range. This method has been developed by our group for the pure isolation of extracellular vesicles devoid of ribonucleoprotein complexes [515].

4.3.7.1 Collection of C2C12 myoblast and myotube conditioned cell culture media

C2C12 myoblasts were seeded at a density of 4.7×10^5 cells in ten 100 mm plates (Thermo Fisher Scientific, MA, USA), each with 20 mL of GM. For the collection of myoblast-conditioned media (i), cells were switched to serum-free media (DMEM without any supplements) 4 days later (referred to as GM day 4, GM4) after removing the culture media and washing the cells gently with PBS to remove residual EVs from the FBS-containing GM. Serum-free media is typically used in EV isolation experiments to avoid co-purification of extraneous particles contained within serum. DMEM containing no supplements was added to the cells, which were then cultured for a further 24 hrs to allow for myoblast-specific EV release. The ~200 mL of CM was harvested by transferring into eight 50 mL tubes (Thermo Fisher Scientific), which were centrifuged at 300 *g* for 5 mins at 4°C, the supernatant then

transferred to fresh tubes, and centrifuged again at 2,000 *g* for 10 mins at 4°C. The supernatant was pooled and filtered using a 0.22 µm filter unit. The CM was kept at 4°C until the ultrafiltration (UF) step was carried out by TFF (described in **Section 4.3.7.2**).

For the collection of myotube-conditioned media (ii), cells were switched to serum-free media (DMEM only) at day 5 when they reached cellular confluency, referred to as DM day 0 (DM0), to initiate myoblast to myotube differentiation. The DMEM was replaced at DM2, and CM harvested from myotubes at DM2 and DM4 in the same manner as described above for the myoblast-derived CM. Protein lysates from the EV producer C2C12 cells of GM4, DM2 and DM4 cultures were obtained by trypsinisation with Tryp-LE Express Enzyme (Thermo Fisher Scientific), and lysis in RIPA buffer (**Section 2.7**). Lysates from cells were included as controls for Western blot characterisation of EVs (described further in **Section 4.3.8.2**).

4.3.7.2 Ultrafiltration of myoblast- and myotube-CM by tangential flow filtration (TFF) and spin filtration

Myoblast- and myotube-derived CM was processed in the same manner from hereafter. CM was separated from particles <10 kDa in size, and concentrated from ~200 mL to <15 mL by tangential flow filtration (TFF). This volume was further concentrated to <2 mL using a 10 kDa Amicon Ultra-15 centrifugal unit (MilliporeSigma) and centrifugation at 2,300 *g* (typically for ~30 mins) as the SEC column used has a maximum sample loop capacity of 2 mL.

4.3.7.3 Size exclusion liquid chromatography

Size exclusion liquid chromatography (SEC) was carried out using the ÄKTA Prime instrument (GE Healthcare, Pollards Wood, UK), equipped with an ultraviolet flow cell. The concentrated CM was loaded onto a column packed with Sepharose 4 Fast Flow (GE Healthcare, IL, USA). Eluates were measured at 280 nm absorbance to produce a chromatogram of the components within each fraction. The volume of each EV-containing

fraction was set at 0.5 mL, and fractions 5-10 (EV-containing) were pooled prior to downstream analysis to obtain MB- and MT-EVs. The total volume was concentrated to ~2 mL by spin filtration using an Amicon Ultra-2 Centrifugal Filter Unit with Ultracel-100 membrane (MilliporeSigma, MA, USA). The non-vesicular protein fraction volumes were each collected in 2 mL of PBS, and were used without further processing in downstream experiments.

4.3.8 Characterisation of UF-SEC purified EVs

4.3.8.1 Nanoparticle tracking analysis (NTA)

Nanoparticle tracking analysis (NTA) is a method used to determine the size distribution and particle count of EVs within a solution. A camera captures a video of the particles moving by Brownian motion through the path of a laser beam, and light scattering is used to determine the hydrodynamic diameter of individual particles. NTA was performed to determine the concentration and size range of EVs obtained by UF-SEC using the NanoSight NS500 instrument (Malvern Instruments, Worcestershire, UK), and NTA 2.3 analytical software. Particles were counted and a mean of triplicate 30 sec recordings was obtained using an automated script. Measurements were analysed using the following settings: the camera level was set at 14, and the detection threshold set at 5. All other post-acquisition settings were automated. Samples were diluted 1:2,000 for count detection within the range of 2×10^8 to 2×10^9 particles per mL.

4.3.8.2 Western blotting

Protein concentration was determined using the Micro BCA Protein Assay Kit (Thermo Fisher Scientific, MA, USA) as according to the manufacturer's instructions. A total of 10 µg of protein was added per well, and Western blotting carried out as described in **Section 2.7**. Briefly, membranes were blocked for 1 hr at RT in Intercept (PBS) Blocking Buffer (LI-COR, NE, USA). Primary and secondary antibody incubations were carried out in Intercept Blocking

Buffer supplemented with 0.1% Tween-20 (Sigma-Aldrich, MO, USA) overnight at 4°C, or for 1 hr at room temperature respectively. Blots were washed for 3×10 mins with PBS-T after antibody incubations. Blots were probed for two typical human cell-derived exosome markers, namely Alix (PDCD6IP, 96 kDa), and CD9 (26 kDa), as well as the cell-specific marker calnexin (CANX, 90 kDa). HRP-conjugated secondary antibodies were used for visualisation using chemiluminescence. Antibody information is described in **Table 2.2** and **Table 2.3**.

4.3.9 Transfer of EVs to C2C12 myoblasts

C2C12 myoblasts were seeded at a density of 1×10^5 cells in 0.5 mL of GM per well in a 24 well multiplate, and cultured for 2 days. The supernatant was removed, and 0.5 mL of DM supplemented with 0.1 mL of PBS containing MB-EVs (obtained from GM4 cultures) or MT-EVs (obtained from DM4 cultures) was added in doses ranging from 2×10^2 to 2×10^{11} particles/mL. These doses were selected to represent the physiological range of EVs produced by C2C12 cells [306], as well as higher concentrations which are typically used in *in vitro* EV transfer experiments [251,522]. Another measure of EV dose is protein content, however, this was not used here as myoblast and myotube cultures have greatly altered secretory profiles [245], and EV particle count was believed to be a more relevant measure. Cultures treated with DMEM supplemented with PBS only were included as controls. MHC IF was carried out 72 hrs post treatment (as described in **Section 4.3.2**) to determine the effect of EV dose, as well as cellular origin (i.e. myoblast or myotube cultures), on myogenic differentiation.

4.3.10 Transfer of non-vesicular factors from C2C12 myotubes isolated by UF-SEC

C2C12 myoblasts were seeded at a density of 4.7×10^5 cells in ten 150 mm plates (Thermo Fisher Scientific, MA, USA), each with 20 mL of GM. Cells were grown in GM for 4 days, DM for 2 days, and Opti-MEM I Reduced Serum Media (Thermo Fisher Scientific, MA, USA) for 2 days. The same methods as described in **Section 4.3.7.2 and 4.3.7.3** were employed to

isolate the non-vesicular fractions (i.e. the secreted protein) from C2C12 myotubes by UF-SEC. C2C12 myoblasts were seeded at a density of 1×10^5 cells in 0.5 mL of GM per well in a 24 well multiplate, and cultured for 2 days. Each protein fraction (#1-15) was transferred to C2C12 cultures at the time of switching to DM at a final concentration of 1 µg/mL. Cells were fixed at DM1, DM2, and DM3 to assess the level of myogenic differentiation by MHC IF (described in **Section 4.3.2**).

4.3.11 Treatment of C2C12 cultures with BSA, insulin, and transferrin

For treatment of differentiating myoblast cultures with non-specific protein (i.e. bovine serum albumin, BSA), insulin or transferrin, C2C12 cells were seeded at a density of 1×10^5 cells in 0.5 mL of GM per well in a 24 well multiplate. Cells were treated at the time of switching to DM (i.e. when cells are at 100% confluency in order to promote efficient myogenic differentiation) with 5 µg/mL of BSA, or two different doses of insulin or transferrin (i.e. 1 and 10 µg/mL). Cells were fixed at DM2 to assess the level of myogenic differentiation by MHC IF (as described in **Section 4.3.2**).

4.4 Results

Our group has previously shown that pro-myogenic factors, such as myomiRs, are detectable in myotube-derived EVs as well as the non-EV-associated extracellular milieu [233,234,523]. It has also been suggested previously that encapsulation of these factors within EVs is required for functional transfer to recipient cells [306,308,309,312,510]. Other studies have, however, employed ultracentrifugation or commercially available polymer precipitation kits to purify EVs, both methods which result in high yields, but low specificity [306,308,309,312,510]. Furthermore, no study to our knowledge has directly compared the effects of myoblast- versus myotube-derived EVs on myogenic differentiation, or tested the protein fractions for pro-myogenic effects. As such, this chapter aims to explore the role of myotube-conditioned media on the process of myogenic differentiation using the C2C12 model of myogenic differentiation.

4.4.1 Small molecule inhibitors of exosome release and EV uptake impairs myogenic differentiation

In order to determine whether EV release and uptake is required for the transfer of pro-myogenic signals within the same culture system, exosome release via the ceramide-dependent pathway was inhibited using the nSMase2 inhibitor GW4869, and EV uptake blocked using heparin, which interacts with EV surface molecules required for binding to target cells. C2C12 myoblasts were cultured for two days in GM, after which myogenic differentiation was initiated by switching to DM (**Figure 4.1A**). The timepoint at which myoblast to myotube differentiation is initiated is referred to from this point on as DM day 0 (DM0). Myoblasts were treated with either 10 μM GW4869 or 10 μg/mL heparin at DM0, and myogenic differentiation assessed 48 hrs later by IF staining for expression of MHC (**Figure 4.1A**). Untreated (DM only) and DMSO-treated cultures were included as negative controls.

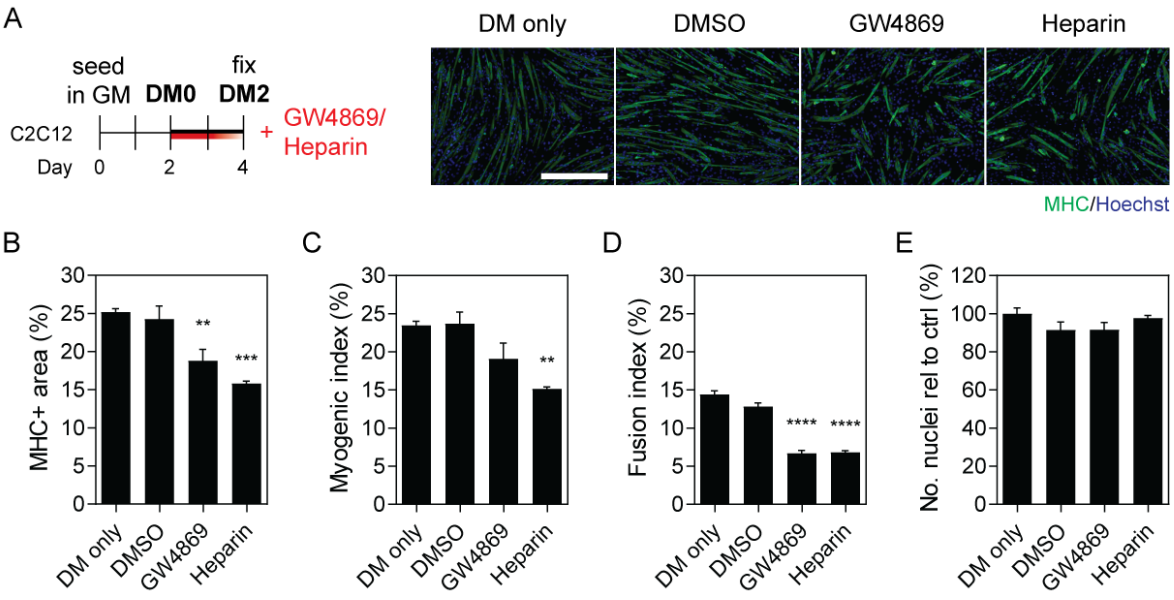


Figure 4.1 Inhibition of EV release and uptake impairs myoblast differentiation.

C2C12 cells were cultured in GM for two days and then switched to DM for two days. Cultures were treated with 10 µM GW4869 (exosome biogenesis inhibitor), or 10 µg/mL heparin (EV uptake inhibitor), at the time of switching to DM. Untreated (DM only) and DMSO-treated cultures were included as negative controls. (**A and B**) Myogenic differentiation was assessed by immunofluorescence (IF) staining for Myosin Heavy Chain (MHC), and quantified by calculating the (**C**) myogenic, and (**D**) fusion indices. (**E**) The total number of nuclei per representative field of view are shown as a percentage relative to the control group. All microscopy images were taken at 10× magnification. Scale bar represents 400 µm. All values are means shown + SEM from four fields of view obtained in one independent experiment, representative of two independent experiments. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the DM only control. ** = $P < 0.01$, *** = $P < 0.001$, and **** = $P < 0.0001$.

Both heparin and GW4869 treatment had an inhibitory effect on myogenic differentiation (**Figure 4.1A-D**). Heparin resulted in the reduction of the MHC+ area, as well as the myogenic and fusion indices, by ~40% ($P < 0.01$), ~40% ($P < 0.01$) and ~60% ($P < 0.0001$) respectively, relative to the untreated control (**Figure 4.1B-D**). The effect of GW4869 treatment was similarly inhibitory, resulting in a ~30% ($P < 0.01$), and ~60% ($P < 0.0001$) reduction in the MHC+ area and fusion index, respectively (**Figure 4.1B and D**). The myogenic index was reduced by ~25%, however, this did not reach statistical significance relative to the untreated control (**Figure 4.1C**). Conversely, C2C12 myocyte proliferation was unaffected by either treatment as there was no difference in the total number of nuclei between these groups and the untreated control (**Figure 4.1E**). Furthermore, DMSO did not affect myogenic differentiation or cell proliferation (**Figure 4.1A-E**). These results demonstrate that small molecule-mediated inhibition of exosome release and EV uptake within the same culture dish impairs myogenic differentiation, and suppresses myotube maturation. These data support the notion that C2C12 myotube-derived EVs play an important pro-myogenic role through paracrine signalling during myogenic differentiation.

4.4.2 Combined siRNA knockdown of *Rab27a* and *Rab27b* exosome biogenesis factors inhibits myogenic differentiation

To assess the involvement of the specific exosomal subpopulation of EVs in the functional transfer of pro-myogenic factors between differentiating C2C12 myoblasts, the transcripts of two key factors associated with exosome biogenesis, RAB27A and RAB27B [273,280,286,524–527], were knocked down by RNA interference using small interfering RNAs (siRNAs). C2C12 cells were treated with siRNAs specific to *Rab27a* and *Rab27b*, separately or in combination, one day prior to switching to DM (**Figure 4.2A**). The extent of C2C12 myogenic differentiation was assessed two days later by MHC IF (**Figure 4.2A**). Transcript levels of *Rab27a* and *Rab27b* were significantly ($P < 0.01$) reduced by ~70% and ~90% relative to the non-targeting

control respectively (**Figure 4.2B**). There was no significant effect on the overall MHC+ area between the treatments, relative to the DM only control (**Figure 4.2C**). Individual siRNA-mediated *Rab27a* and *Rab27b* knockdown did not result in differences in the myogenic index of the myotubes, however, a significant ~20% reduction ($P < 0.01$) was observed for the combined siRNA treatments targeting both transcripts simultaneously (**Figure 4.2D**). Similarly, the combination treatment resulted in a ~40% reduction ($P < 0.01$) in the fusion index, relative to the control, while the individual treatments alone did not significantly alter the fusion of myotubes (**Figure 4.2E**). Notably, there was no change in the number of nuclei relative to the control for any of the siRNA treatment groups and, as such, the effect of si*Rab27a* and si*Rab27b* co-treatment on inhibiting myogenic differentiation is most likely a result of direct regulation of the myogenic differentiation pathway (**Figure 4.2F**). These results suggest that RAB27A and RAB27B are both required for the production of exosomes that possess transferrable pro-myogenic properties, but that the functions of each individually may be redundant. This is supported by reports elsewhere that the activity of RAB GTPases in EV secretion is cell-type-dependent [528]. The level of inhibition of myotube fusion with knockdown of both *Rab27a* and *Rab27b* was similar to that achieved with GW4869 or heparin treatment (**Figure 4.1**). Heparin, however, was found to exert a greater inhibitory effect on the myogenic index than both GW4869 treatment and combined *Rab27a* and *Rab27b* transcript knockdown which is unsurprising given that the effects of heparin on EV signalling are more pleiotropic.

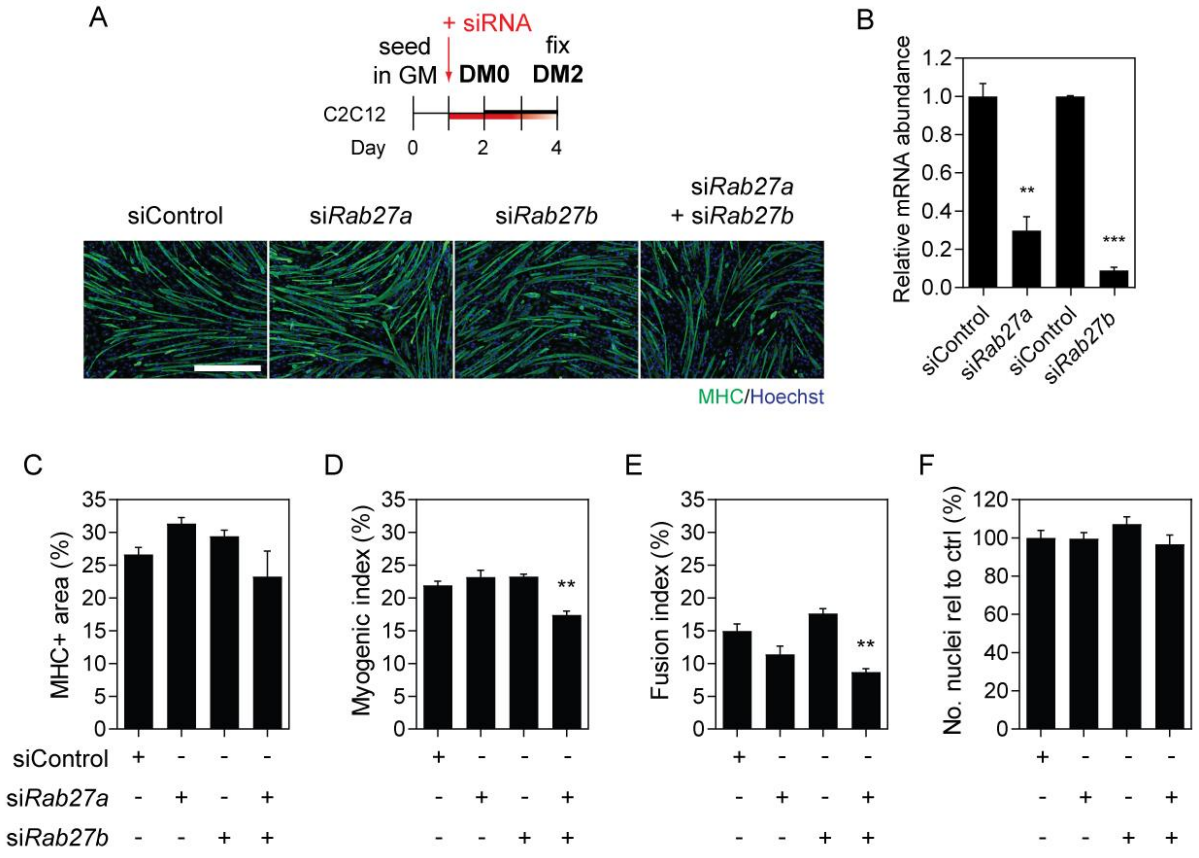


Figure 4.2 Combined siRNA-mediated knock down of *Rab27a* and *Rab27b* exosome biogenesis factors reduces myogenic differentiation.

C2C12 cells were cultured in GM for two days and then switched to DM for two days. Cultures were treated with siRNAs targeted to two exosome biogenesis factors, *Rab27a* and *Rab27b*, separately or in combination, at a final concentration of 100 nM one day before switching to DM. Treatment with a non-targeting siRNA was included as a negative control. (**A**) Myogenic differentiation was assessed by MHC IF. (**B**) *Rab27a* and *Rab27b* mRNA levels were determined by RT-qPCR, normalised to the *Rplp0* reference gene. The data were scaled such that the mean of the control group was returned to a value of 1. Myogenic differentiation was quantified by (**C**) measuring the MHC+ area, and by calculating the (**D**) myogenic and (**E**) fusion indices. (**F**) The total number of nuclei per representative field of view are shown as a percentage relative to the control group. All microscopy images were taken at 10× magnification. Scale bar represents 400 µm. All values are means shown + SEM from four fields of view obtained in one independent experiment, representative of two independent experiments. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the non-targeting siRNA control. ** = $P < 0.01$.

4.4.3 EV-containing myotube-conditioned media enhances myogenic differentiation

To determine whether pro-myogenic paracrine factors secreted by C2C12 myotubes could enhance myogenic differentiation when transferred between cell culture dishes, C2C12 myotube-conditioned-media (CM) was harvested from cultures in the late stages of differentiation (i.e. at DM7). The supernatant was subjected to sequential ultrafiltration (UF) using 300, 100 and 10 kDa molecular weight (Mr) cut-off filters. The four different treatment fractions obtained from this process included: (i) crude unfractionated CM, (ii) CM containing EVs and other particles >300 kDa in size, and CM depleted of EVs with particles that are (iii) 100-300 kDa, or (iv) 10-100 kDa, in size (**Figure 4.3A**). The resulting fractions were mixed with an equal volume (50:50) of fresh complete DM so as not to deprive the recipient cultures of vital nutrients depleted by the metabolic activity of the producer cells, as well as through sequential filtration of the CM (**Figure 4.3A and B**), and transferred to C2C12 myoblasts at DM0. The effect of each treatment fraction on cell proliferation and myogenic differentiation was assessed 24 and 48 hrs later respectively in parallel experiments (**Figure 4.3A and B)**. A major contributing factor towards the initiation of myogenic differentiation is cell-cell contact [529]. The effect of transferred media on cell proliferation was therefore monitored at an earlier timepoint than myogenic differentiation to exclude the possibility that the threshold for cell-cell contact was not caused by a particular treatment fraction, thereby advancing the myogenic differentiation trajectory independently of transferred paracrine signalling. An assay that measures EdU incorporation into actively dividing cells was used to measure cell proliferation (**Figure 4.3B**).

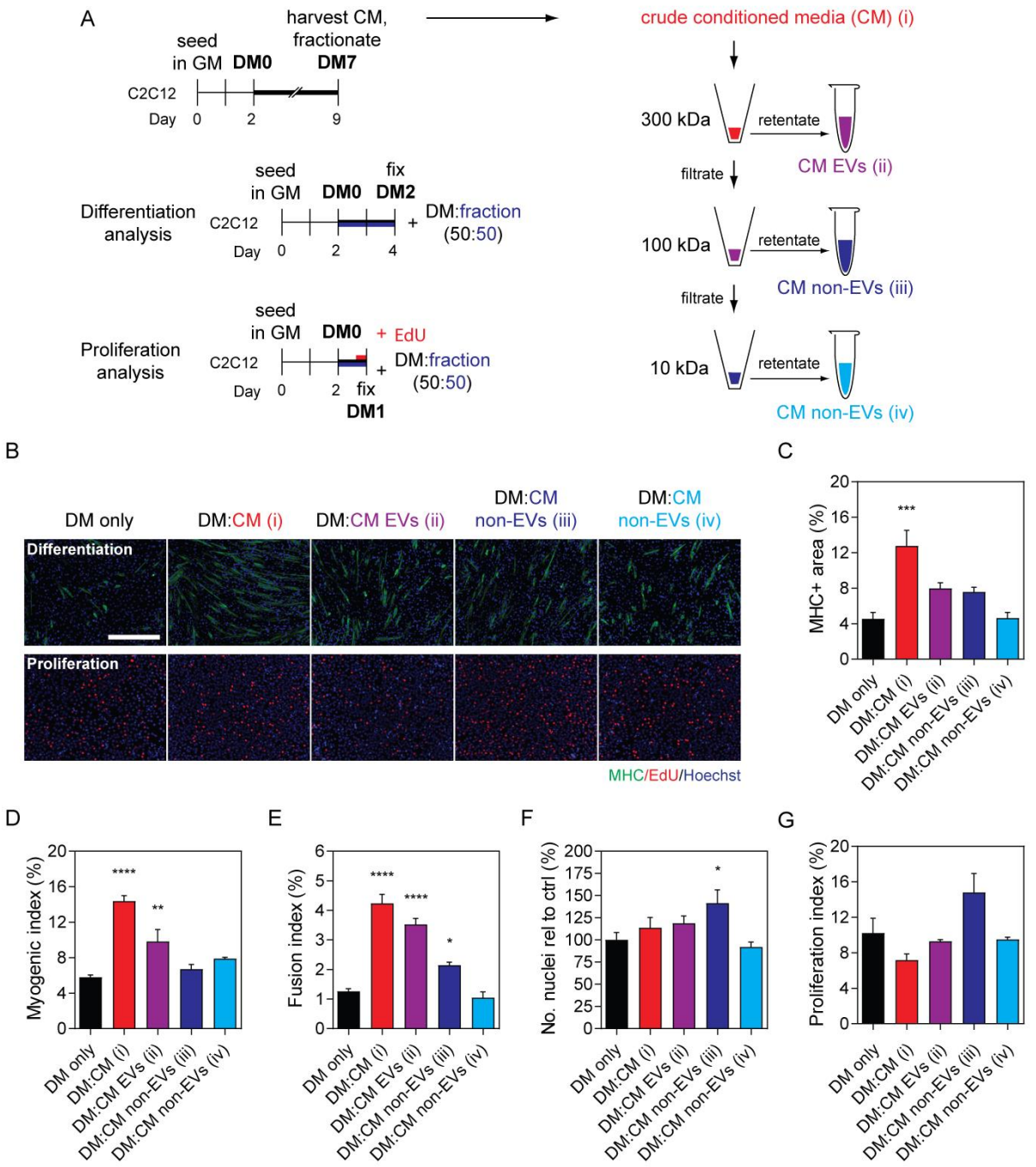

A
seed in GM
DM0
harvest CM, fractionate
DM7
C2C12
Day 0 2 9
Differentiation analysis
seed in GM
DM0
fix
DM2
C2C12
Day 0 2 4
DM:fraction (50:50)
Proliferation analysis
seed in GM
DM0
+ EdU
C2C12
Day 0 2
fix
DM1
DM:fraction (50:50)
crude conditioned media (CM) (i)
300 kDa
retentate
CM EVs (ii)
filtrate
100 kDa
retentate
CM non-EVs (iii)
filtrate
10 kDa
retentate
CM non-EVs (iv)
B
DM only
DM:CM (i)
DM:CM EVs (ii)
DM:CM non-EVs (iii)
DM:CM non-EVs (iv)
Differentiation
Proliferation
MHC/EdU/Hoechst
C
MHC+ area (%)
D
Myogenic index (%)
E
Fusion index (%)
F
No. nuclei rel to ctrl (%)
G
Proliferation index (%)

Figure 4.3 EV-containing myotube-conditioned media enhances myoblast differentiation.

(**A**) C2C12 cells were cultured in GM for two days and then switched to DM for one week. Myotube-conditioned-media (CM) from DM7 cultures was fractionated using sequential molecular weight (Mr) cut-off filters of 300, 100 and 10 kDa. Treatment fractions obtained included: (i) crude CM, (ii) EV-containing CM, and (iii and iv) EV-depleted CM. C2C12 cells were cultured in two separate wells per treatment group for two days in GM, and treated with each fraction at the same time of switching to DM to assess myogenic differentiation by MHC IF, and cell proliferation by measuring the incorporation of EdU into newly synthesised DNA. Samples were treated with a 50:50 mixture of fractionated CM and complete DM. Treatment with an equal volume of complete DM was included as the DM only control. (**B**) Myogenic differentiation was assessed by MHC IF, and was quantified by (**C**) measuring the MHC+ area, and by calculating the (**D**) myogenic index, and (**E**) fusion index. (**F**) The total number of nuclei per representative field of view are shown as a percentage relative to the control group. (**G**) The proliferation index was calculated. All microscopy images were taken at 10× magnification. Scale bar represents 400 μm. All values are means shown + SEM from four fields of view obtained in one independent experiment, representative of two independent experiments. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the DM only control. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$, and **** = $P < 0.0001$.

Treatment with the crude myotube-CM, containing EVs as well as secreted proteins and soluble factors, enhanced myogenic differentiation substantially (**panel 1 of Figure 4.3B**) as demonstrated by an increase in the MHC+ area (**Figure 4.3B and C**) and myogenic index (**Figure 4.3D**) by ~300% ($P < 0.001$), as well in the fusion index (**Figure 4.3E**) by ~400% ($P < 0.0001$), relative to the untreated control. This pro-myogenic effect could not be explained by enhanced cell proliferation, as no significant difference in the total number of nuclei at DM2 (**Figure 4.3F**), or the proliferation index at DM1 (**panel 2 of Figure 4.3B, Figure 4.3G**), was observed. These findings suggest that differentiating myotubes are capable of releasing pro-myogenic factors into the culture medium, which can be transferred to recipient cultures.

Treatment with the EV-containing fraction (CM EVs) also enhanced myogenic differentiation (**panel 1 of Figure 4.3B**), resulting in an increase in the myogenic and fusion indices by ~200% ($P < 0.01$) (**Figure 4.3D**) and ~300% ($P < 0.0001$) (**Figure 4.3E**) respectively. A ~200% increase in MHC+ area was observed which is consistent with the other two measures of myogenic differentiation for this treatment fraction, although this did not reach statistical significance (**Figure 4.3C**). The pro-myogenic effect of the EV-containing CM fraction could also not be explained by an increase in cell proliferation (**Figure 4.3F and G**).

In general, both of the EV-depleted fractions had minimal effect on myogenic differentiation. A marginal increase ($P < 0.05$) in the fusion index was observed for the treatment group enriched in particles 100-300 kDa in size (**Figure 4.3E**). This may be accounted for by indirect effects of increased cell density as a result of a greater number of nuclei observed only in this treatment group ($P < 0.05$) (**Figure 4.3F**). Furthermore, the proliferation index for this group was similarly increased, although this did not reach statistical significance at the $P < 0.05$ level (**Figure 4.3G**).

The findings from this experiment demonstrate that the pro-myogenic effect of unfractionated, as well as EV-containing, myotube-CM is diminished through the sequential depletion of larger biomolecules by spin filtration. This is highly suggestive of a role of EVs in the functional transfer of pro-myogenic factors between donor and recipient cells. It is not possible, however, to conclude that the effect was purely EV-borne as other larger protein complexes would also have been present within the EV-containing treatment fraction. Our lab has previously developed a scalable method for the isolation of EV populations that are devoid of protein contamination by combining UF with size exclusion liquid chromatography (UF-SEC) [515]. This method provides a means to delve more deeply into the nuances of the effects of the myotube-derived secretome on myogenic differentiation in downstream experiments.

4.4.4 Myotube-derived EVs exert a pro-myogenic effect at low doses

CM was harvested from C2C12 myoblasts and myotubes to obtain EVs representative of cell cultures in the growth phase (MB-EVs), or undergoing myogenic differentiation (MT-EVs), respectively. The effects of these two pure EV subtypes on myogenic differentiation was assessed and compared to further elucidate the role of myotube-derived EVs in the functional transfer of pro-myogenic factors. UF-SEC was employed to obtain MB-EVs from the CM of C2C12 myoblasts cultured in GM for 4 days, and then low serum media for 1 day, and MT-EVs from the CM of myotubes cultured in GM for 4 days, and DMEM for 4 days (**Figure 4.4A**).

During EV purification by size exclusion chromatography (SEC), a concentrated sample is injected into a column packed with a porous resin. Smaller molecules are able to enter these pores and elute more slowly from the column, while larger particles elute more quickly. The volume of each fraction collected can be controlled for broader, or narrower, molecular size separation. The absorbance of eluting molecules is measured at 280 nm which

generates a chromatogram trace read-out, depicting the relative protein content across the range of eluted fractions. EVs (a heterogenous pool consisting of exosomes as well as small microvesicles) are among the largest particles within the UF-prepared samples, and are thus represented by the first peak within the chromatogram trace (**Figure 4.4B**). Soluble factors and extracellular proteins correspond to the larger and broader second peak recorded on the SEC trace. The EV-containing fractions were pooled, and nanoparticle tracking analysis (NTA) performed to assess the distribution in sizes of the purified particles, as well as to determine the concentration of EVs in each sample (**Figure 4.4C**). The modal sizes of MB- and MT-EVs were 108.9 and 83.9 nm respectively, and these values were not significantly different to one another (Student's *t*-test, P=0.0793). The size of the isolated EVs was within the expected range for exosomes and other small microvesicles (i.e. ~30-150 nm). Western blot analysis confirmed expression of exosome-specific markers, Alix (PDCD6IP, 97 kDa) and CD9 (26 kDa), within the EV lysates, but not in the corresponding producer cell lysates (**Figure 4.4D**). Furthermore, the cell-specific marker, Calnexin (CANX, 90 kDa), was present in the producer cells but absent in the EVs (**Figure 4.4D**). The results obtained from NTA and Western blotting indicate the successful isolation of pure EVs from the CM of myoblasts and myotubes.

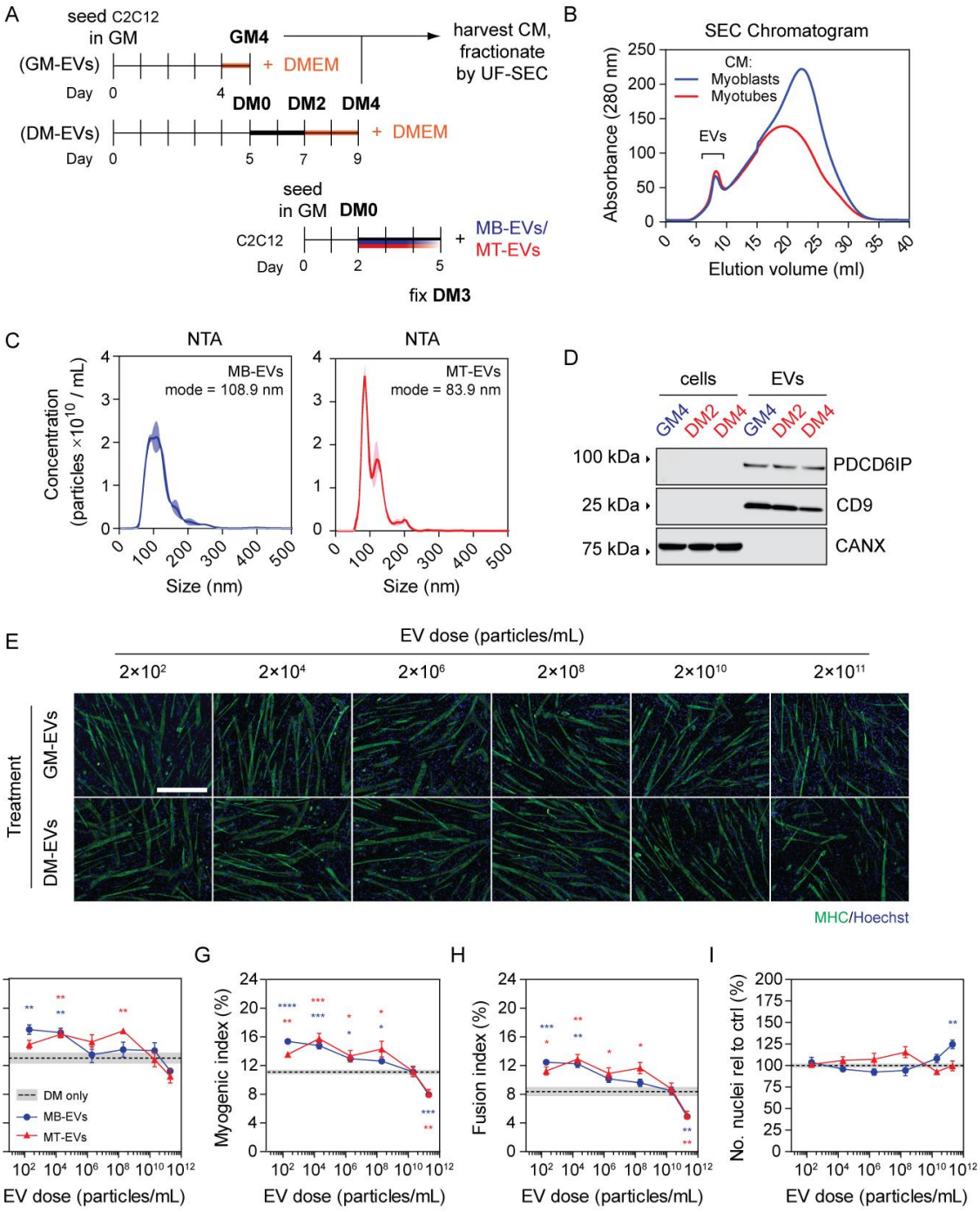

A
seed C2C12 in GM
GM4
harvest CM, fractionate by UF-SEC
(GM-EVs)
+ DMEM
Day 0
4
DM0 DM2 DM4
(DM-EVs)
+ DMEM
Day 0
5
7
9
seed in GM DM0
C2C12
+ MB-EVs/ MT-EVs
Day 0
2
5
fix DM3
B
SEC Chromatogram
CM:
Myoblasts
Myotubes
EVs
Absorbance (280 nm)
Elution volume (ml)
C
NTA
MB-EVs
mode = 108.9 nm
NTA
MT-EVs
mode = 83.9 nm
Concentration (particles ×10^10 / mL)
Size (nm)
Size (nm)
D
cells
EVs
GM4
DM2
DM4
GM4
DM2
DM4
100 kDa
PDCD6IP
25 kDa
CD9
75 kDa
CANX
E
EV dose (particles/mL)
2×10^2
2×10^4
2×10^6
2×10^8
2×10^10
2×10^11
Treatment
GM-EVs
DM-EVs
MHC/Hoechst
F
MHC+ area (%)
DM only
MB-EVs
MT-EVs
EV dose (particles/mL)
G
Myogenic index (%)
EV dose (particles/mL)
H
Fusion index (%)
EV dose (particles/mL)
I
No. nuclei rel to ctrl (%)
EV dose (particles/mL)

Figure 4.4 Myoblast- and myotube-derived EVs enhance myogenesis at lower doses. (**A**) EVs were isolated by ultrafiltration size exclusion liquid chromatography (UF-SEC) from CM obtained from C2C12 myoblasts (MB-EVs) or myotubes (MT-EVs). Myoblasts were grown in GM for three days and serum-free media (DMEM) for one day. Myotubes were grown in GM for five days, then switched to DM for two days and serum-free media for a further two days. (**B**) The EV-containing eluates from the SEC column were determined by UV spectrophotometry absorbance at 280 nm, and pooled before treating C2C12 cells grown in GM for 2 days at the time of switching to DM. (**C**) The modal size and concentration of EVs was determined by nanoparticle tracking analysis (NTA). Values are the mean distribution from three separate experiments, shown ± SEM (the shaded area). (**D**) Exosome-specific (PDCD6IP/Alix and CD9) and cell-specific (CANX/Calnexin) markers were detected by Western blot. (**E**) Myogenic differentiation was assessed by MHC IF at DM3 and quantified by (**F**) measuring the MHC+ area, and using (**G**) myogenic, and (**H**) fusion indices. (**I**) The total number of nuclei per representative field of view are shown as a percentage relative to the control group. All microscopy images were taken at 10× magnification. Scale bar represents 400 μm. All values are means shown ± SEM from four fields of view obtained in one independent experiment, representative of two independent experiments. Cultures treated with complete DM were used as the DM only controls. Statistical significance was determined by a Student's *t*-test relative to the DM only control. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$, and **** = $P < 0.0001$. The colour of the significance indicators corresponds to the respective comparisons (i.e. blue for MB-EVs, red for MT-EVs). MB: myoblast. MT: myotube.

Forterre and colleagues conducted proteomic analysis of EVs produced by C2C12 myoblasts and myotubes, and reported that differentiating C2C12 cells release 0.42 ± 0.01 µg/mL of EVs per 24 hr period [306]. This value was therefore considered to be an approximate 'physiological dose' for C2C12 cells in the present study. The majority of studies that have assessed the functional transfer of EVs in culture used protein concentration as a measure of dose, and effects have been reported with 10-100 ug/mL of exogenously derived EVs [251,308,522]. These studies, however, employed polymer precipitation isolation methods which typically co-purify high amounts of non-vesicular protein. Dose comparison may therefore be misleading when based on the existing literature. As such, C2C12 cells in the study reported here were treated with MB- and MT-EVs in a dose response, ranging from 2×10^2 to 2×10^{11} particles/mL. When converted to protein concentrations, these doses included three that were below, and three that were above, the ~0.5 µg/mL 'physiological dose' [306].

The effect on myogenic differentiation of C2C12 myoblasts treated at the time of switching to DM with EVs was assessed by MHC IF (**Figure 4.4E**). The transfer of lower doses of both MB- and MT-EVs (2×10^2 - 2×10^8 particles/mL) to myoblast cultures enhanced myogenic differentiation as measured by the MHC+ area (**Figure 4.4F**), myogenic index (**Figure 4.4G**), and fusion index (**Figure 4.4H**), relative to the untreated control. The greatest effect of the MB-EV treatment group was observed at the lowest dose (2×10^2 particles/mL), achieving a ~30% increase ($P < 0.01$) in MHC+ area (**Figure 4.4F**), ~35% increase ($P < 0.0001$) in myogenic index (**Figure 4.4G**), and ~50% increase ($P < 0.001$) in the fusion index (**Figure 4.4H**). MT-EVs resulted in the most significant promotion of myogenic differentiation at the second lowest dose tested (2×10^4 particles/mL), with increases of ~25% ($P < 0.01$), ~40% ($P < 0.001$), and ~50% ($P < 0.01$) in the MHC+ area (**Figure 4.4F**), myogenic index (**Figure 4.4G**), and fusion index (**Figure 4.4H**) respectively. Conversely, both MB- and MT-EVs significantly inhibited myogenic differentiation by ~30% ($P < 0.01$ for MT-EVs and 0.001 for

MB-EVs), and ~60% ($P < 0.01$) at the highest dose (2×10^{11} particles/mL), as measured by the myogenic and fusion indices respectively.

The total number of nuclei following treatment with MT-EVs was not significantly different at any dose, relative to the control (**Figure 4.4I**), while the same was true for the MB-EV treatment group at all but the highest dose, where a ~25% ($P < 0.01$) increase in nuclei was observed (**Figure 4.4I**). The effects of MB- and MT-EV transfer on myogenic differentiation therefore cannot be explained by alterations in cell density. Instead, there appear to be both pro- and anti-myogenic factors within the vesicular compartment of C2C12 cells which are independent of the growth phase of the donor cell. Our lab has reported that myomiRs are expressed almost exclusively during myogenic differentiation [234] and, as such, it is unlikely that the pro-myogenic effects observed with low dose EV treatment were as a result of the functional transfer of EV-contained extracellular myomiRs. Furthermore, the increase in myogenic differentiation observed even with the most successful MT-EV treatment was only ~50%, whereas the maximal effect with treatment of EV-containing CM fractions (>300 kDa) was ~200% in the previous experiment (**Figure 4.3**). Pro-myogenic effects in these other experiments may be the result of the combination of EV- and other non-EV-mediated effects.

4.4.5 The choice of isolation medium determines the effect of myotube-derived secreted protein transfer on myogenic differentiation

To determine whether high Mr non-vesicular fractions could also possess transferrable pro-myogenic characteristics, UF-SEC purification was carried out to obtain pure secreted protein fractions from myotube-CM. The non-vesicular secreted 'protein' fraction from late stage (DM7) differentiating C2C12 myotubes was purified by UF-SEC, and transferred to myoblasts to determine the effect on myogenic differentiation (**Figure 4.5A**). Donor cultures were grown in GM for two days, DM for two days, and Opti-MEM for 2 days as this is a low serum media (i.e. contains minimal protein that can be carried over during the LC fractionation step) that

facilitates myoblast to myotube differentiation (**Figure 4.5A**). Fifteen protein fractions were collected, ranging from high to low Mr based on the temporal order of elution from the column (**Figure 4.5B**). C2C12 myoblasts were treated with each fraction (1 μg/mL) at the time of switching to DM. The effect on myogenic differentiation was assessed by MHC IF 24, 48, and 72 hrs post treatment, of which representative microscopy images are shown for one high, and one low, Mr protein fraction (#3 and #12 respectively, grey bars indicated on the SEC chromatogram) (**Figure 4.5B**). The MHC+ area was progressively increased over time, and significantly increased (ranging from $P < 0.01$ to $P < 0.0001$) at 48 and 72 hrs following treatment with the lower Mr 'protein' fractions in the range of #7 to #15 (**Figure 4.5C and D**). This effect was greatest at 72 hrs post treatment with fraction #12, resulting in a >200% increase in MHC expression relative to the control. The total number of nuclei was also significantly increased at each timepoint for treatment fraction #7-15, peaking with #12 at >250% relative to the control (**Figure 4.5E**). Furthermore, cell proliferation was increased even at 24 hrs following treatment (ranging from $P < 0.05$ to $P < 0.0001$) with fraction #7-15, peaking at a >150% increase in the number of nuclei, whereas the effects on MHC expression manifested only 48 hrs post treatment. It is therefore difficult to distinguish between direct and indirect effects on myogenic differentiation due to cell density. These data show that a low Mr non-vesicular component of these fractions is inducing a profound increase in myogenic differentiation.

No pro-myogenic effects were observed when myoblast culture were treated with 5 μg/mL of bovine serum albumin (BSA) at the time of switching to DM, thereby showing that this effect is not a consequence of non-specific increase in extracellular protein *per se* (**Figure 4.6A**).

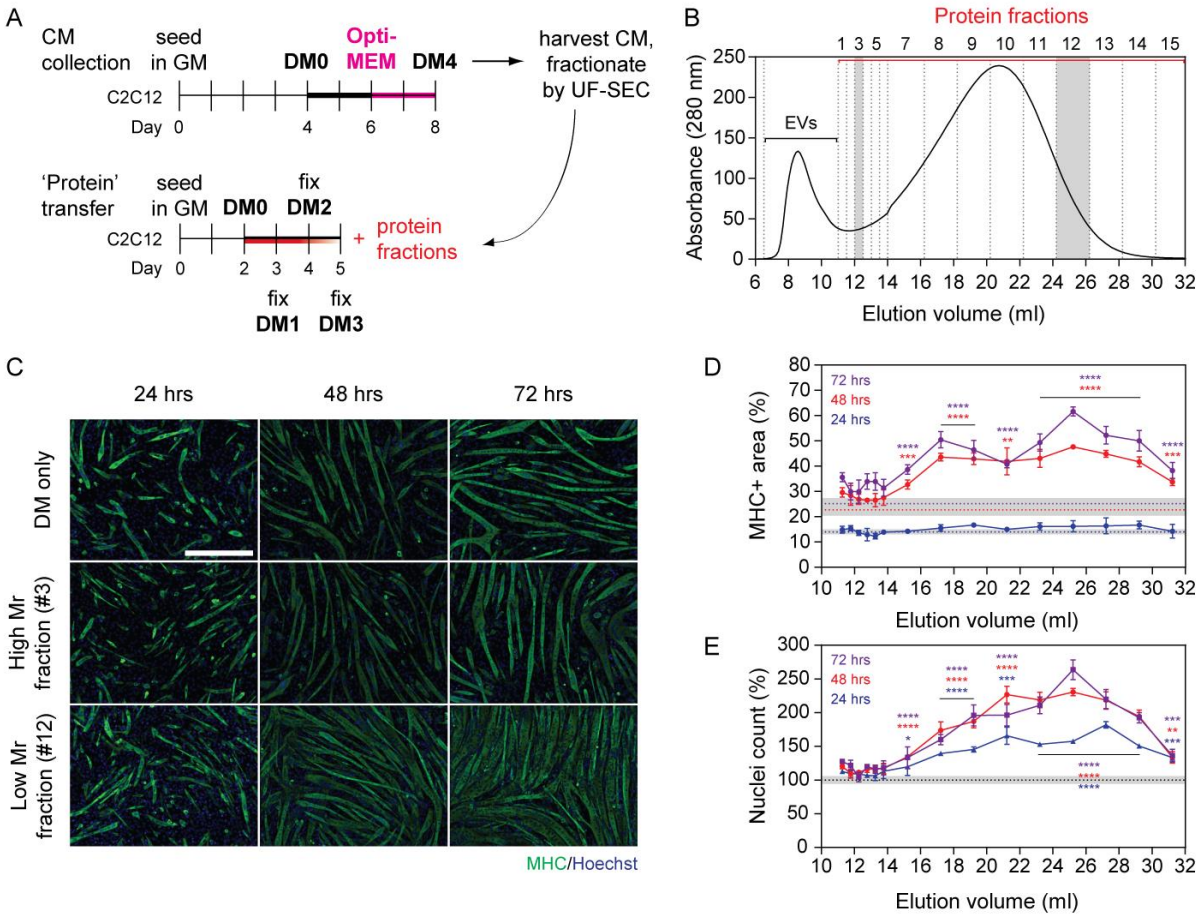

A
CM collection
seed in GM
Opti-MEM
DM0
DM4
C2C12
Day 0
4
6
8
harvest CM, fractionate by UF-SEC
'Protein' transfer
seed in GM
DM0
fix DM2
protein fractions
2
3
4
5
fix DM1
fix DM3
B
Protein fractions
1 3 5 7 8 9 10 11 12 13 14 15
Absorbance (280 nm)
EVs
Elution volume (ml)
C
24 hrs
48 hrs
72 hrs
DM only
High Mr fraction (#3)
Low Mr fraction (#12)
MHC/Hoechst
D
72 hrs
48 hrs
24 hrs
MHC+ area (%)
Elution volume (ml)
E
Nuclei count (%)
Elution volume (ml)

Figure 4.5 Low molecular weight myotube-derived secreted proteins enhance myogenic differentiation when purified from Opti-MEM CM.

(**A**) C2C12 myotube-derived secreted protein fractions were isolated by ultrafiltration size exclusion liquid chromatography (UF-SEC) from CM collected in Opti-MEM. Myoblasts were grown in GM for four days, followed by DM for two days, and then serum-free media (Opti-MEM) for a further two days. (**B**) The non-vesicular-associated protein eluates from the SEC column were measured by UV spectrophotometry absorbance at 280 nm. C2C12 cells were grown in GM for 2 days and treated with 1 μg/mL of each of the 15 protein fractions at the time of switching to DM. (**C**) Myogenic differentiation was assessed by MHC IF at DM1 (24 hrs), DM2 (48 hrs), and DM3 (72 hrs). Representative images from one high Mr, and one low Mr protein fraction (#3 and #12 respectively, highlighted in grey on the LC trace) are shown from each timepoint. (**D**) Myogenic differentiation was quantified by measuring the MHC+ area. The means of corresponding DM only control groups from each timepoint are shown ± SEM (grey area). (**E**) The total number of nuclei per representative field of view are shown as a percentage relative to the control group (data were scaled such that the mean of the control group was returned to a value of 100%). Cultures treated with complete DM were used as the DM only controls. All microscopy images were taken at 10× magnification. Scale bar represents 400 μm. All values are means shown ± SEM from four fields of view obtained in one independent experiment. Statistical significance was determined by a Student's *t*-test (only shown for fraction #7-15) relative to the DM only control at each corresponding timepoint. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$, and **** = $P < 0.0001$. The colour of the significance indicators corresponds to the respective comparisons (i.e. blue for the 24 hr timepoint, red for 48 hrs, and purple for 72 hrs). MB: myoblast. MT: myotube.

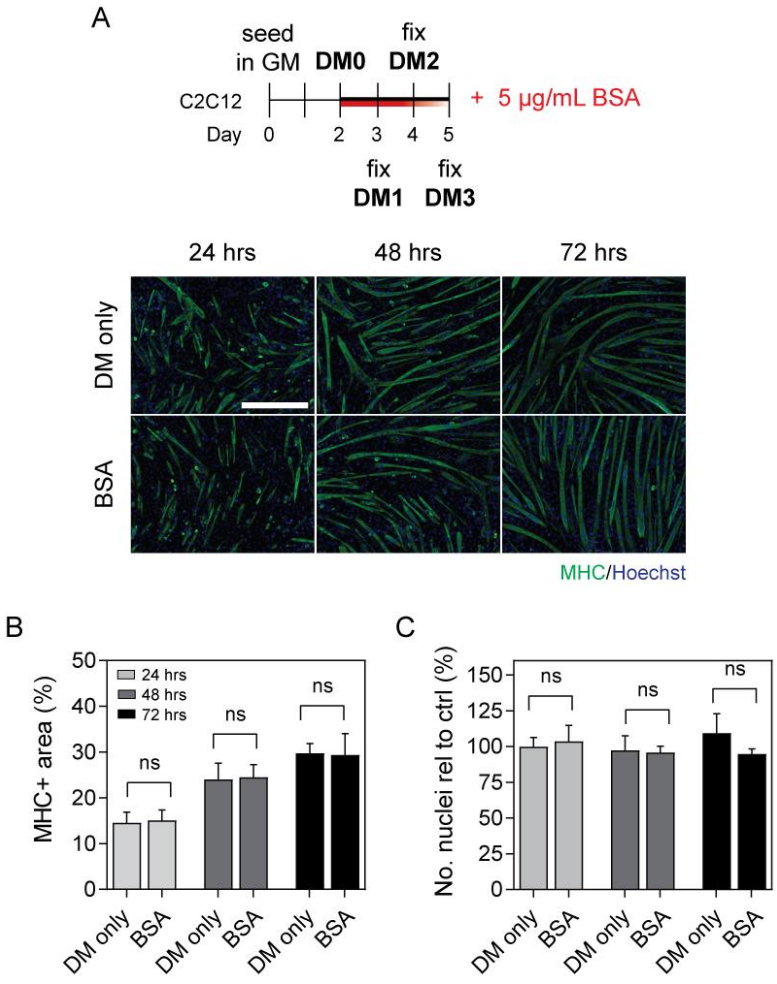


Figure 4.6 A non-specific increase in extracellular protein does not enhance myogenic differentiation.

(**A**) C2C12 myoblasts were grown in GM for two days, and treated with 5 µg/mL of bovine serum albumin (BSA) at the time of switching to DM. Myogenic differentiation was assessed by MHC IF at DM1, DM2 and DM3, and (**B**) quantified by measuring the MHC+ area. Untreated (DM only) cultures were included as negative controls. (**C**) The total number of nuclei per representative field of view are shown as a percentage relative to the control group. All microscopy images were taken at 10× magnification. Scale bar represents 400 µm. All values are means shown + SEM from four fields of view obtained in one independent experiment. Statistical significance was determined by a Student's *t* test. ns = not significant.

Opti-MEM was used as the isolation media which is typical for EV transfer experiments, however, this media contains additives which may be confounding factors in this experiment. To determine whether the pro-myogenic effects of non-vesicular protein transfer to differentiating myoblast cells was directly caused by a myotube-specific secreted factor, or a co-purified component from the isolation medium, myotube-derived secreted 'protein' fractions were obtained in the same manner as previously described (**Figure 4.4A**), but under different culture media conditions (i.e. (i) DM/Opti-MEM, (ii) Opti-MEM/DMEM, and (iii) DM/DMEM) (**Figure 4.7A**). The SEC chromatogram traces from each condition were similar (**Figure 4.7B**). Furthermore, the protein components of Opti-MEM were also separated by SEC which showed a 280 nm peak in the low Mr range (fractions #8-13) (**Figure 4.7B**). C2C12 myoblasts were treated with one high Mr and one low Mr fraction (#3 and #12 respectively) from each of the three growth conditions. Only the low Mr protein fraction from the DM/Opti-MEM treatment condition (i) resulted in highly pro-myogenic effects (**Figure 4.7C**), consistent with our previous results, described above (**Figure 4.5**). Furthermore, cultures treated with a 50:50 mix of DM and Opti-MEM exhibited similar enhanced effects on myogenic differentiation (**Figure 4.7C**). Pre-treatment with Opti-MEM did not produce the same effects, suggesting it is the media itself and not Opti-MEM-stimulated release of a pro-myogenic factor. Similarly, when DM or DMEM was used as collection media, no enhancement in myogenic differentiation was observed. These results demonstrate that the myotube-derived non-vesicular secretome (which includes free myomiR:protein complexes) does not induce myogenic differentiation. Instead, the use of Opti-MEM is a major potential confounding factor in such experiments. This is supported by the overlap of the SEC chromatogram trace obtained from Opti-MEM and the fractions that transferred the greatest pro-myogenic effects (**Figure 4.7B**).

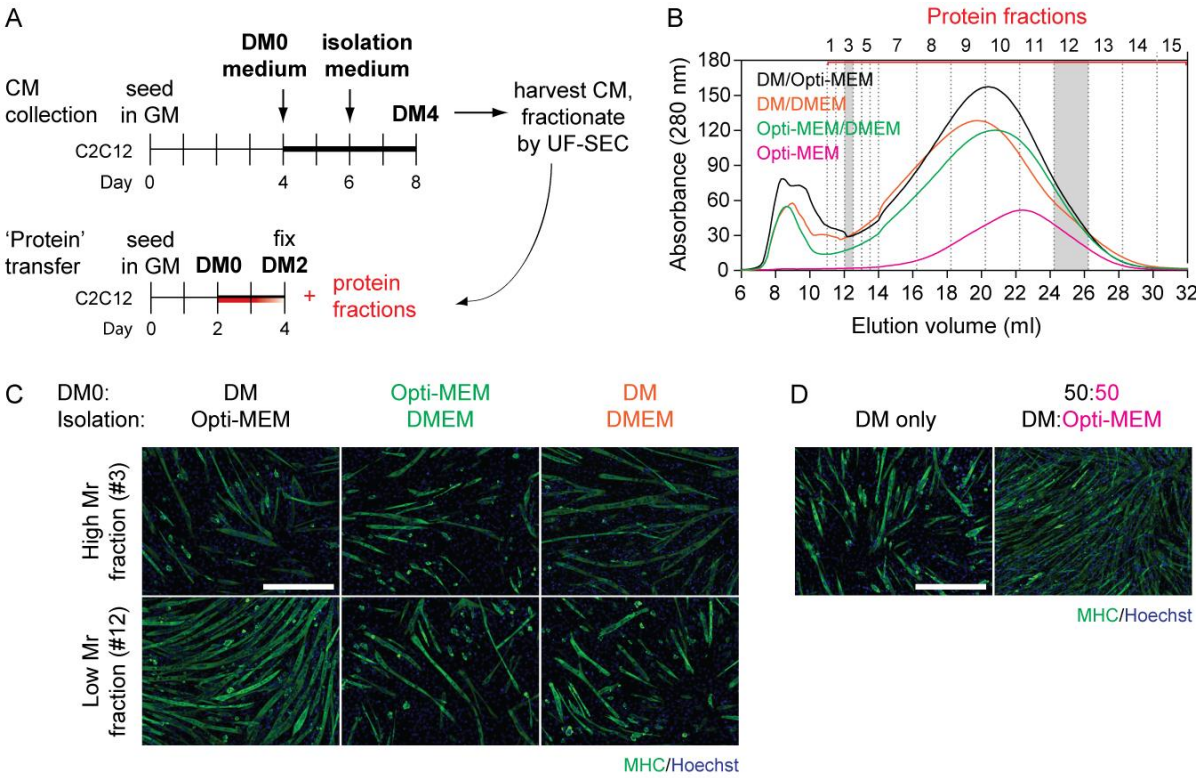


Figure 4.7 Pro-myogenic factors in Opti-MEM confound results from myotube-derived secreted protein transfer experiments.

(**A**) C2C12 myotube-derived secreted protein fractions were isolated by UF-SEC from CM obtained from cell growth in different combinations of DM0 and isolation media. Myoblasts were grown in GM for four days, DM0 medium (DM or Opti-MEM) for two days, and isolation medium (Opti-MEM, DMEM or DM) for a further two days. DM0 and isolation media combinations were: (i) DM followed by Opti-MEM, (ii) Opti-MEM followed by DM, and (iii) DM followed by DMEM. (**B**) The protein eluates from the SEC column were measured by UV spectrophotometry absorbance at 280 nm. An equal volume of pure Opti-MEM was processed by SEC for visualisation of the absorbance of proteins contained within this culture medium. C2C12 cells were grown in GM for 2 days and then treated with (**C**) 1 μg/mL of one high Mr protein fraction (#3) or one low Mr protein fraction (#12) from each growth condition group at the time of switching to DM, or (**D**) directly with a 50:50 mix of DM and Opti-MEM. Cells treated with DM only were included as a control. Myogenic differentiation was assessed by MHC IF at DM2. All microscopy images were taken at 10× magnification. Scale bar represents 400 μm.

4.4.6 Insulin within Opti-MEM confounds myogenic differentiation experiments

Closer inspection of the components within Opti-MEM revealed potential candidates which could be responsible for enhancing myogenic differentiation in the purified non-vesicular fraction transfer experiments. The precise composition and concentrations of this culture medium is proprietary, however, according to the product datasheet it contains 15 µg/mL of protein in total, which includes insulin and transferrin [530]. These factors are both incidentally also components of ITS (Insulin Transferrin Selenite) medium [531], a supplement used in a number of studies to promote myogenic differentiation of cells *in vitro* [532–539]. ITS medium at 1× concentration in culture media contains 10 µg/mL of insulin, and 5.5 µg/mL of transferrin. Sodium selenite, the third component of ITS supplement, exists at a much lower concentration (0.0067 µg/mL at 1× concentration), and was not further studied. C2C12 cells were grown in GM for two days, and then treated with either insulin or transferrin (at concentrations of 1 and 10 µg/mL) at the time of switching to DM (**Figure 4.8A**). These doses were selected as UF-SEC treatment fractions were transferred at a final concentration of 1 µg/mL, which is thus presumed to be the maximum dose of insulin or transferrin that would have been present in these samples (**Figure 4.5),** and 10 µg/mL is comparable with the highest dose of insulin or transferrin that is expected to be in Opti-MEM. Differentiating myoblast cultures treated with insulin displayed massively elevated levels of MHC expression at DM2, at >200% relative to the untreated control (**Figure 4.8B**). These pro-myogenic effects were not significantly different between the doses, suggesting that the concentrations far exceeded the threshold of insulin treatment required to promote myogenic differentiation. Conversely, transferrin had no effect at either concentration. The number of nuclei were greatly elevated to ~250% ($P < 0.0001$) relative to the DM only control following treatment with insulin at both concentrations, while transferrin had no effect on cell proliferation (**Figure 4.8C**). These findings demonstrate that one or more factors within Opti-MEM, including insulin, were co-purified with the non-

vesicular myotube-CM fraction by UF-SEC thereby confounding the highly pro-myogenic effects of secreted non-vesicular fractions on differentiating myoblasts which were observed in the experiment described above (**Figure 4.5**).

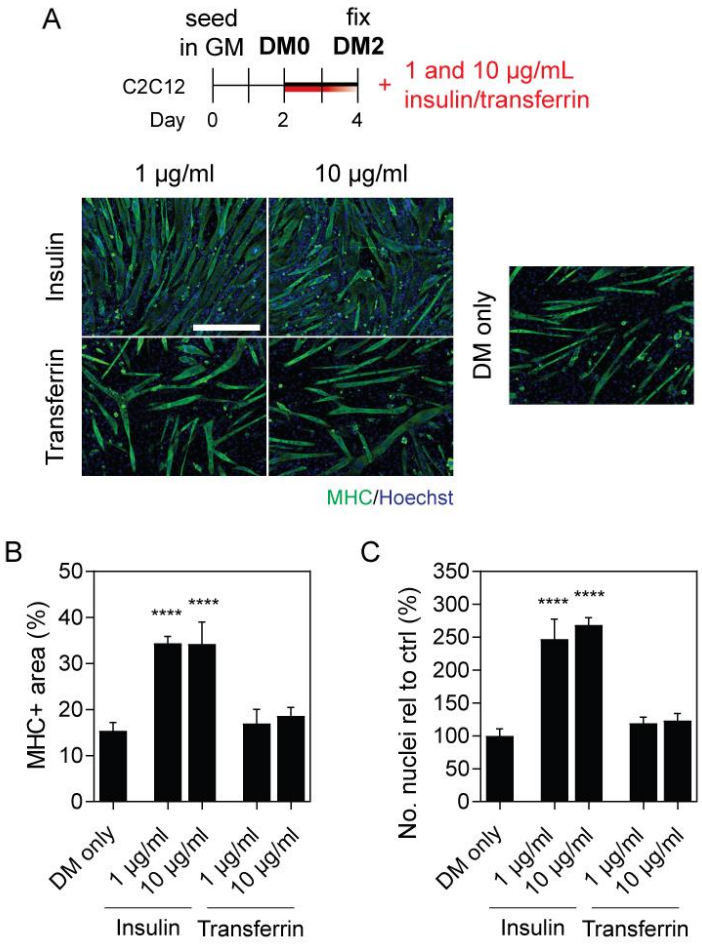


Figure 4.8 Insulin, but not transferrin, enhances myogenic differentiation in C2C12 cell cultures.

(**A**) C2C12 myoblasts were grown in GM for two days, and treated with insulin or transferrin (1 or 10 µg/mL) at the time of switching to DM. Myogenic differentiation was assessed by MHC IF at DM2, and (**B**) quantified by measuring the MHC+ area. Untreated (DM only) cultures were included as negative controls. (**C**) The total number of nuclei per representative field of view are shown as a percentage relative to the control group. All microscopy images were taken at 10× magnification. Scale bar represents 400 µm. All values are means shown + SEM from four fields of view obtained in one independent experiment. Statistical significance was determined by a one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the DM only control. **** = $P < 0.0001$.

4.5 Discussion

The results from this chapter demonstrate that pure muscle cell-derived EVs (**Figure 4.4**), but not the non-vesicular secreted 'protein' fraction (**Figure 4.5**), obtained using UF-SEC exhibits a pro-myogenic potential that is independent of the stage of myogenic differentiation in the donor cell culture (i.e. C2C12 myoblasts vs myotubes). This effect is likely co-ordinated by a complex pool of heterogenous molecules as was suggested by the further enhanced pro-myogenic effects observed following the transfer of EV-containing crude CM from differentiating C2C12 myotubes (**Figure 4.3**) as compared to the highly pure EV population obtained by UF-SEC (**Figure 4.4**). Furthermore, C2C12 myogenic differentiation was inhibited at the highest dose of EVs, suggesting that there is a range of functional doses that should be carefully considered in future experiments (**Figure 4.4**). Additionally, the non-vesicular fractions did not influence myogenic differentiation unless EVs were collected after culturing the C2C12 myotubes in Opti-MEM.

The field of study based on naturally-derived EVs offers several exciting avenues to developing non-viral methods of therapeutic delivery. Furthermore, employing EVs to deliver drugs and/or particular qualities derived from their cell- or tissue-of-origin provides a means for regenerative medicine without the need for cell therapy, and thereby avoids issues associated with immune rejection provided that human leukocyte antigens (HLAs) are not presented on the EV surface [297]. Trafficking of biomolecules using EVs also offers several other advantages over viral-mediated delivery, including flexibility in cargo size and biotype (i.e. DNA, RNA, and protein), stability within the extracellular environment, and the potential for customisable targeting through engineering of the surface receptors. This chapter aimed to determine whether EVs derived from fully differentiated C2C12 myotubes possess the innate potential to promote myogenic differentiation in recipient myoblast cultures at the time of initiating differentiation, and by extension if paracrine signalling mediated by EVs might be

contributing towards muscle growth and repair. The data presented here have revealed a role for muscle-cell-derived EVs in the transfer of pro-myogenic properties to recipient myoblast cells which was independent of the stage of myogenic differentiation of the donor cell culture, thereby arguing against a myomiR-related mechanism. Furthermore, important technical considerations were highlighted as components within the Opti-MEM isolation medium were co-purified with the non-vesicular SEC fractions, and transferred highly pro-myogenic signals to recipient myoblast cultures. In particular, insulin was identified to be responsible for this highly pro-myogenic effect, although the involvement of additional factors present in Opti-MEM cannot be excluded (**Figure 4.7**).

The endocrine system plays an important role in skeletal muscle metabolism through the targeted action of receptor-mediated signalling by growth factors (e.g. insulin-like growth factors 1 and 2, IGF1 and IGF2), and hormones (e.g. growth hormone, insulin, androgens) [540]. Insulin and IGF1 are both low molecular weight (5.8 and 7.6 kDa respectively) proteins that play important roles in skeletal muscle functioning and share ~50% structural similarity [541]. Furthermore, the insulin and IGF1 receptors (IR and IGF1R respectively) both belong to the superfamily of receptor tyrosine kinases that activate similar intracellular signalling cascades involving phosphoinositide 3-kinase (PI3K)/Akt and Shc/Ras/MAP kinase axes [542]. Despite the high degree of homology between the insulin/IGF1R ligands and their cognate receptors, as well as similarities in the intracellular signalling transduction pathways, the resulting processes mediated by these two receptors are highly distinct [543]. Insulin mediates glucose uptake and metabolism, whereas IGF1 has a profound effect on myoblast proliferation and differentiation (reviewed recently elsewhere [544]). Importantly, while insulin has a high affinity for IR, it also has a low affinity for the IGF1R (and vice versa for IGF1 and the IR). Notably, the range of circulating insulin in the body is naturally maintained at picomolar concentrations [545] and, as such, the highly pro-myogenic effect of insulin and Opti-MEM on C2C12 myoblasts

observed in the present study is likely a result of aberrant signalling via insulin binding to IGF1R (resulting in greatly enhanced proliferation and myogenic differentiation) through supraphysiological levels of insulin provided in the culture medium (i.e. at micromolar concentrations). A deeper understanding of the mechanisms of cross-talk of the insulin and IGF1 pathways in muscle cell signalling is warranted, however, it is clear that the confounding effect of insulin supplementation should be carefully considered in the future design of EV transfer experiments. This is particularly relevant for methods of EV purification that employ polymer precipitation which are known to co-purify other non-vesicular proteins and soluble factors. It is important to avoid over-interpretations of data in EV transfer experiments, considering the numerous technical aspects that contribute towards the isolated product, and to consider artefacts that may be introduced, depending on the methodology used.

4.5.1 The functional role of EVs in the transfer of pro-myogenic signals

Heparin-mediated blockade of EV uptake in differentiating C2C12 myoblasts resulted in decreased myogenic differentiation, suggesting that pro-myogenic signals contained within EVs are released during myotube maturation, and mechanisms of uptake/re-uptake are involved, at least within the same culture dish. This finding was supported by the impairment of myoblast differentiation following reduction of exosome biogenesis by treatment with the NSMase2 (SMPD2) small molecule inhibitor GW4869 (**Figure 4.1**). Additionally, siRNA-mediated knock down of *Rab27a* and *Rab27b*, two key players in the exosome biogenesis pathway [280], recapitulates this effect (**Figure 4.2**). GW4869 interferes with the ceramide-dependent exosome biogenesis pathway, while RAB GTPases are involved in the ESCRT-dependent mode of exosome release [280,524–526]. NSMase2 (SMPD2) has also been shown to regulate the export of miRNAs into non-vesicular HDL [253]. Indeed, one study carried out using HEK293T cells demonstrated that treatment with GW4869 and NSMase2-targeting siRNA inhibited the release of both exosomes as well as ex-miRNAs, whereas siRNA-mediated knock

down of *PDCD6IP* (Alix, an accessory protein in the ESCRT machinery) decreased exosome release but had no effect on secreted miRNAs [519]. Furthermore, RAB11 is believed to be dispensable for exosome release in K562 cells [280], and it is possible that this could be true of other RAB GTPases in different cell model contexts as a combination knockdown of *Rab27a* and *Rab27b* was required to mediate the inhibitory effects on myogenic differentiation (**Figure 4.2**). RAB27A and RAB27B have been implicated in exosome release in many different cancer cells types [273,526,546–549], however, future studies are required to understand the precise individual role of these exosome biogenesis factors in C2C12 cells, and other models of skeletal muscle cell differentiation.

The pro-myogenic effect of EV transfer on C2C12 myogenic differentiation demonstrated in this chapter supports another publication using the same *in vitro* cell model, but a different method for EV isolation (i.e. polymer precipitation) [306]. Functional trafficking of extracellular components has also been described using other *in vitro* models of cellular differentiation, including from donor mesenchymal stem cell (MSC) to recipient C2C12 or human umbilical-vein endothelial cell (HUVEC) cultures [309], and human skeletal myoblast (HSkM) to human adipocyte-derived stem cell (HASC) cultures [308]. We [234,523], and others [307,309], have hypothesised that the protection of ex-myomiRs within EVs enables cell-cell communication of pro-myogenic signals in a paracrine manner. Data presented in this chapter have demonstrated that both MB- and MT-EVs possess transferable, pro-myogenic factors, whereas data from our group has shown that only myotubes express the myomiRs that are believed to be capable of promoting this effect [234]. It is therefore unlikely that the pro-myogenic characteristic of the MB- and MT- EVs derived from C2C12 cell culture is myomiR-specific. The same cannot be concluded for the pro-myogenic effect observed following treatment with the EV-containing fraction from crude myotube-CM fractionation as the effects of myoblast-CM fractions were not assessed here. Even so, the treatment with EV-containing crude

fractionated CM (which also retained non-vesicular particles >300 kDa in size) resulted in the transfer of highly pro-myogenic signals, resulting in up to 200% increases in the measures used to assess myogenic differentiation (**Figure 4.3**), whereas the pure EV fraction resulted in a more modest increase of up to 50% (**Figure 4.4**). While these experiments were not carried out in parallel, it is suggestive of a synergistic function of vesicular and larger non-vesicular components in myotube-CM, or perhaps the existence of an intermediary effector molecule, required for cell-cell communication and pro-myogenic signalling.

The most important consideration is that the majority of papers describing the functional role of EVs in myogenic differentiation (described in **Table 1.2**) [245,306–309,311,312] employed methods of EV isolation that result in high yields, and which are fast and efficient (i.e. either by ultracentrifugation or using commercially available polymer precipitation kits), but which are also known to co-purify non-vesicular proteins and soluble factors [267]. The use of these methods make it difficult to conclusively attribute a functional effect observed to EV-specific paracrine signalling, rather than some other extracellular contaminant. This is in contrast to UF-SEC employed in the study reported here, which compromises on yields but results in a higher purity of EV subpopulations isolated [267]. These findings are important as there is a risk that biological effects could be misattributed to EVs when these are in reality caused by co-purified factors when employing polymer precipitation and ultracentrifugation methods for EV isolation.

4.5.2 The effective dose of EVs in transfer experiments

The stoichiometry of miRNA loading into EVs warrants further investigation, particularly as it has been suggested that natural levels of miRNAs in plasma-derived EVs from humans may not be sufficient to induce gene regulatory effects in target cells [550,551]. A recent review published by our group has hypothesised that circulating ex-myomiRs in plasma may not be

particularly physiologically relevant, and instead that the functional transfer of ex-myomiRs in EVs occurs locally within the muscle tissue interstitium where miRNA concentrations are likely to be higher (and possibly physiologically meaningful) [523]. The proposed mechanism is through a directionality effect created between mature myofibres expressing high levels of ex-myomiRs protected within EVs into the proximal interstitial fluid, and quiescent satellite cells or newly regenerating fibres that do not express these pro-myogenic signals [523]. Close-range paracrine signalling in skeletal muscle tissue by EV-encapsulated myomiRs has recently been reported *in vivo* by Mytidou and colleagues, wherein different skeletal muscle groups within the mouse hindlimb were shown to express unique patterns of myomiRs (miR-1, miR-133, and miR-206) [312]. Future *in vivo* studies that use EV isolation methods that do not co-precipitate non-vesicular factors are required to fully elucidate the role of the functional transfer of muscle-specific ex-miRNAs in EVs.

An interesting finding from the experiments described in this chapter was the suppressive effect of both MB- and MT-EVs at the highest dose tested (2×10^{11} particles/mL, corresponding to ~1 mg/mL of protein) (**Figure 4.4**). Choi and colleagues reported that a high dose of EVs (200 µg/mL of total protein) isolated from differentiating HSkMs resulted in decreased proliferation when transferred to HASC cultures (**Table 1.2**) [308]. The authors of this study did not, however, assess the effect of this dose on myogenic differentiation. Notably, a recent study has reported that inflammatory signals are transferable within EVs derived from fully differentiated C2C12 myotubes (isolated using polymer precipitation) following stimulation with tumour necrosis factor-α (TNF-α) and interferon- γ (INF-γ) cytokines, which resulted in the suppression of myogenic differentiation in recipient myoblast cultures [552]. The data from these studies [308,552], and that reported here, suggest that additional antagonistic factors may exist within MB- and MT-EVs, which could exhibit a threshold of effect that is not apparent at all EV concentrations.

4.5.3 Conclusions

The results reported in this thesis, and by other studies described in **Table 1.2** [245,306–309,311,312], highlight the complex nature of research in the field of EV-mediated transfer of bioactive molecules. To date, no perfect technique has been developed for isolating and assessing the function of highly specific subpopulations of EVs from the heterogenous pool of extracellular biomolecules. The importance of determining the dose of EVs required to induce a given physiological response was highlighted a decade ago [553], but remains unknown in many biological contexts. The ability to translate research into understanding the role of EVs in the functional transfer of biomolecules that is conducted in simple two-dimensional cell culture experiments to that of the complex muscle tissue system remains a significant challenge. Nonetheless, the work in this chapter demonstrates that highly pure preparations of EVs derived from differentiating muscle cells have the potential to transfer pro-myogenic signals to myoblasts when provided at appropriate concentrations, while the non-vesicular fraction is unable to transfer pro-myogenic signals to recipient cells. These findings are promising for the future development of engineered vesicles loaded with relevant therapeutic cargo that, once identified, additionally possess innate pro-regenerative properties which could be useful for neuromuscular disease.

5 Results III - Elucidating the functional role of upstream open reading frames

5.1 Introduction

A plethora of diseases are caused by loss-of-function mutations, and development of therapeutic strategies focused towards increasing protein expression is highly desirable. A small molecule compound has been developed by PTC Therapeutics called Translarna (Ataluren) for the promotion of stop codon readthrough in patients with nonsense mutations, and currently has conditional approval for use by the European Medicines Agency (but not the FDA) in several countries mainly within Europe [554,555]. A current exploratory approach to DMD therapeutic development is candidate screening for small molecule activators of eEF1A (a human translation initiation factor) that facilitate upregulation of utrophin protein expression [556]. Transcript-specific, rather than global activators, of protein translation are highly attractive alternative therapeutic strategies, however, to date progress has been minimal. The *in situ* manipulation of upstream open reading frames (uORFs) has the potential to upregulate protein expression in a highly targeted manner, and offers a novel approach to therapy development for a great number of human diseases, including neuromuscular disorders.

Several studies have carried out global predictions of uORFs that are encoded within the 5′ UTR of human [324,353,359,361,386,557,558] or mouse [324,353,359,361] genomes, as well as in other animal species [353]. While these global predictive studies each provide valuable insights into the possible extent of the uORFome, there are differences in the bioinformatics methodologies used, and the data files are not easily accessible. Ribosome profiling (also known as ribosome sequencing, Ribo-Seq) is a high-throughput approach to monitoring the peptide-coding potential of open reading frames across the entire transcriptome [559,560]. This method maps the exact positions of actively translating ribosomes through the use of harringtonine to inhibit

initiation within the first few peptide bond formation events [561,562]. Ribosomes occupy ~30 nt of mRNA sequence and, as such, nuclease footprinting enables isolation of ribosome-protected fragments (RPFs) to which deep sequencing is applied for mapping actively translated sequences to the transcriptome. Such approaches are useful for providing a layer of integrated information to suggest candidate uORFs which are likely to be functional. Further experimental validation of uORF activity is required to confirm the functionality of predicted uORFs in downstream ORF translational repression. A simple means to achieve this end is through the use of reporter gene assays, whereby the 5′ UTR containing the putative uORF is cloned upstream of the reporter gene pORF (e.g. luciferase), and differences in expression compared between the construct containing the native uORF (i.e. the WT construct with an ATG start codon) and that with a start codon mutation (i.e. an ATG > TTG mutation) [359,362,563].

Multiple features have been predicted to be involved in dictating uORF activity [564]. These include uORF proximity to the m^7G cap and positioning within the 5′ UTR [564,565], the primary sequence context around the AUG (i.e. Kozak consensus sequence) [318,566], amino acid identity of the micropeptide and its length [376,377,567], as well as nucleic acid sequence secondary structures [568]. Understanding the relative contribution of such uORF features on translational repressive activity is important to fully elucidate the functional role of uORFs. The overarching aim of the research presented here is to expand our current knowledge surrounding uORF-mediated translational control in the human genome, and to obtain new insights into the key features that contribute towards uORF functional activity as a step towards the development of targeted protein upregulation therapies for neuromuscular diseases.

5.2 Hypothesis and aims

I hypothesise that uORF activity is influenced by specific structural and biophysical features which affect uORF capacity to repress translation of the downstream coding sequence.

This hypothesis was tested by:

- Validating uORF activity in a panel of genes.
- Investigating whether uORF micropeptides are able to act in *trans*.
- Assessing the effect of modifying particular uORF features on uORF activity, including uORF:
    - Kozak sequence context,
    - length,
    - amino acid composition,
    - nucleic acid sequence and structure, and
    - position within the 5′ UTR.
- Applying RNA editing to disrupt uORFs *in situ* using CRISPR-mediated RNA editing.

5.3 Materials and Methods

5.3.1 Cell culture

Human embryonic kidney 293T (HEK293T) cells were grown in Dulbecco's Modified Eagle Medium containing glucose (DMEM; Thermo Fisher Scientific, MA, USA), supplemented with 10% Foetal Bovine Serum (FBS; Thermo Fisher Scientific) and 1% Antibiotic-Antimycotic (PSA: Penicillin, Streptomycin and Amphotericin B; Thermo Fisher Scientific).

5.3.2 uORF prediction

A bioinformatics analysis of the human and mouse genomes was performed by Dr Thomas Roberts to identify all putative uORFs within the 5′ UTR of coding transcripts (i.e. the 'NM_' RefSeq IDs), upstream of the primary ORF (pORF). Below is a brief summary of how this was carried out.

BED files were downloaded using the UCSC Table Browser data retrieval tool for the human (GRCh38/hg38) and mouse (GRCm38/mm10) genomes. A search was conducted to identify all canonical 'start' codons (i.e. upstream ATG trinucleotides, uATGs) occurring within 5′ UTR sequences, and the entire transcript searched for potential 'stop' codons (TAA, TAG, or TGA). Output was limited to in-frame stop codons (using the modulo operator, i.e. mod 3 = 0). The minimal possible uORF was defined as 6 nt in length (i.e. methionine-Stop).

The strength of the Kozak context sequence surrounding the uAUG codon was analysed, and scored as being 'strong' where a guanine nucleotide (G) was present at the +4 position relative to the start of the uORF, and either an adenine (A) or a G at the -3 position [318]. Sequence logos were generated using WebLogo [569].

Other outputs from the bioinformatic search included the RNA space co-ordinates (i.e. position along the length of the mature transcript) which enable the determination of the uORF

length and whether it spans the translation initiation site (TIS) of the pORF. The gDNA co-ordinates were calculated such that BED files of uORF predictions could be generated for visualisation, and the bigWigAverageOverBed v2 (https://genome.ucsc.edu/goldenpath/help/bigWig.html) programme used to calculate the mean phastCons scores for each.

Lists of oncogenes and tumour suppressor genes (TSGs) used to validate these putative uORFs were downloaded from the Cancer Gene Census (COSMIC, accessed March 2018) [570].

5.3.3 Luciferase assays

Successful cloning of the dual luciferase constructs (described in detail in **Section 5.4**) was confirmed by Sanger sequencing using the primer listed in **Table 5.1**.

Table 5.1 Sequencing primer for confirmation of pDLR vector creation

ID	Sequence (5' to 3')
pSV40_Seq_Fwd	CCGCCTCGGCCTCTGAG

5.3.3.1 Dual-Glo Luciferase Assay System

HEK293T cells were seeded at 2×10^4 cells in 0.1 mL of DMEM (10% FBS and 1% PSA) per well in a 96 well multiplate. Cells were transfected 24 hrs later with 100 ng of dual-luciferase reporter plasmid DNA (pDLR) per well in quadruplicate using Polyethylenimine (PEI, Sigma-Aldrich, MO, USA) at a weight/weight (w/w) ratio of 1:4 (DNA:PEI). Complexes were formed in DMEM containing no supplements for 20 mins at room temperature, and the complex mixture (20% of the total well volume) was added drop-wise to the cells. A Dual-Glo Luciferase Assay (Promega, WI, USA) was performed 24 hrs later to assess the effect of uORF activity on *Renilla* luciferase expression. Firefly luciferase expression was measured as an internal normalisation control. Briefly, cells and reagents were equilibrated to room temperature, half of the well volume was removed, and an equal volume of Dual-Glo Reagent

was added to each well. Samples were mixed thoroughly by pipetting, and incubated at room temperature for 20 mins before the lysate was transferred to a Greiner 96 well plate (Merck, NJ, USA). Black plates were used to reduce luminescence cross-talk between wells. Firefly luciferase expression was measured using the CLARIOstar plate reader (BMG Labtech, Aylesbury, UK; at a focal height of 9.0 mm, with no filter), and then quenched by the addition of an equal volume of Stop & Glo Reagent (prepared as a 1:100 dilution in Stop & Glo buffer) to the initial culture medium. After 10 mins at room temperature, *Renilla* luciferase expression was measured using the plate reader (with the same parameters as above), and these values were normalised to the levels of Firefly luciferase. All luciferase assay experiments were carried out at least twice.

5.3.3.2 Nano-Glo HiBiT Lytic Detection System

HEK293T cells were seeded at 1×10^5 cells in 0.5 mL of DMEM (10% FBS and 1% PSA) per well in a 24 well multiplate. Cells were transfected 24 hrs later with 500 ng of pDLR per well using PEI transfection reagent, as described in **Section 5.3.3.1**. Levels of HiBiT-tagged uORF-encoded micropeptides were measured 48 hrs later using the Nano-Glo HiBiT Lytic Detection System (Promega). Cells were treated at various timepoints prior to HiBiT protein detection with 10 μM of MG132 proteasome inhibitor (Sigma Aldrich). MG132 was included to facilitate protein accumulation with the aim to improving detection of the tagged micropeptide. Plates and buffers were first equilibrated to room temperature before half of the well volume was removed, and an equal volume of Nano-Glo HiBiT Lytic Reagent (1:100 of LbBiT Protein, and 1:50 of Nano-Glo HiBiT Lytic Substrate in Lytic Buffer) was added to the culture medium remaining. Samples were mixed thoroughly by pipetting and incubated for 10 mins at room temperature. NanoLuc expression was measured in triplicate, using the CLARIOstar plate reader (at a focal height of 9.0 mm, and an integration time of 1s). Background expression from untreated cells was subtracted from the values obtained.

5.3.4 RT-qPCR

HEK293T cells were transfected at 1×10^5 cells in 0.5 mL of DMEM (10% FBS and 1% PSA) per well in a 24 well multiplate, treated with 0.5 ng of the appropriate pDLR construct by transfection, and cells harvested 24 hrs later. Transfections were carried out in triplicate (i.e. three independent experiments) to obtain replicates for RT-qPCR. RNA was extracted using the Maxwell RSC simplyRNA Tissue Kit (Promega), as described in **Section 2.5.1**. RT-qPCR was carried out using Power SYBR Green Master Mix (Thermo Fisher Scientific) as described in **Section 2.6.2** using the primers listed in **Table 5.2**.

Table 5.2 Primer sequences for RT-qPCR detection of *RLuc* and *FLuc* mRNA transcripts

ID	Sequence (5' to 3')
qRLuc_Fwd	GTAACGCTGCCTCCAGCTAC
qRLuc_Rev	CCAAGCGGTGAGGTACTTGT
qFLuc_Fwd	ACTCTAAGACCGACTACCAGG
qFLuc_Rev	GTAGACCCAGAGCTGTTCATG

5.3.5 Transgene attenuation by synthetic uORFs

A series of recombinant 5′ UTR-iRFP (a scAAV vector expressing a near infrared protein from a CMV promoter) constructs were generated by Gibson assembly cloning using the NEBuilder HiFi DNA Assembly Kit (New England Biolabs, MA, USA), as per the manufacturer's instructions. Bacterial transformation and plasmid isolation was carried out as described in **Section 2.2.3**.

HEK293T cells were seeded at 2×10^4 cells in 0.1 mL of DMEM (with 10% FBS) per well in a 96 well multiplate. Cells were transfected 24 hrs later with 100 ng of scAAV-iRFP vector per well in quadruplicate using PEI transfection reagent, as described in **Section 5.3.3.1**. Cells were washed with PBS and lysed in RIPA buffer (Pierce, Thermo Fisher Scientific)

before levels of iRFP expression were measured using the CLARIOstar plate reader (at a focal height of 9.0 mm, with the Cy5.5 detection setting).

5.3.6 CRISPR/ADAR-mediated RNA editing

The ADAR-expressing plasmid used was pC0053-CMV-dPspCas13b-GS-ADAR2DD(E488Q)-delta-984-1090 (Addgene plasmid # 103869 ; http://n2t.net/addgene:103869 ; RRID:Addgene_103869), and the crRNA guide-expressing plasmid used was C0043-PspCas13b crRNA backbone (Addgene plasmid # 103854 ; http://n2t.net/addgene:103854 ; RRID:Addgene_103854).

DNA oligonucleotides encoding the plus and minus strands (IDT, NJ, USA) (**Table 2.1**) of the guide targeting regions for the CRISPR/ADAR system (i.e. crRNAs) were cloned into the *BbsI*-digested crRNA backbone vector, according to the protocol in **Section 2.2**. The crRNA guide sequences are listed in (**Table 5.3**).

Table 5.3 Sequences of CRISPR/ADAR crRNAs targeted to the *HOXA11* 5′ UTR

ID	Sequence (5' to 3')
crRNA1	CGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTGAAG
crRNA2	CCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTGAA
crRNA3	CGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTGAA
crRNA4	TCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTGA
crRNA5	CTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTG
crRNA6	TCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTTG
crRNA7	TCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTTT
crRNA8	CTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCTT
crRNA9	GGGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCT
crRNA10	GCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCT
crRNA11	CTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTCT
crRNA12	GGGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTC
crRNA13	GGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCTC
crRNA14	GGGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCCT
crRNA15	TGGGCTCTCCGCAGTAGCCGAGCTTAACACGATTCTCCACTGCAGCTGCC

Note: cytosine (C) nucleotides indicated in red are mismatched with the target adenosine (A) to facilitate adenosine to inosine (A > I) conversion by CRISPR-ADAR-mediated RNA editing.

HEK293T cells were seeded at 2×10^4 cells in 0.1 mL of DMEM (with 10% FBS) per well in a 96 well multiplate. Cells were transfected in quadruplicate with separate constructs expressing ADAR (37.5 ng/well), *HOXA11*-uORF targeting crRNA (75 ng/well), and *HOXA11* WT or *HOXA11* TTG plasmids (25 ng/well) using PEI transfection reagent, as described in **Section 5.3.3.1**. Treatments with a non-specific crRNA sequence were included as negative controls. Dual-luciferase assays were carried out 72 hrs post transfection (as described in **Section 5.3.3**).

5.4 Results

5.4.1 Prediction of uORFs in the human and mouse transcriptomes

Prediction of putative uORFs within the 5′ UTR of human and mouse coding transcripts (i.e. the 'NM_' RefSeq IDs) was carried out computationally using a custom script written and executed by Dr Thomas C. Roberts. A total of 25,503 (59.3%) human, and 14,479 (48.5%) mouse genome transcripts were predicted to encode at least one uORF (**Figure 5.1A**). This supports previous estimates from studies using earlier genome builds [359,361,571,572]. The majority of all uORF-containing transcripts had only a single uORF, while ~6% of human, and ~4% of mouse, transcripts had 10 or more uORFs (**Figure 5.1B**). Approximately half of the predicted uORFs were ~6-30 amino acids in length for both the human and mouse genomes, and overall the these ranged from 1 to ~ 5,000 amino acids (**Figure 5.1C**). More than 80% of uORFs were found to exist solely within the 5′ UTR, while the remainder overlapped with the translation initiation site (TIS) of the primary ORF (pORF) (**Figure 5.1D**). The distance between the transcription start site (TSS) and the start of the uORF (**Figure 5.1E**), as well as between the uORF and the pORF (**Figure 5.1F**), was ~100-400 nt for the majority of transcripts in each group, and ranged from 1 to ~6,000 nt for the human genome, and 1 to ~3000 nt for the mouse genome (**Figure 5.1E and F**). Approximately half of all uORFs were predicted to be terminated by the UGA stop codon, and 20-30% by UAA or UAG (**Figure 5.1G**). Generally, the Kozak consensus sequence of uORFs was weaker than that of pORFs [318] (Kozak contexts with a +4 G, and a -3 A or G, relative to the start of the ORF, were classified as strong, and otherwise as weak) (**Figure 5.1H**). The Kozak consensus sequence is represented byWebLogo plots, where the height of the letter corresponds with the relative frequency, and with the most frequent nucleotides placed higher on the stack [569] (**Figure 5.1I**). The genomic co-ordinates of the predicted uORFs were used to determine the average phastCons scores for each uORF, and this was compared with the coding DNA sequence (CDS), 5′ and 3′ UTRs, and intronic regions

[573]. The inter-quartile range of uORF conservation values spanned almost the entire possible range, indicating that some are highly conserved, while others are poorly conserved across species (**Figure 5.1J**). The characteristics of the predicted uORFs were similar for both the mouse and human genomes (**Figure 5.1A-J**). The predicted uORFs were mapped onto the UCSC Genome Browser Web interface, and an example of this is shown for the *Huntingtin* (*HTT*) gene along with publicly available riboSeq data for the region (**Figure 5.1K**). A peak of initiating ribosomes at the predicted uORF start suggests that the uORF micropeptide is translated and likely to be functional.

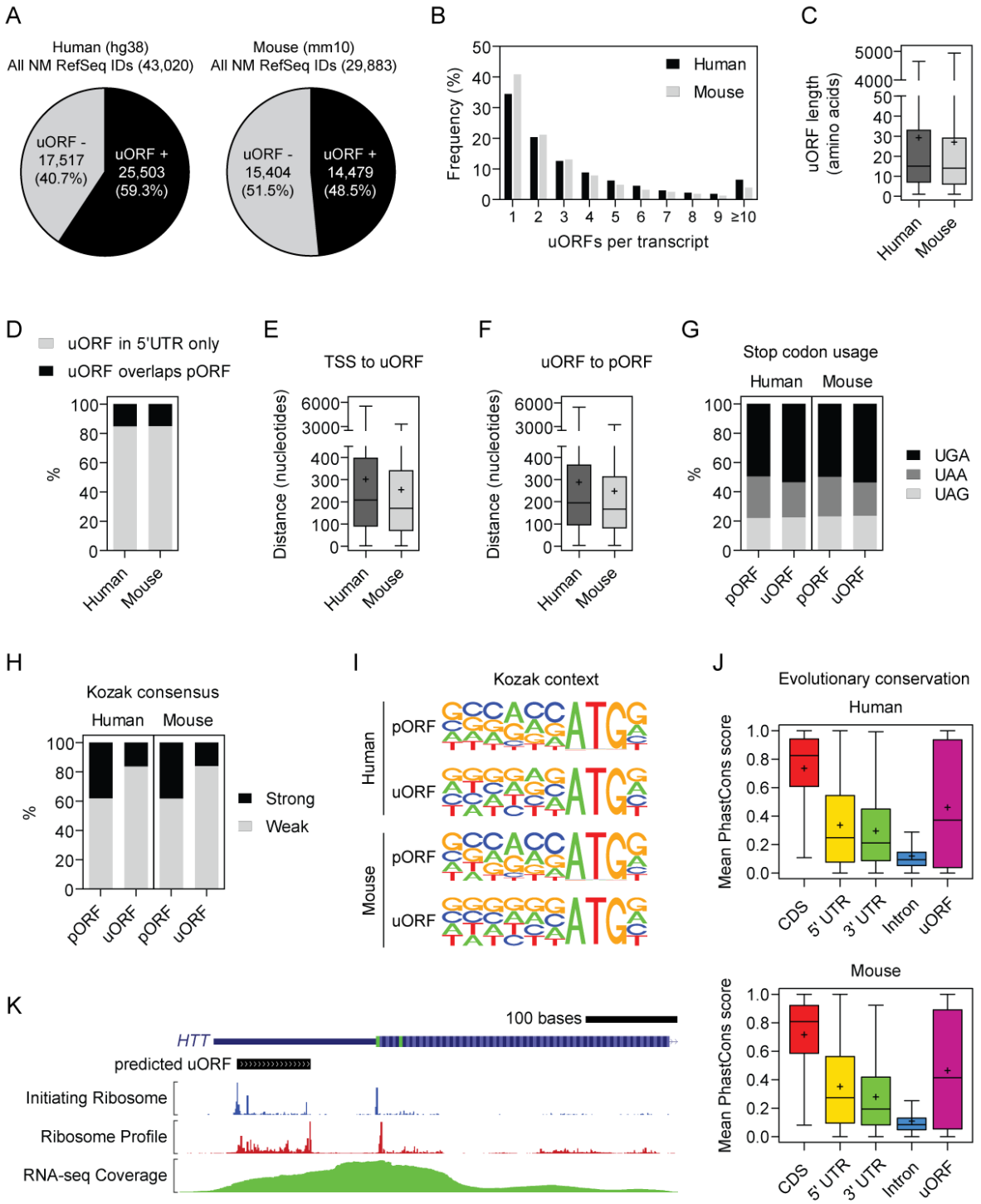

A
Human (hg38)
All NM RefSeq IDs (43,020)
uORF - 17,517 (40.7%)
uORF + 25,503 (59.3%)
Mouse (mm10)
All NM RefSeq IDs (29,883)
uORF - 15,404 (51.5%)
uORF + 14,479 (48.5%)
B
Frequency (%)
Human
Mouse
uORFs per transcript
C
uORF length (amino acids)
Human
Mouse
D
uORF in 5'UTR only
uORF overlaps pORF
%
E
TSS to uORF
Distance (nucleotides)
F
uORF to pORF
G
Stop codon usage
UGA
UAA
UAG
pORF
uORF
H
Kozak consensus
Strong
Weak
I
Kozak context
J
Evolutionary conservation
Mean PhastCons score
CDS
5' UTR
3' UTR
Intron
K
100 bases
HTT
predicted uORF
Initiating Ribosome
Ribosome Profile
RNA-seq Coverage

Figure 5.1 Properties of computationally predicted uORFs.

A bioinformatic pipeline designed and executed by Dr Thomas C. Roberts was used to identify and characterise putative upstream open reading frames (uORFs) within the human (GRCh38/hg38) and mouse (GRCm38/mm10) genomes using 'NM_' RefSeq IDs. (**A**) Pie charts illustrating the proportion of transcripts containing at least one predicted uORF. (**B**) Distribution of the number of predicted uORFs per transcript. (**C**) Box plots of the distribution of uORF length (number of amino acids). (**D**) Proportion of transcripts containing a uORF within the 5′ UTR only, and those with uORF(s) that span the translation initiation site (TIS) of the primary ORF (pORF). Box plots showing the distribution of the distance (number of nucleotides) between (**E**) the transcription start site (TSS) and the start codon of the uORF, and (**F**) the start of the uORF and the start of the pORF. (**G**) Stop codon usage frequencies (UGA, UAA, or UAG) of the pORF and cognate uORF. The predicted strength of the Kozak consensus within the start codon context of pORFs and uORFss were (**H**) scored as 'weak' or 'strong' and (**I**) represented as sequence logos. The height of the letter represents its proportional frequency. (**J**) Box plots of the phastCons scores of evolutionary conservation of the coding DNA sequence (CDS), 5′ UTR, 3′ UTR, intronic sequences, and uORFs for uORF-containing transcripts. (**K**) Representation of the predicted uORF within the *Huntingtin* (*HTT*) 5′ UTR obtained from the UCSC Browser. The initiating ribosome, ribosome profile and RNA-seq coverage are shown below for the *HTT* uORF and pORF. Scale bar represents 100 bp. Box plots show the median and interquartile range, whiskers represent the minimum and maximum values, and the mean value is represented by '+'.

5.4.2 Validation of predicted uORF functionality

The dual-luciferase reporter (DLR) assay provided a simple means to validate the activity of a panel of putative uORFs, requiring only transient transfection of plasmid constructs, and light-based detection of reporter gene expression using a luminometer. Furthermore, the DLR assay is highly sensitive, and includes an internal control which enables sequential measurement of both luminescent reporters from the same sample. This provides more accurate data for transient transfection experiments than other commonly used normalisation methods, such as total protein content or co-transfection with reporter plasmids. The dual-luciferase reporter vector (psiCheck2.2, referred to here as pDLR) expressing human codon optimised versions of the *Renilla* and Firefly luciferase genes (RLuc and FLuc respectively) was employed to assess the activity of individual uORFs for the repression of pORF translation. The pDLR vector was modified to introduce a 5′ UTR cloning site downstream of the simian virus 40 (SV40) promoter, and upstream of the RLuc gene, creating the pDLR 5′ UTR vector (**Figure 5.2**). Double digestion with *StuI* and *DraIII* restriction enzymes facilitated the removal of the chimeric intron and T7 promoter regions within the original vector, and allowed for the introduction of a *BsmBI* cloning site by Gibson assembly. *BsmBI* is a type II restriction enzyme that cleaves outside of its recognition sequence, and thus was selected to enable seamless integration of the 5′ UTR of interest immediately upstream of the RLuc start codon via the Golden Gate assembly method [450].

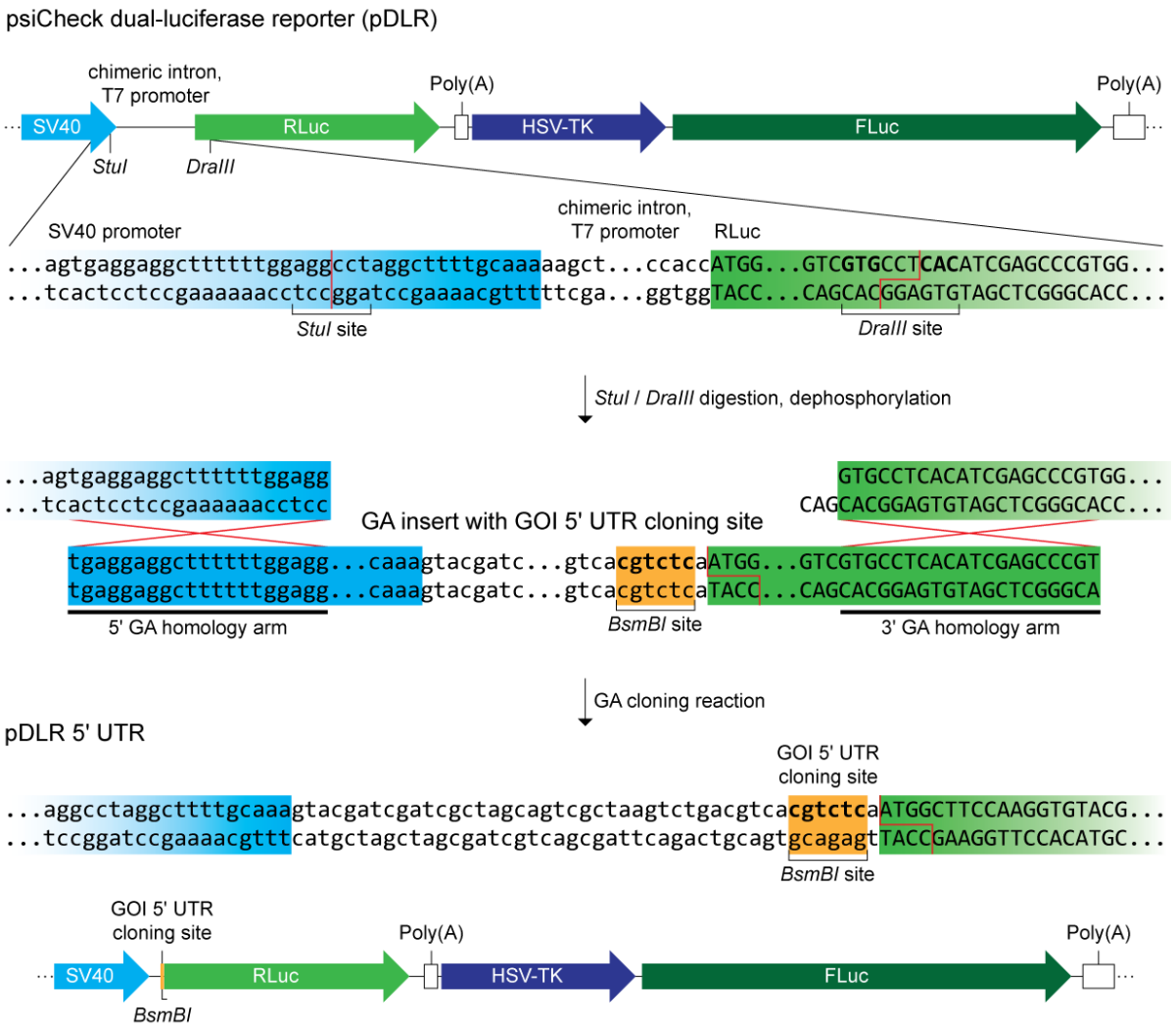

psiCheck dual-luciferase reporter (pDLR)
chimeric intron, T7 promoter
Poly(A)
Poly(A)
SV40
RLuc
HSV-TK
FLuc
StuI
DraIII
SV40 promoter
chimeric intron, T7 promoter
RLuc
...agtgaggaggctttttggaggcctaggcttttgcaaaaagct...ccaccATGG...GTCGTGCCTCACATCGAGCCCGTGG...
...tcactcctccgaaaaaacctccggatccgaaaacgtttttcga...ggtggTACC...CAGCACGGAGTGTAGCTCGGGCACC...
StuI site
DraIII site
StuI / DraIII digestion, dephosphorylation
...agtgaggaggctttttggagg
...tcactcctccgaaaaaacctcc
GTGCCTCACATCGAGCCCGTGG...
CAGCACGGAGTGTAGCTCGGGCACC...
GA insert with GOI 5' UTR cloning site
tgaggaggctttttggagg...caaagtacgatc...gtcacgtctcaATGG...GTCGTGCCTCACATCGAGCCCGT
tgaggaggctttttggagg...caaagtacgatc...gtcacgtctcaTACC...CAGCACGGAGTGTAGCTCGGGCA
5' GA homology arm
BsmBI site
3' GA homology arm
GA cloning reaction
pDLR 5' UTR
GOI 5' UTR cloning site
...aggcctaggcttttgcaaagtacgatcgatcgctagcagtcgctaagtctgacgtcacgtctcaATGGCTTCCAAGGTGTACG...
...tccggatccgaaaacgtttcatgctagctagcgatcgtcagcgattcagactgcagtgcagagtTACCGAAGGTTCCACATGC...
BsmBI site
GOI 5' UTR cloning site
Poly(A)
Poly(A)
SV40
RLuc
HSV-TK
FLuc
BsmBI

Figure 5.2 Generation of a dual-luciferase reporter vector containing a gene-of-interest 5′ UTR cloning site.

The psiCheck dual-luciferase reporter (pDLR) vector encodes a human codon optimised *Renilla* luciferase (RLuc) gene driven by the simian virus 40 (SV40) promoter, and a Firefly luciferase gene (FLuc, also human codon optimised) driven by the HSV-TK promoter. In this study, FLuc was utilised as the internal control for transient transfection experiments while the expression of RLuc was modulated by the activity of an upstream open reading frame (uORF), present within a synthetic 5′ UTR sequence. This schematic depicts the Gibson assembly (GA) cloning strategy employed to generate a pDLR vector containing a *BsmBI* (type IIs) restriction cloning site for insertion of the 5′ untranslated region (UTR) of a gene-of-interest (GOI) immediately upstream of the RLuc start codon. The pDLR backbone was double-digested with *StuI* and *DraIII* restriction enzymes (and dephosphorylated to prevent vector re-ligation), and a GA insert containing the GOI 5′ UTR cloning site and 20 bp of flanking sequence homologous to the digested backbone was introduced by a GA cloning reaction. The novel cloning vector generated was named pDLR 5′ UTR. Coding and non-coding sequences are indicated in upper- and lowercase respectively.

The list of predicted uORFs (25,503) was refined to obtain experimental validation candidates first by intersecting it with the Cancer Gene Consensus database [574], then limiting to those with only a single uORF within the 5′ UTR (8,778 transcripts), and thereafter to those which do not overlap with the pORF (5,923 transcripts). Oncogenes and tumour suppressor genes (TSGs) with mutations that have been causally implicated in cancer were identified (130 genes), and a panel of 11 genes of interest (GOIs) manually selected from this list to represent both oncogenes and TSGs. The panel of genes included those classified as oncogenes (*JUN*, *IKBKB*, *KDR*, *MAP2K2*, and *SMO*), oncogenes and fusions (*KDM5A*), TSGs (*NCOR1*), oncogenes and TSGs (*FOXL2*, and *MAP3K1*), or oncogenes, TSGs, and fusions (*HOXA11*, and *HOXA9*). Notably, the activity of uORFs in the 5′ UTRs of *KDR*, *MAP2K2*, and *MAP3K1* have been assessed in a previous study [381]. The reason for selecting cancer-associated genes was that this provided a relatively long list of potentially disease-relevant candidates with which to assess the functional role of uORFs. Furthermore, cancer-associated genes are typically important for cell-cycle regulation and, as such, growth characteristics in cell culture could potentially be a useful and simple phenotypic readout of experimental modulation of uORF activity.

The full-length 5′ UTR sequences from the selected panel of GOIs were cloned into pDLR 5′ UTR vectors by Gibson assembly [575] (**Figure 5.3A**). Two versions of each pDLR GOI 5′ UTR vector were created, one with the wild-type (WT) uORF sequence, and another with an ATG > TTG start codon mutation to disrupt initiation of uORF translation, thereby knocking out uORF activity. Representative WT and TTG pDLR vector sequences are shown for the *HOXA11* 5′ UTR (**Figure 5.3B**). The *HOXA11* 5′ UTR is 72 nt in length, and the putative uORF encodes an 11 amino acid micropeptide which is indicated below the DNA sequence.

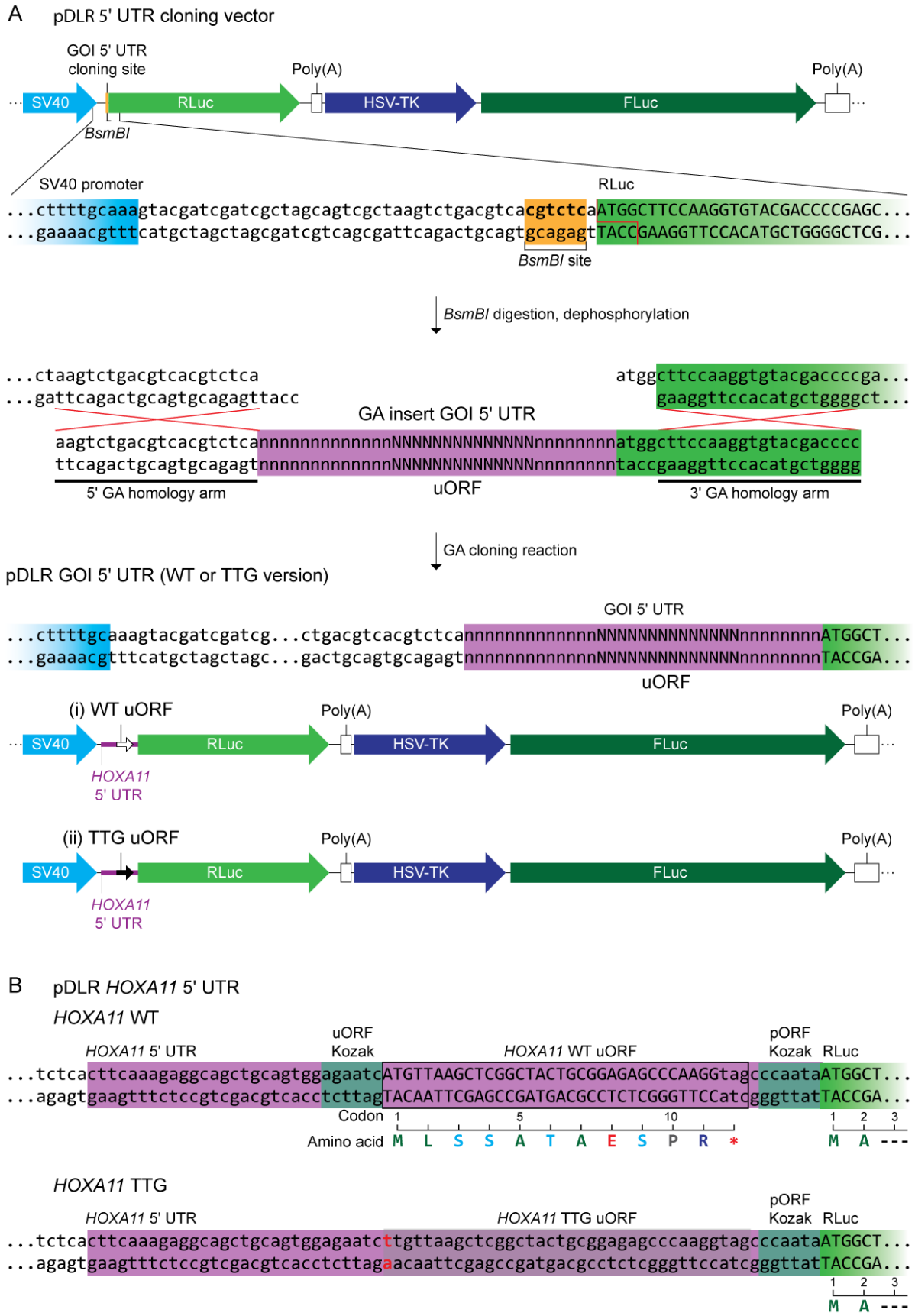

A
pDLR 5' UTR cloning vector
GOI 5' UTR cloning site
Poly(A)
Poly(A)
SV40
RLuc
HSV-TK
FLuc
BsmBI
SV40 promoter
RLuc
...cttttgcaaagtacgatcgatcgctagcagtcgctaagtctgacgtcacgtctcaATGGCTTCCAAGGTGTACGACCCCGAGC...
...gaaaacgtttcatgctagctagcgatcgtcagcgattcagactgcagtgcagagtTACCGAAGGTTCCACATGCTGGGGCTCG...
BsmBI site
BsmBI digestion, dephosphorylation
...ctaagtctgacgtcacgtctca
...gattcagactgcagtgcagagttacc
atggcttccaaggtgtacgaccccga...
gaaggttccacatgctggggct...
GA insert GOI 5' UTR
aagtctgacgtcacgtctcannnnnnnnnnnnnNNNNNNNNNNNNNNnnnnnnnnatggcttccaaggtgtacgacccc
ttcagactgcagtgcagagtnnnnnnnnnnnnnNNNNNNNNNNNNNNnnnnnnnntaccgaaggttccacatgctgggg
5' GA homology arm
uORF
3' GA homology arm
GA cloning reaction
pDLR GOI 5' UTR (WT or TTG version)
GOI 5' UTR
...cttttgcaaagtacgatcgatcg...ctgacgtcacgtctcannnnnnnnnnnnnNNNNNNNNNNNNNNnnnnnnnnATGGCT...
...gaaaacgtttcatgctagctagc...gactgcagtgcagagtnnnnnnnnnnnnnNNNNNNNNNNNNNNnnnnnnnnTACCGA...
uORF
(i) WT uORF
Poly(A)
Poly(A)
SV40
RLuc
HSV-TK
FLuc
HOXA11 5' UTR
(ii) TTG uORF
Poly(A)
Poly(A)
SV40
RLuc
HSV-TK
FLuc
HOXA11 5' UTR
B
pDLR HOXA11 5' UTR
HOXA11 WT
HOXA11 5' UTR
uORF Kozak
HOXA11 WT uORF
pORF Kozak
RLuc
...tctcacttcaaagaggcagctgcagtggagaatcATGTTAAGCTCGGCTACTGCGGAGAGCCCAAGGtagcccaataATGGCT...
...agagtgaagtttctccgtcgacgtcacctcttagTACAATTCGAGCCGATGACGCCTCTCGGGTTCCatcgggttatTACCGA...
Codon 1 5 10
Amino acid M L S S A T A E S P R *
1 2 3
M A ---
HOXA11 TTG
HOXA11 5' UTR
HOXA11 TTG uORF
pORF Kozak
RLuc
...tctcacttcaaagaggcagctgcagtggagaatcttgttaagctcggctactgcggagagcccaaggtagccaataATGGCT...
...agagtgaagtttctccgtcgacgtcacctcttagaacaattcgagccgatgacgcctctcgggttccatcgggttatTACCGA...
1 2 3
M A ---

Figure 5.3 Generation of pDLR GOI 5′ UTR vectors.

(**A**) Schematic depicting the cloning strategy employed to generate a pDLR GOI 5′ UTR vector containing the 5′ UTR sequence of a gene that has a single uORF that does not span the translation initiation site (TIS). RLuc expression was thereafter under control of the recombinant uORF-containing 5′ UTR. The pDLR 5′ UTR cloning vector was digested with *BsmBI* which cleaves within the RLuc ORF (this was dephosphorylated to prevent vector re-ligation), and a GA insert containing the 5′ UTR sequence of a GOI with 20 bp homology arms on either side was introduced into the linearised vector using a GA cloning reaction. (**B**) An example of the wild-type (WT) and TTG mutant versions of the 72 nt *HOXA11* 5′ UTR cloned into the pDLR 5′ UTR vector. The *HOXA11* WT plasmid encodes an 11 amino acid uORF (depicted below the nucleic acid sequence, translation termination is shown with an asterisk). The 6 nt Kozak context sequence upstream of the start codon for the putative *HOXA11* uORF and RLuc are highlighted, as well as the predicted amino acid sequence of the encoded micropeptide (M-L-S-S-A-T-A-E-S-P-R-STOP, where equal or similar colours were assigned manually to amino acids that share similar biochemical properties). Amino acid codes: M = methionine, L = leucine, S = serine, A = alanine, T = threonine, E = glutamic acid, P = proline, R = arginine. Coding and non-coding sequences are indicated in upper- and lowercases respectively.

A number of features that are predicted to contribute towards uORF activity are described for the 5′ UTRs and uORFs of the cancer-associated gene panel (**Figure 5.4A**). These features include the Kozak context strength, the length of the uORF in amino acids, the distance in nucleotides between the 5′ cap (TSS) and the start of the uORF, the distance between the start of the uORF and start of the pORF, and the STOP codon usage. The full sequences of each 5′ UTR and translated uORF micropeptide are listed in **Table 5.4**. The effect of uORF knockout across the panel of oncogenes and TSGs was assessed by measuring changes in RLuc expression, normalised to FLuc, using a DLR assay (**Figure 5.4B**). A significant ($P < 0.001$) increase in RLuc expression was detected as a result of uORF knockout for 10 of 11 of the gene-specific 5′ UTRs tested. The *SMO* 5′ UTR resulted in the highest level of uORF-mediated translational repression, whereby removing uORF activity increased RLuc expression by 9.3-fold relative to the WT control ($P < 0.0001$). Repression of RLuc expression was relieved following uORF knockout by 4.8-fold ($P < 0.0001$) for the *HOXA11* 5′ UTR, and between 1.5 and 3-fold for 8 of the remaining genes assessed (in decreasing order of activity: *NCOR1*, *MAP2K2*, *KDM5A*, *FOXL2*, *HOXA9*, *KDR*, *IKBKB*, and *JUN*). Knockout of the *MAP3K1* uORF had no effect on luciferase reporter expression. These changes in RLuc expression cannot be explained by concomitant effects on transcript stability, such as by the induction of nonsense-mediated decay, as the RLuc mRNA levels were not significantly different between the WT and TTG versions of each pDLR GOI 5′ UTR vector (**Figure 5.4C**). Overall, the functionality of uORFs was highly variable across the panel of genes tested here, ranging from non-functional (*MAP2K2*) to highly active (*SMO*). These data validate the translational repressive activity of the computationally predicted uORFs, and consequently also the prediction pipeline itself.

No significant correlation was found between the fold-change in RLuc expression, and the uORF length ($r = 0.1892$, $P = 0.5774$), distance from TSS to uORF ($r = -0.2328$, $P =$

0.4910), or distance from uORF to pORF (r = -0.1681, P = 0.6212) (**Figure 5.4A and B**). This is perhaps unsurprising given the small sample size, and the large number of factors that have been proposed to influence uORF activity [362].

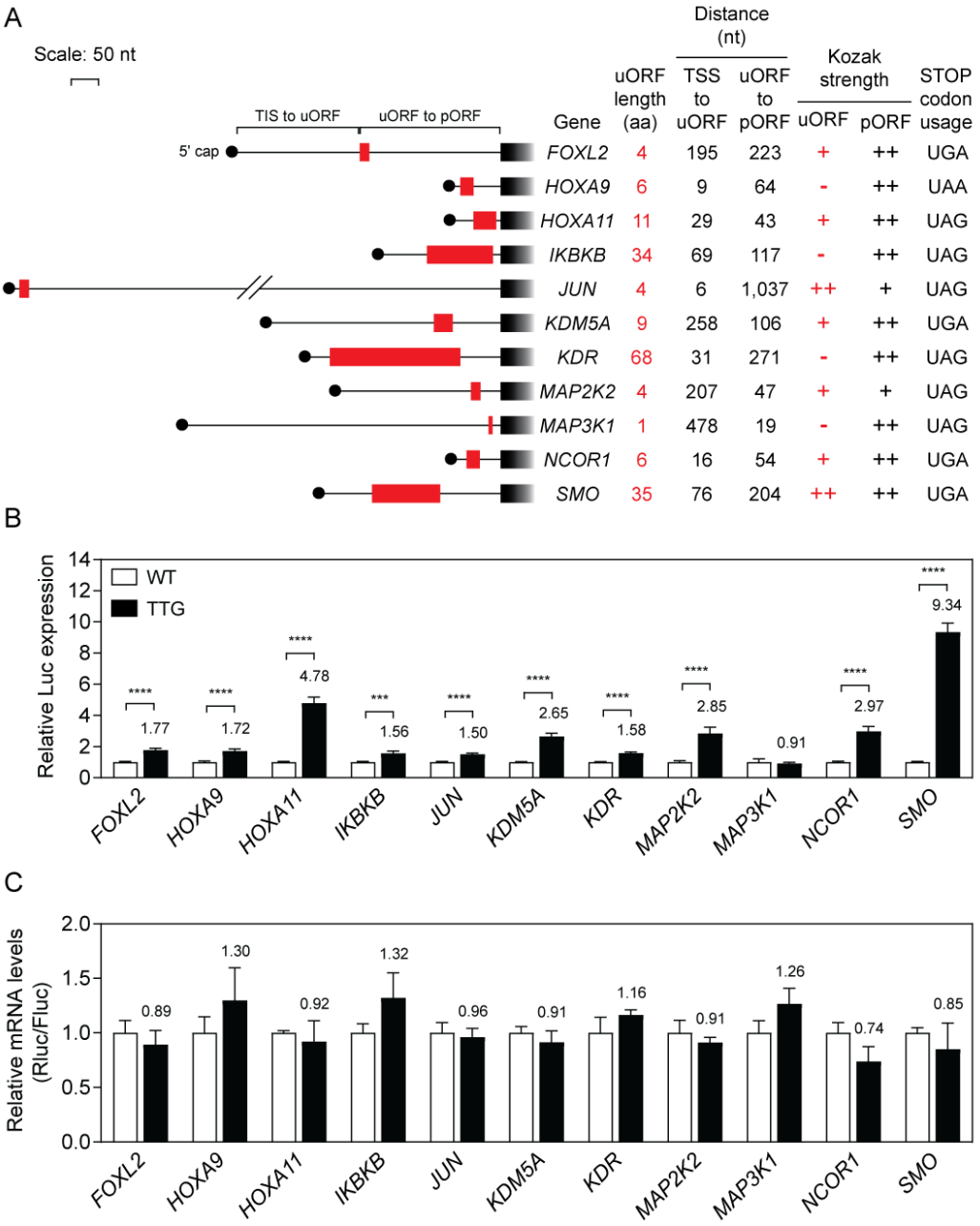

A
Scale: 50 nt
TIS to uORF
uORF to pORF
5' cap
Gene
uORF length (aa)
Distance (nt)
TSS to uORF
uORF to pORF
Kozak strength
uORF
pORF
STOP codon usage
FOXL2 4 195 223 + ++ UGA
HOXA9 6 9 64 - ++ UAA
HOXA11 11 29 43 + ++ UAG
IKBKB 34 69 117 - ++ UAG
JUN 4 6 1,037 ++ + UAG
KDM5A 9 258 106 + ++ UGA
KDR 68 31 271 - ++ UAG
MAP2K2 4 207 47 + + UAG
MAP3K1 1 478 19 - ++ UAG
NCOR1 6 16 54 + ++ UGA
SMO 35 76 204 ++ ++ UGA
B
Relative Luc expression
WT
TTG
1.77
1.72
4.78
1.56
1.50
2.65
1.58
2.85
0.91
2.97
9.34
FOXL2
HOXA9
HOXA11
IKBKB
JUN
KDM5A
KDR
MAP2K2
MAP3K1
NCOR1
SMO
C
Relative mRNA levels (Rluc/Fluc)
0.89
1.30
0.92
1.32
0.96
0.91
1.16
0.91
1.26
0.74
0.85

Figure 5.4 Validation of predicted uORFs in a panel of oncogenes and TSGs.

(**A**) Features of the 5′ UTRs of a panel of oncogenes and tumour suppressor genes (TSGs) cloned upstream of RLuc in a DLR vector. Each 5′ UTR contains a single uORF (red box) that does not span the pORF start. uORF-containing (WT) and uORF-knockout (TTG) versions of each pDLR vector were transfected into HEK293T cells (n=4 wells) and (**B**) a DLR assay carried out to assess the effect of uORF activity on RLuc protein expression. Values were normalised to FLuc expression (the internal control). (**C**) RT-qPCR of RLuc mRNA transcript levels, normalised to FLuc. Values of the WT and TTG version of each uORF are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Scale bar represents 50 nucleotides. Statistical significance was determined by a Student's *t*-test relative to the corresponding gene-specific WT control. *** = $P < 0.001$, and **** = $P < 0.0001$. TSS, transcription start site.

Table 5.4 5′ UTR and translated uORF sequences of a panel of oncogenes and TSGs

Gene	5' UTR sequence (5'-3')	uORF peptide sequence
FOXL2	atcagaacagagcgaggctcctggcgcactagggactccaggaggcag ctccgccagagacgcgggtcgtgcttcgggaaaccgggggcgggggg aggggaagagcgcagaaaagaaaacccaccaaggcggggactggcctg agcggggaggggcggcgaggccggagcccctctctgttgggcggactc ccc**ATGGCCAGAGGCtga**gctccactcccgccggccgctccctagggg aaggggaaggagaggggagagcagcgacaggcctccagcaagcaagcg cgggcggcatccgcagtctccagaagtttgagacttggccgtaagcgg actcgtgcgccccaactctttgccgcgccagcgcctggagcggagagc agaggcggcccggccgcggcgcgccggctttgtc	MARG*
HOXA9	agttgttac**ATGAAATCTGCAGTTTCAtaa**tttccgtgggtcgggccg ggcgggccaggcgctgggcacggtg	MKSAVS*
HOXA11	cttcaaagaggcagctgcagtggagaatc**ATGTTAAGCTCGGCTACTG CGGAGAGCCCAAGGtag**cccaata	MLSSATAESPR*
IKBKB	gtgacgtcagagcaggaagtgtttgaggaagtcgcgccgcgctgcccg cgttaagattcccgcatttta**ATGTTTTCAGGGGGGTGTCATAGCCCC GGGTTTGGCCGCCCCAGCCCCGCCTTCCCCGCCCCGGGGAGCCCGCCC CCTGCCCCGCGTCCCTGCCGACAGAGTtag**cacgacatcagt	MFSGGCHSPGFGRPSP AFPAPGSPPPAPRPCRQ S*
JUN	gacatc**ATGGGCTATTTTtag**ggttgactggtagcagataagtgttg agctcgggctggataagggctcagagttgcactgagtgtggctgaagc agcgaggcgggagtggaggtgcgcggagtcaggcagacagacagacac agccagccagccaggtcggcagtatagtccgaactgcaaatcttattt tcttttcaccttctctctaactgcccagagctagcgcctgtggctccc gggctggtgtttcgggagtgtccagagagcctggtctccagccgcccc cgggaggagagccctgctgcccaggcgctgttgacagcggcggaaagc agcggtacccacgcgcccgccgggggaagtcggcgagcggctgcagca gcaaagaactttcccggctgggaggaccggagacaagtggcagagtcc cggagccaacttttgcaagcctttcctgcgtcttaggcttctccacgg cggtaaagaccagaaggcggcggagagccacgcaagagaagaaggacg tgcgctcagcttcgctcgcaccggttgttgaacttgggcgagcgcgag ccgcggctgccgggcgccccctcccctagcagcggaggaggggacaa gtcgtcggagtccgggcggccaagacccgccgccggccggccactgca gggtccgcactgatccgctccgcggggagagccgctgctctgggaagt gagttcgcctgcggactccgaggaaccgctgcgcacgaagagcgctca gtgagtgaccgcgacttttcaaagccgggtagcgcgcgcgagtcgaca agtaagagtgcgggaggcatcttaattaaccctgcgctccctggagcg agctggtgaggagggcgcagcggggacgacagccagcgggtgcgtgcg ctcttagagaaactttccctgtcaaaggctccgggggggcgcgggtgtc ccccgcttgccacagccctgttgcggccccgaaacttgtgcgcgcagc ccaaactaacctcacgtgaagtgacggactgttct	MGYF*
KDM5A	ggcagaactttggcgggcaaggagacaaacagccacactgtgctgccg gggaagagaagaagaaggaaaggccagacacacaggcgtcgtgcagct gcggaaggattccgtttcctgcagggggcggagggatttgacagcacg gtaacttccggaaaagtcctgaacctcagtggaaaagccttgcaagaa acaataacaaacgggaggaggaagcgattctggggtttctgtgttgaa cggttcttgtccgcgaag**ATGCGCTTCGGCCTCTGTCAGGGGACTtga** acgggcttagtgggcttcagccagcttttctccaccggttccccacgg ggaccccccccccccccggccgttgca	MRFGLCQGT*

Gene	5' UTR sequence (5'-3')	uORF peptide sequence
KDR	actgagtcccgggaccccgggagagcggtcaATGTGTGGTCGCTGCGTTTCCTCTGCCTGCGCCGGGCATCACTTGCGCGCCGCAGAAAGTCCGTCTGGCAGCCTGGATATCCTCTCCTACCGGCACCCGCAGACGCCCCTGCAGCCGCGGTCGGCGCCCGGGCTCCCTAGCCCTGTGCGCTCAACTGTCCTGCGCTGCGGGGTGCCGCGAGTTCCACCTCCGCGCCTCCTTCTCTagacaggcgctgggagaaagaaccggctcccgagttctgggcatttcgcccggctcgaggtgcagg	MCGRCVSSACAGHHLRAAESPSGSLDILSYRHPQTPLQPRSAPGLPSPVRSTVLRCGVPRVPPPRLLL*
MAP2K2	cccctgcctctcggactcgggctgcggcgtcagccttcttcgggcctcggcagcggtagcggctcgctcgcctcagccccagcgcccctcggctacccctcggcccaggcccgcagcgccgcccgccctcggccgccccgacgccggcctgggccgcggccgcagccccgggctcgcgtaggcgccgaccgctcccggcccgcccccctATGGGCCCCGGCtagaggcgccgccgccgccggcccgcggagccccg	MGPG*
MAP3K1	agaaacccaagttggaactaattctttctttcggaaggtgcaactcccctcccgcgagctccgcggtgccgggccgagattgccgagaggaagcggcgcagcgctgccgccaaggctcctcctgtcgccggtgcggccgggactacctggcggcgcggcgcgtgcagcgcgcagagtcccgggagcccacgcctccgcctccgcccccgcccccctccgcctcccagtccacctcgcccgccgccctctcgcccggcggagagcacagcccactccctcccacctgcggccgccgggccgccctccacccacacctctgccgcaggccggacccagtgcgcccgcccgtcggtcagtccaggccaggcgcccggcgggccgcgctcacgcagttggcgcaggaggccttacgctggcggcgcagtgcccgccccctgcgctctccccgcccccctccctccctcgcaggggccgagcgaATGtagcccgcgagagaaa	M*
NCOR1	gtggtttcctagaaacATGATTGTTTATTGGCATtgatctctcagtctggtgaggacttctttactgata	MIVYWH*
SMO	gggctgaagacaacttggattgcgaggctagggcttggggagtcgtgcatcccgttccgggcctccgcagcccaacATGGGCCCCGGGTTCCAAAGTTTGCGAAGTTGGGCGCCGAGGGGCCGGGGCGCGCGGAGCGTCCGGGGGGGGCCCGGGCCCGGATTCTCTGGGCGCACAGGTCGCCtgagccgcctccgcggccgccgaggtcgtgcgtgtggccgggggctccgaggagcaggcgggggcgccggggcttttgctgagttggcggggttggcc	MGPGFQSLRSWAPRGRGARSVRGGPGPDSLGAQVA*

Note: the uORF-encoding sequence is highlighted in red, coding sequence is indicated in uppercase and non-coding sequence in lowercase.

The uORFs predicted to be within the 5′ UTRs of two genes with neuromuscular disease relevance where upregulation of expression of the associated protein is therapeutically beneficial were also validated. For example, upregulation of survivor motor neuron 2 (SMN2) and utrophin (UTRN) are expected to be therapeutically beneficial in the cases of spinal muscular atrophy (SMA) and DMD respectively. SMA is another devastating neuromuscular disease caused by loss-of-function mutations in the survival motor neuron 1 (*SMN1*) gene, resulting in impaired synaptic transmission at the neuromuscular junction [576–578]. This autosomal recessive disease is the leading cause of death in infancy, with an incidence of ~1:10,000-1:28,000 [579,580]. The loss of *SMN1* is compensated for to a certain extent by the paralogue *SMN2* which arose from a genetic duplication event. A C>T single nucleotide polymorphism (SNP) in *SMN2* exon 7 results in aberrant splicing during messenger RNA maturation, and consequently only 10-20% of normal functional protein is produced. The number of copies of *SMN2* varies naturally in the population (typically ranging from 1-8 copies per cell) and thus the differences in total protein levels accounts for the variations in severity and classification of SMA into four types (OMIM ID Type I #25330, Type II #253550, Type III #253400, and Type IV #271150). Genetic therapies that promote proper *SMN2* splicing (i.e. the ASO drug nusinersen, and small molecule risdiplam), or replace the *SMN1* gene via AAV-mediated expression (i.e. zolgensma, onasemnogene abeparvovec), are FDA-approved and currently the leading therapeutic strategies for SMA. uORF-mediated upregulation of SMN2 might therefor constitute a novel therapeutic strategy.

Both *SMN2* and *UTRN* 5′ UTRs encode predicted uORFs near the m^7G cap (**Figure 5.5A**). uORF knockout in both cases resulted in an 11% increase in luciferase reporter expression which nonetheless were statistically significant, although of debatable biological significance ($P < 0.01$ and $P < 0.001$) (**Figure 5.5B**). The functionality of the *SMN2* and *UTRN* uORFs is therefore considered to be negligible and unlikely to be therapeutically relevant. A

possible explanation for the lack of uORF activity in these contexts is that recognition of an AUG positioned close to the m^7G cap (i.e. within ~30 nt) by the scanning ribosome has been shown to be significantly impaired [565,581,582]. These data further support the idea that not all predicted uORFs are functional.

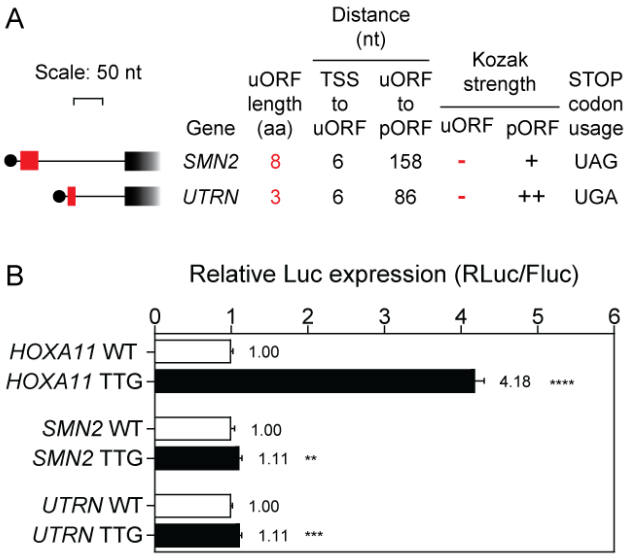


Figure 5.5 Validation of predicted uORFs in *SMN2* and *UTRN* genes with relevance to neuromuscular disease.

(**A**) Features of the 5′ UTRs of survivor motor neuron 2 (*SMN2*) and utrophin (*UTRN*) which were cloned upstream of RLuc in a dual-luciferase reporter (DLR) vector. uORFs are depicted by red boxes. uORF-containing (WT) and uORF-knockout (TTG) versions of each pDLR vector were transfected into HEK293T cells (n=4 wells) and (**B**) a DLR assay carried out to assess the effect of uORF activity on *Renilla* luciferase (RLuc) protein expression. The *HOXA11* WT and TTG uORF constructs were included as controls. Values were normalised to FLuc expression. Values of the WT and TTG version of each uORF are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Scale bar represents 50 nucleotides. Statistical significance was determined by a Student's t-test relative to the corresponding gene-specific WT control. ** = $P < 0.01$, *** = $P < 0.001$, and **** = $P < 0.0001$. TSS, transcription start site.

5.4.3 Analysis of uORF structure and function

Subsequent experiments were designed to investigate the various properties of uORFs in relation to their translation repressive effects using a variety of functional mutants. The *HOXA11* gene was selected for these studies as the resident uORF conferred a strong translational repressive effect (~4-fold) in DLR studies, and both the uORF and the 5′ UTR were of convenient lengths for facile experimental manipulation (**Figure 5.4A**). The pDLR *HOXA11* 5′ UTR vector was modified to replace the uORF sequence with a Golden Gate cloning site (**Figure 5.6**), enabling subsequent seamless insertion of various candidate uORF mutants upstream of the RLuc reporter gene in the context of the 5′ UTR of *HOXA11* (**Figure 5.7**).

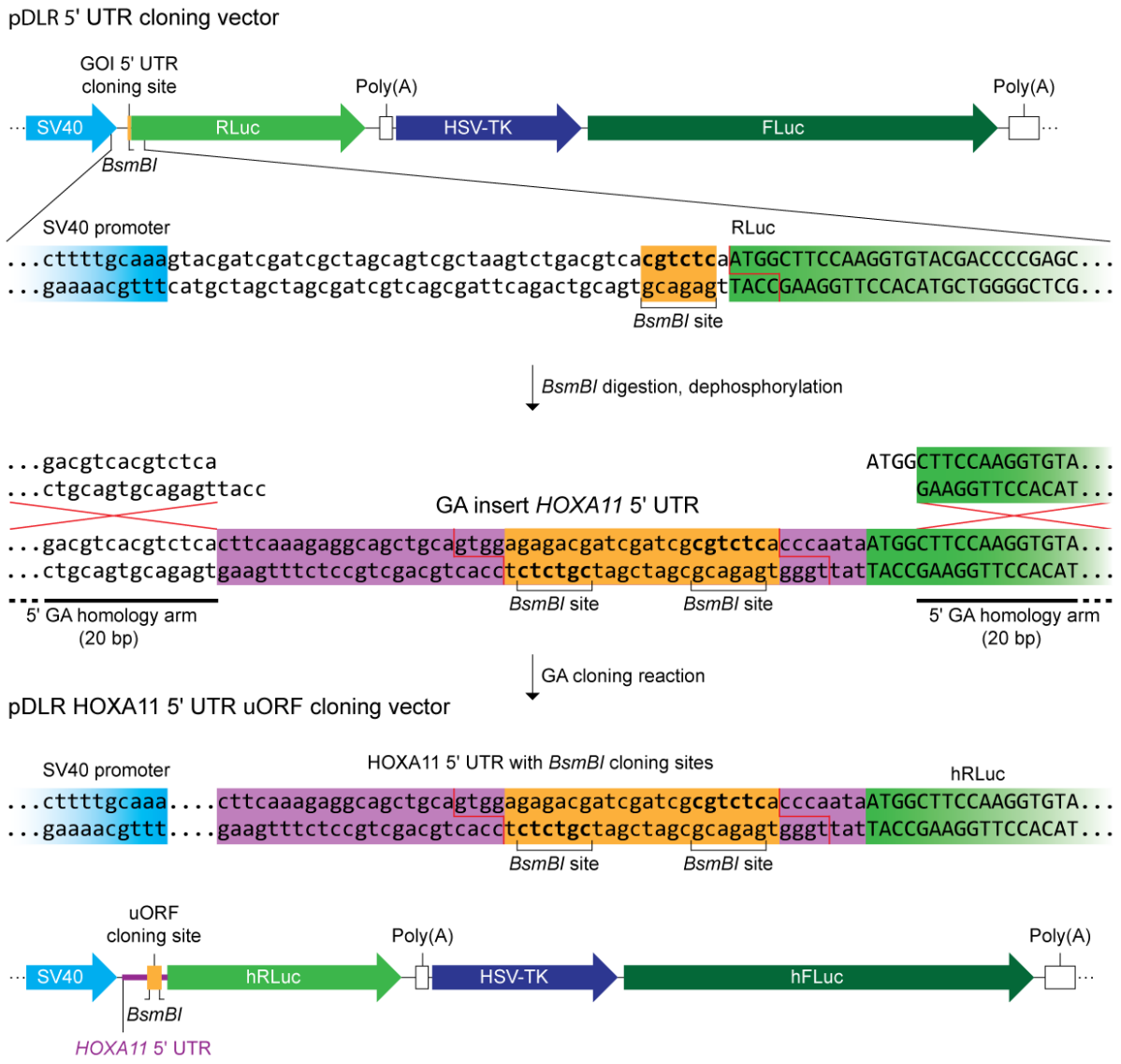


Figure 5.6 Generation of *HOXA11* uORF cloning site in the pDLR vector.

A uORF cloning site within the pDLR *HOXA11* 5′ UTR vector was introduced to enable insertion of various uORF mutants to assess the effect of various properties on uORF functionality. The pDLR 5′ UTR cloning vector was digested with *BsmBI*, and a GA insert containing the *HOXA11* 5′ UTR sequence where the uORF was replaced with a Golden Gate cloning site (two *BsmBI* recognition sequences for the seamless insertion of a mutant uORF sequence, including the Kozak context sequence and termination codon) was introduced. Coding and non-coding sequences are indicated in upper- and lowercase respectively.

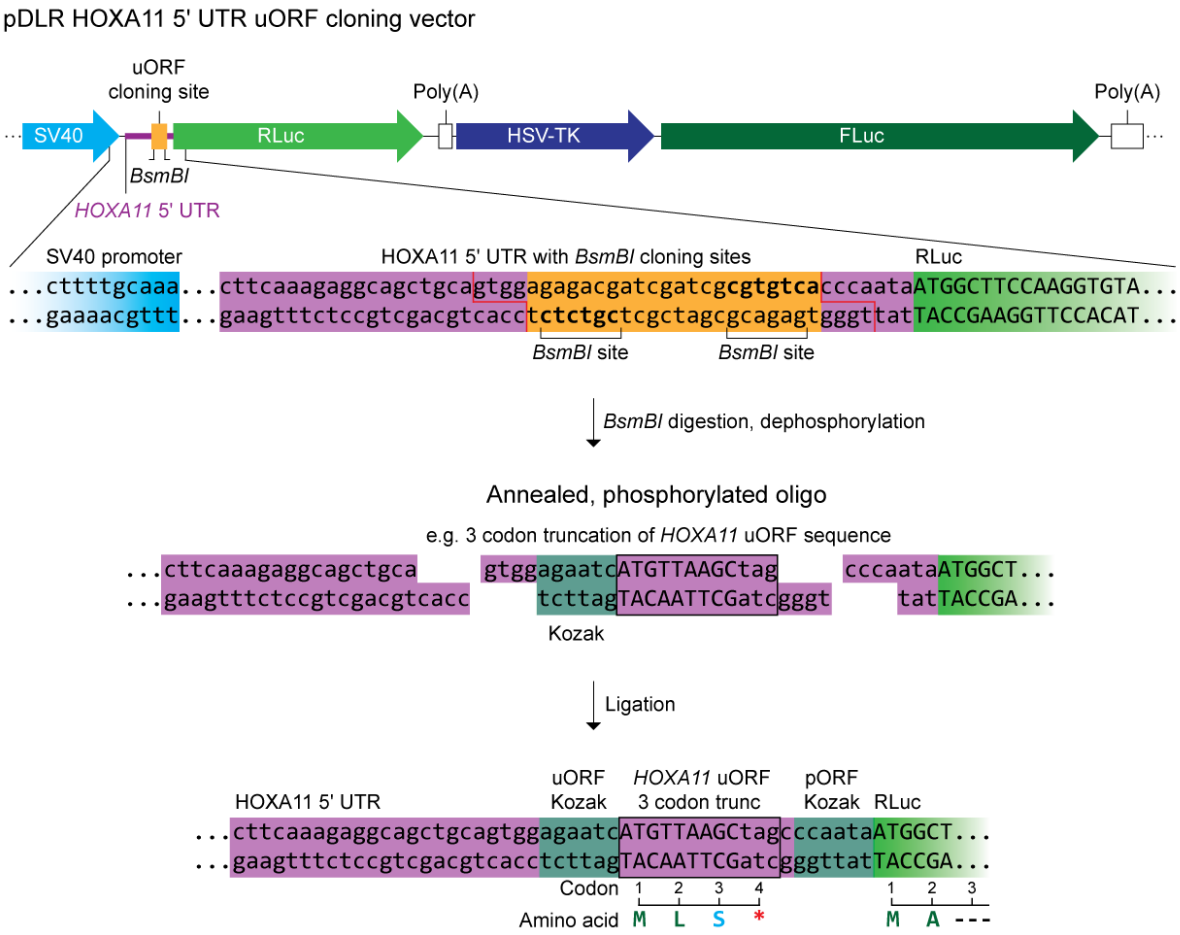


Figure 5.7 Generation of *HOXA11* mutant uORF pDLR vectors.

Mutants of the *HOXA11* uORF representative of a property of interest (i.e. length, Kozak consensus strength, etc.) were inserted into the pDLR *HOXA11* 5′ UTR cloning vector to assess the resulting effect on RLuc expression. The pDLR *HOXA11* 5′ UTR cloning vector was digested with *BsmBI* to create 5′ overhangs compatible with those of annealed oligos harbouring the desired uORF sequence. Insertion of a truncated *HOXA11* functional mutant (3 amino acid-encoding codons in length) is included as an example here. Coding and non-coding sequences are indicated in upper- and lowercase respectively.

5.4.3.1 The effect of uORF length and Kozak consensus on uORF activity

The probability of AUG codon recognition and initiation of translation by the scanning ribosome is dependent in part on the surrounding sequence context (i.e. the Kozak consensus) [318]. The optimal Kozak consensus sequence has been identified as GCCRCCAUGG (where R = A or G, and AUG is the start codon) [318]. The strength of the *HOXA11* uORF Kozak sequence is intermediate (i.e. an A is present in the -3 position, and a U at +4), as this fulfils only one of the two conditions required to be classified as strong [317,318]. Altering the Kozak context sequence upstream of the uORF start codon (i.e. to the optimal consensus [318]) resulted in enhancement of translational repression by 22%, although this did not reach statistical significance at the $P < 0.05$ level (**Figure 5.8A**). It should be noted that both the original and 'enhanced' Kozak sequence would be classified as intermediate in strength as only the preceding 6 DNA nucleotides to the uATG were optimised, whereas the +4 position was within the uORF coding sequence and was not altered. Furthermore, the -3 nucleotide was unchanged as this was already a purine in the WT sequence.

The importance of peptide length was investigated through a series of C-terminal truncations (i.e. to generate uORFs that were 9, 7, 5, 3, 2, and 1 amino acids in length). Progressive truncation of the *HOXA11* uORF resulted in a partial loss of repressive activity that was proportional to the degree of truncation (**Figure 5.8A**). The minimal uORF (i.e. uATG-STOP, M*, methionine-STOP) did not fit this pattern, and instead exhibited repression that was not statistically different from the *HOXA11* WT uORF. Schulz and colleagues demonstrated that the uORF within the 5′ UTR of *MAP2K6* (a proto-oncogene) which is similarly 'minimal' was also functional in a luciferase reporter assay [384]. An unoccupied ribosome E site has been shown to result in translational arrest and consequently dissociation of the ribosomal complex [583]. It is likely that the methionine-Stop combination of codons has a similar arresting effect, and therefore results in translational repression of the pORF by

promoting ribsosome recycling. These data are suggestive of at least two mechanisms of translational repression within the *HOXA11* 5′ UTR sequence context. Notably, such minimal uORFs occur frequently in in our predictions; 5,447 (6% of all uORFs) in the human genome, and 2,759 (6.5%) in mouse.

Increasing the length of the *HOXA11* WT uORF through the generation of a tandem repeat of the WT sequence enhanced the repressive effect by 42% ($P < 0.01$), and removing the internal ATG (and the potential for initiation at a novel overlapping uORF) exhibited a comparable repressive effect (**Figure 5.8A**). The addition of tag sequences (i.e. FLAG and HiBiT), however, did not significantly alter the repressive effect of the WT uORF despite these extending the uORF to a similar overall length as the tandem repeats mutants.

These data indicate that the effect of uORF length on activity is fairly complex. It is evident that uORF length plays a role to some extent as the repressive effect of the *HOXA11* uORF was progressively lost through C-terminal truncations, while the tandem repeat mutant exhibited enhanced translational repression. It is likely that other factors, such as the uORF amino acid identity or nucleic acid sequence, are also important for this activity. Importantly, the effects observed could not be explained by changes in mRNA levels, as demonstrated by qPCR data, suggesting that uORF-mediated repression manifests at the translational level (**Figure 5.8B**).

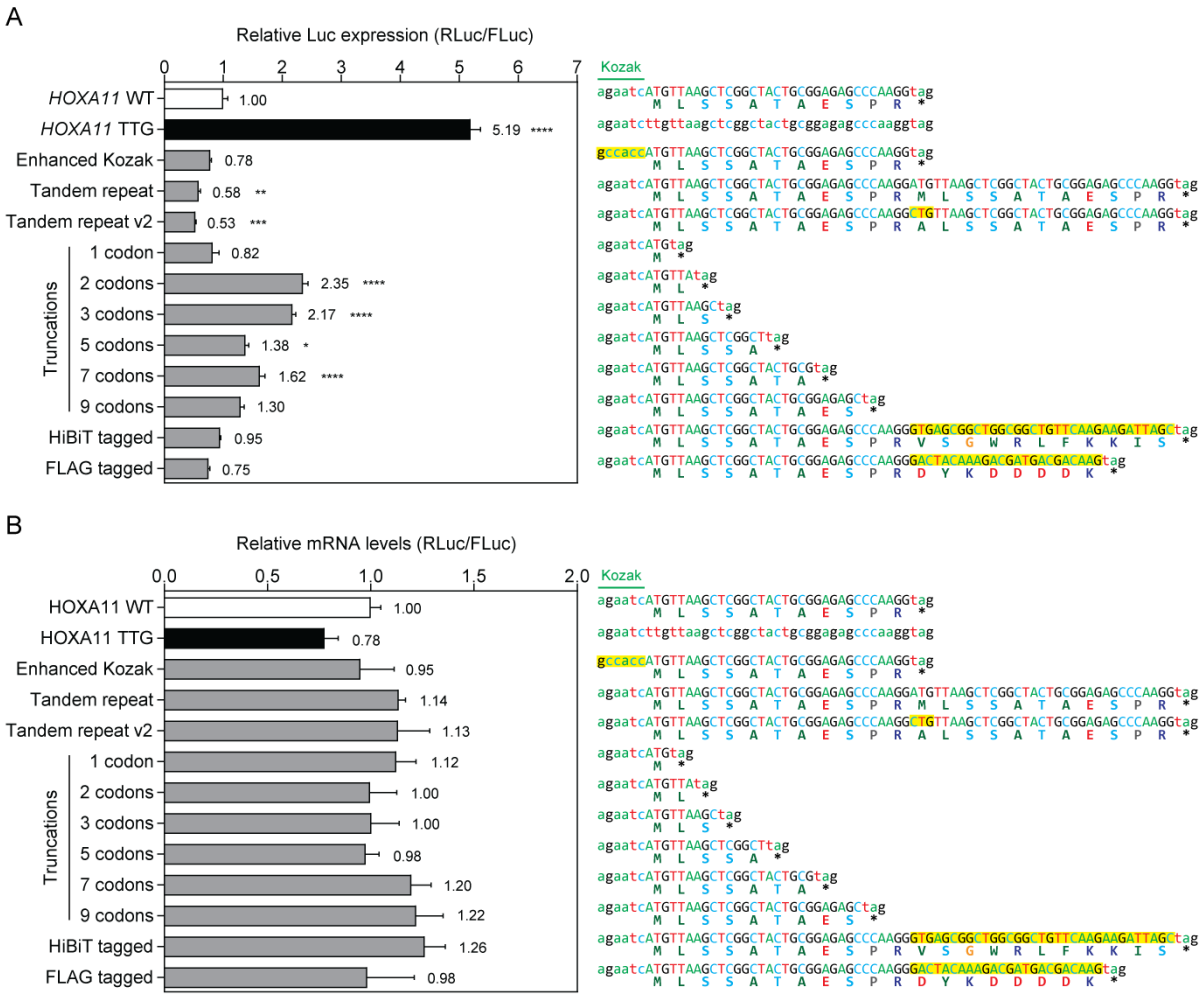

A
Relative Luc expression (RLuc/FLuc)
0 1 2 3 4 5 6 7
HOXA11 WT 1.00
HOXA11 TTG 5.19 ****
Enhanced Kozak 0.78
Tandem repeat 0.58 **
Tandem repeat v2 0.53 ***
Truncations
1 codon 0.82
2 codons 2.35 ****
3 codons 2.17 ****
5 codons 1.38 *
7 codons 1.62 ****
9 codons 1.30
HiBiT tagged 0.95
FLAG tagged 0.75
Kozak
B
Relative mRNA levels (RLuc/FLuc)
0.0 0.5 1.0 1.5 2.0
HOXA11 WT 1.00
HOXA11 TTG 0.78
Enhanced Kozak 0.95
Tandem repeat 1.14
Tandem repeat v2 1.13
Truncations
1 codon 1.12
2 codons 1.00
3 codons 1.00
5 codons 0.98
7 codons 1.20
9 codons 1.22
HiBiT tagged 1.26
FLAG tagged 0.98
Kozak

Figure 5.8 Effect of uORF length and Kozak sequence context on uORF activity.

The *HOXA11* 5′ UTR luciferase reporter was mutated to enhance the Kozak consensus sequence, alter the length of the uORF, or to introduce a FLAG or HiBiT tag to the C-terminus of the uORF micropeptide. HEK293T cells were transiently transfected with these mutant constructs and luciferase activity measured 24 hrs later. The *HOXA11* TTG construct (i.e. uORF knockout) was included as a control, exhibiting the maximal level of alleviation of translational repression. The corresponding uORF DNA and predicted micropeptide sequences for each uORF mutant are depicted to the right of the graphs where non-coding sequences are indicated in lowercase, and coding sequences in uppercase letters. DNA sequences highlighted in yellow are those that differ from the WT uORF. (**A**) Dual-luciferase assay measuring RLuc expression, normalised to FLuc. (**B**) RT-qPCR of RLuc mRNA transcripts, normalised to FLuc. Values are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the *HOXA11* WT control. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$ and **** = $P < 0.0001$.

Oyama and colleagues reported detection of four novel short peptides between 59 and 86 amino acids in length produced from uORFs by mass spectrometry carried out in K562 cells [584]. Notably, these micropeptides are longer than the majority of uORFs in our predicted uORF database (**Figure 5.1**). This highlights the difficulty of detecting a useful range of uORF-expressed micropeptide sizes using currently available high-throughput mass spectrometry methods. In order to determine whether the translated product of short uORFs are directly detectable, a HiBiT-tagged *HOXA11* uORF was expressed from the 5′ UTR of the DLR vector in transiently transfected cells. The HiBiT detection assay utilises nanoluciferase biocomplementation, whereby the HiBiT fragment that is covalently attached to the micropeptide interacts with a LgBiT protein in the presence of the Nano-Glo substrate, creating a bioluminescent Nano-BiT fusion protein (**Figure 5.9A**). The detection of *HOXA11*-uORF micropeptides was increased by ~12-fold in the presence of the 26S proteasome inhibitor MG132 [585,586], suggesting that these peptides are translated, but are expressed at low levels and/or are rapidly degraded (**Figure 5.9B**).

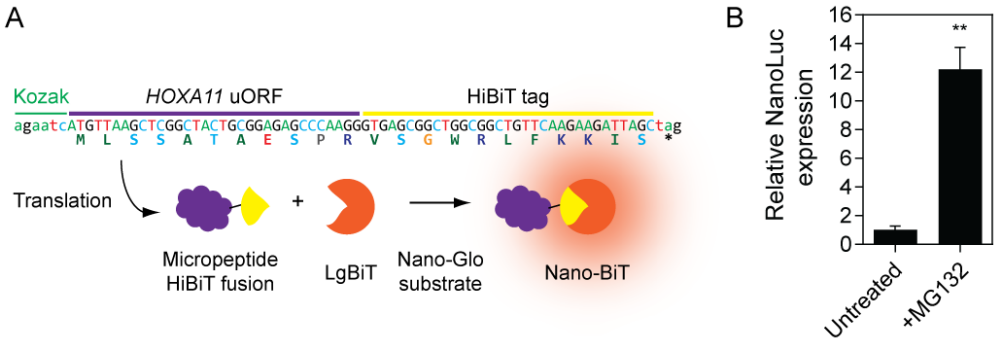


Figure 5.9 Detection of HiBiT-tagged *HOXA11*-uORF micropeptides.

(**A**) The HiBiT lytic detection assay measures levels of the luminescent Nano-BiT enzyme, constituted by the spontaneous interaction between a complementing polypeptide LgBiT and the HiBiT tag. (**B**) HEK293T cells were transiently transfected with the pDLR *HOXA11* HiBiT-tagged uORF vector, and were treated briefly prior to Nano-BiT detection with MG132 (a small molecule protease inhibitor) to facilitate accumulation of the tagged uORF micropeptide in cells. Cells transfected with the pDLR *HOXA11* HiBiT-tagged uORF vector but not treated with MG132 were included as a control. Values are means ± SEM, and were scaled such that the mean of the untreated control group was returned to a value of 1. Statistical significance was determined by a Student's *t* test. ** = $P < 0.01$.

5.4.3.2 The effect of uORF amino acid composition on uORF activity

It is well known that missense mutations within critical domains of annotated proteins can be deleterious to functional activity. Accordingly, a degree of uORF amino acid conservation has been reported by Ji and colleagues in a bioinformatics analysis study [558], and the codons within longer uORFs have been found to exhibit higher levels of evolutionary conservation when compared to 5′ UTR sequences in general [415]. Some examples of uORFs that function with sequence-dependence include those of the receptor-like protein-tyrosine phosphatase J (*PTPRJ*) [587], and S-adenosylmethionine decarboxylase (*AMD1*) [588] genes. To assess the relative contribution of individual amino acids on *HOXA11*-uORF activity, alanine-scanning mutagenesis was carried out. This technique is typically employed for systematic protein function mapping as alanine is a non-bulky amino acid with a chemically inert methyl side chain [589–591]. Alanine-scanning thus provides an insight into the contribution of individual amino acid side chains on the structure and function of the peptide under investigation. The *HOXA11*-uORF micropeptide sequence consists of 11 amino acids (MLSSATAESPR*) for which a series of mutant constructs were generated where each single non-alanine residue was mutated to alanine in turn. In general, the effect of substituting each amino acid with alanine was minimal, with the majority of mutants exhibiting no significant difference in translational repression relative to the WT control (**Figure 5.10A**). However, replacing the leucine at position 2, and serine at position 9, resulted in significant enhancements to translational repression of 44% ($P < 0.0001$) and 23% ($P < 0.01$) respectively, while converting the terminal arginine residue to alanine relieved repression by 44% ($P < 0.0001$). These changes in uORF activity could not be explained by alterations in mRNA transcript levels (**Figure 5.10B**).

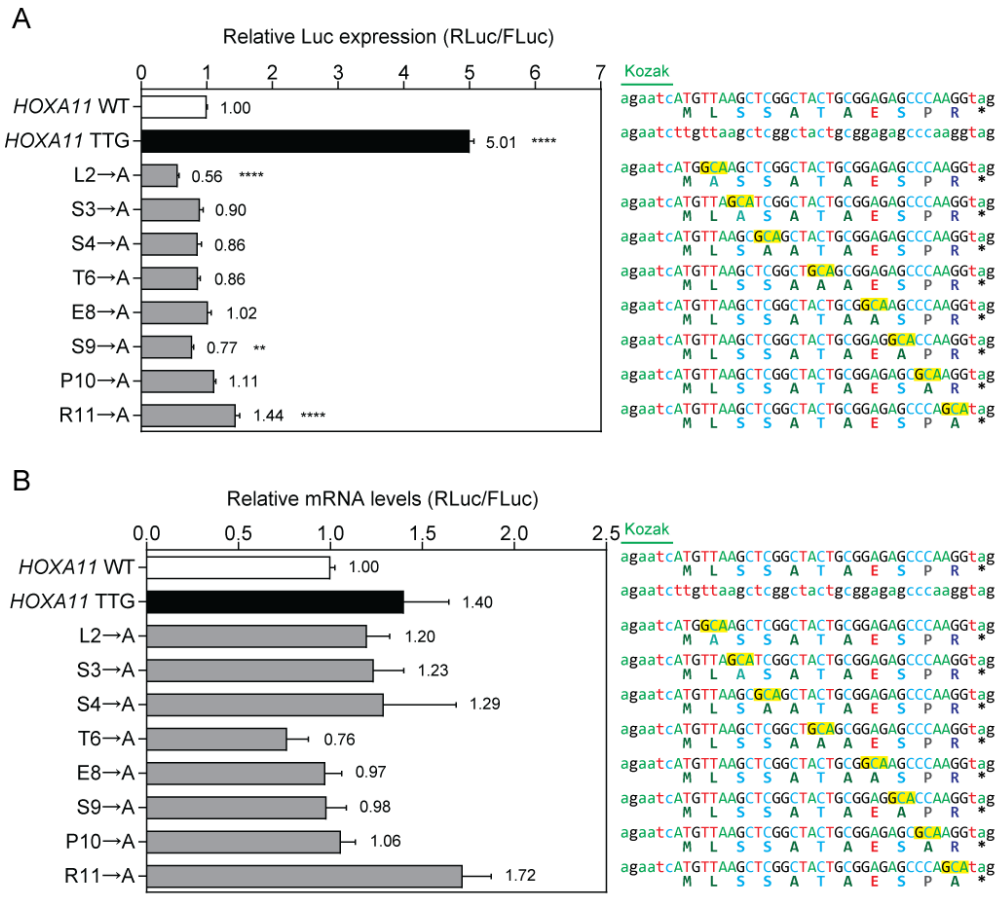


Figure 5.10 The effect of single amino acid changes on *HOXA11* uORF activity.

HEK293T cells were transfected with pDLR *HOXA11 5′* UTR mutant vectors encoding single alanine substitutions at each position along the uORF-encoded micropeptide sequence. The alanine-specific codon, GCA, is highlighted in yellow where different from the WT sequence. The start codon (methionine) and the two original alanine codons were not altered. (**A**) Dual-luciferase assay measuring RLuc expression, normalised to FLuc. (**B**) RT-qPCR of RLuc mRNA transcripts, normalised to FLuc. Values are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the *HOXA11* WT control. ** = $P < 0.01$, and **** = $P < 0.0001$.

When the amino acid sequence of the *HOXA11* uORF was scrambled such that the composition remained the same, but the order was randomised, there was no significant difference in uORF-mediated translational repression relative to the WT control (**Figure 5.11A**). This finding is in accordance with the minimal disruption to uORF activity observed with the majority of the sequential, single alanine substitutions. Taken together, these findings suggest that the effect of order and single positions of amino acid residues within the *HOXA11* uORF micropeptide on uORF activity is minimal, while there may be some exceptions.

To assess the effect of uORF biophysical properties on translational repression, synthetic uORF mutants consisting of homopolymeric amino acid sequences of 10 acidic (i.e. glutamic acid, aspartic acid), basic (i.e. arginine, lysine), aliphatic hydrophobic (i.e. methionine, alanine, leucine), aromatic hydrophobic (i.e. phenylalanine, tryptophan), and 'other' (i.e. proline) residues were created. Notably, rare codons were avoided in the design of the uORF mutants described in this thesis, unless otherwise stated. Altering the biophysical properties of the *HOXA11* uORF had profound effects on luciferase reporter expression. Synthetic homopolymeric uORFs consisting of the acidic residues (Poly-E and Poly-D) exhibited the greatest effects with ~90% increase in translational repression relative to the *HOXA11* WT control (**Figure 5.11A**). Similarly, the basic homopolymeric chain mutants (Poly-R and Poly-K) resulted in 76% and 59% reduction in expression of the luciferase reporter respectively (both $P < 0.001$).

The Poly-M and Poly-A aliphatic hydrophobic mutants moderately increased translational repression of the luciferase reporter by 43% and 25% respectively, although this was not statistically different from the WT and scrambled controls (**Figure 5.11A**). In contrast, the other hydrophobic residues (Poly-L, Poly-F, and Poly-Y) all exhibited reduced translational repression, with the effects being strongest for the aromatic hydrophobic amino acids (Poly-F, and Poly-Y). The Poly-P construct behaved similarly to the acidic and basic chains, enhancing

repression by 77% ($P < 0.001$) relative to the WT control. The greatly enhanced repressive activity of the Poly-R mutant is consistent with the results from the alanine-scan where the R11>A construct partially relieved the *HOXA11* uORF-mediated translation repression (**Figure 5.10A**). Similarly, the increased repressive activity of the L2>A mutant is consistent with the weaker repressive effect of the Poly-L mutant observed in this experiment. The effects observed could not be explained by changes in mRNA levels (**Figure 5.11B**). In fact, moderate increases in transcript levels were observed for the Poly-E, -D, -R, -K, -A, and -P mutants, which is the opposite direction of effect observed at the protein level.

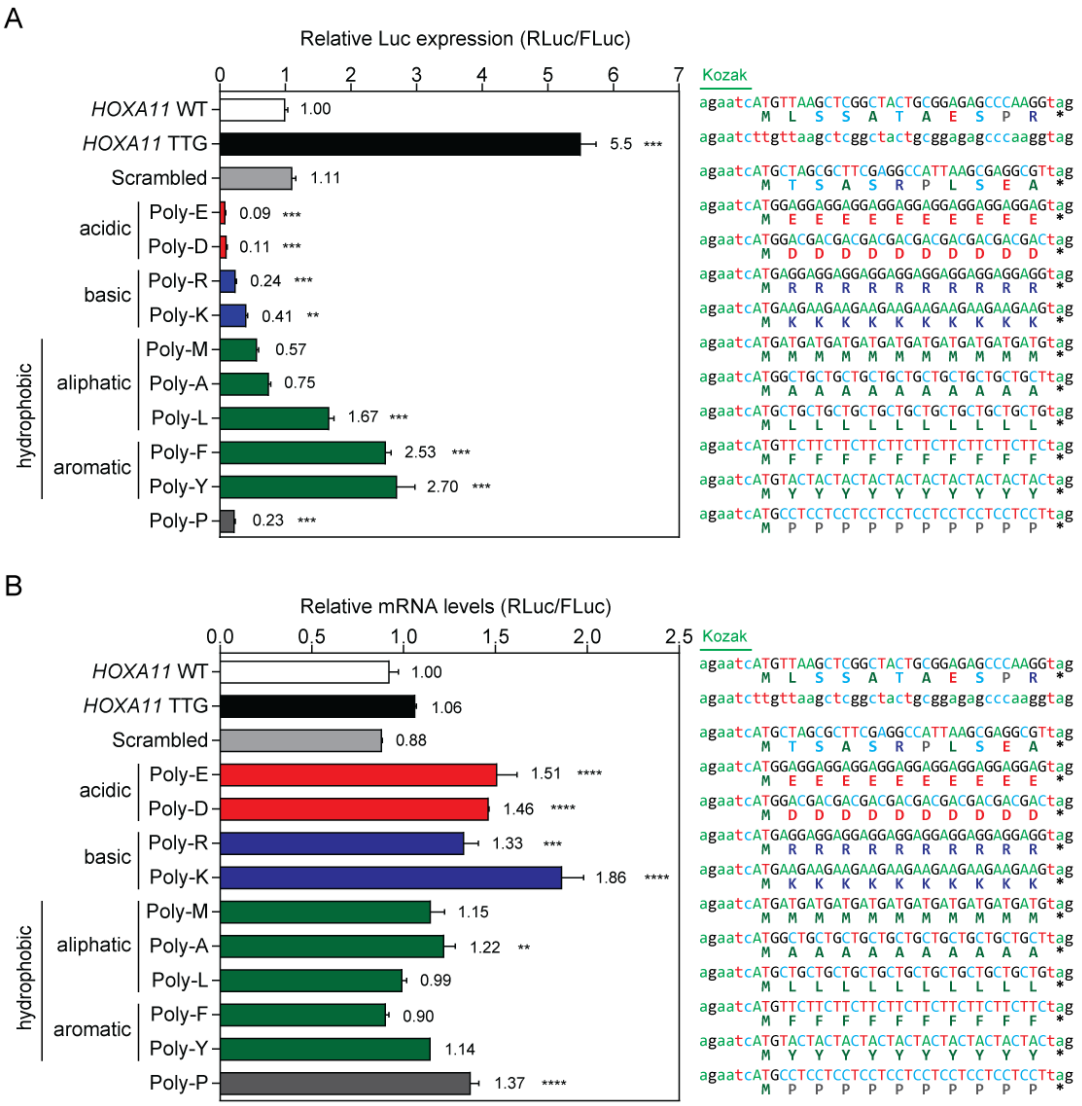


Figure 5.11 The effect of homopolymeric amino acid chains on uORF activity.

HEK293T cells were transfected with pDLR *HOXA11 5′* UTR mutant vectors encoding methionine followed by 10-mer homopolymeric chains of select amino acids, including acidic (E, and D), basic (R, and K), aliphatic hydrophobic (M, A, and L), aromatic hydrophobic (F, and Y), and 'other' (P) residues. A codon-randomised (scrambled) mutant was included as a control. (**A**) Dual-luciferase assay measuring RLuc expression, normalised to FLuc. (**B**) RT-qPCR of RLuc mRNA transcripts, normalised to FLuc. Values are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the *HOXA11* WT control. ** = $P < 0.01$, *** = $P < 0.001$, and **** = $P < 0.0001$.

A series of mutants with truncations to the homopolymeric amino acid chain for Poly-E, Poly-R and Poly-F (i.e. mutants with 1, 2, 4, 6, 8, and 10 amino acid residues) were created to determine if the number of amino acid residues within the uORF micropeptide was proportional to the level of uORF-mediated translational repression. The enhanced repressive effect of the Poly-E uORF was progressively diminished as the uORF length decreased (**Figure 5.12A**), and this could not be attributed to changes in mRNA transcript levels (**Figure 5.12B**). A similar effect was observed for the Poly-R uORF, although the repressive effect plateaued for the 8- and 10-residue mutants (i.e. R_8 and R_{10}) (**Figure 5.12C**). Conversely, the effect of the uORF length of the poly-hydrophobic chain mutant (i.e. Poly-F) was less clear than that of the acid and basic versions (**Figure 5.12D**). The activity of the F_1, F_2, and F_4 uORFs were more similar to the WT control, while the F_6, F_8, and F_{10} mutants exhibited similarly high levels of alleviation of translation repression relative to the full-length version.

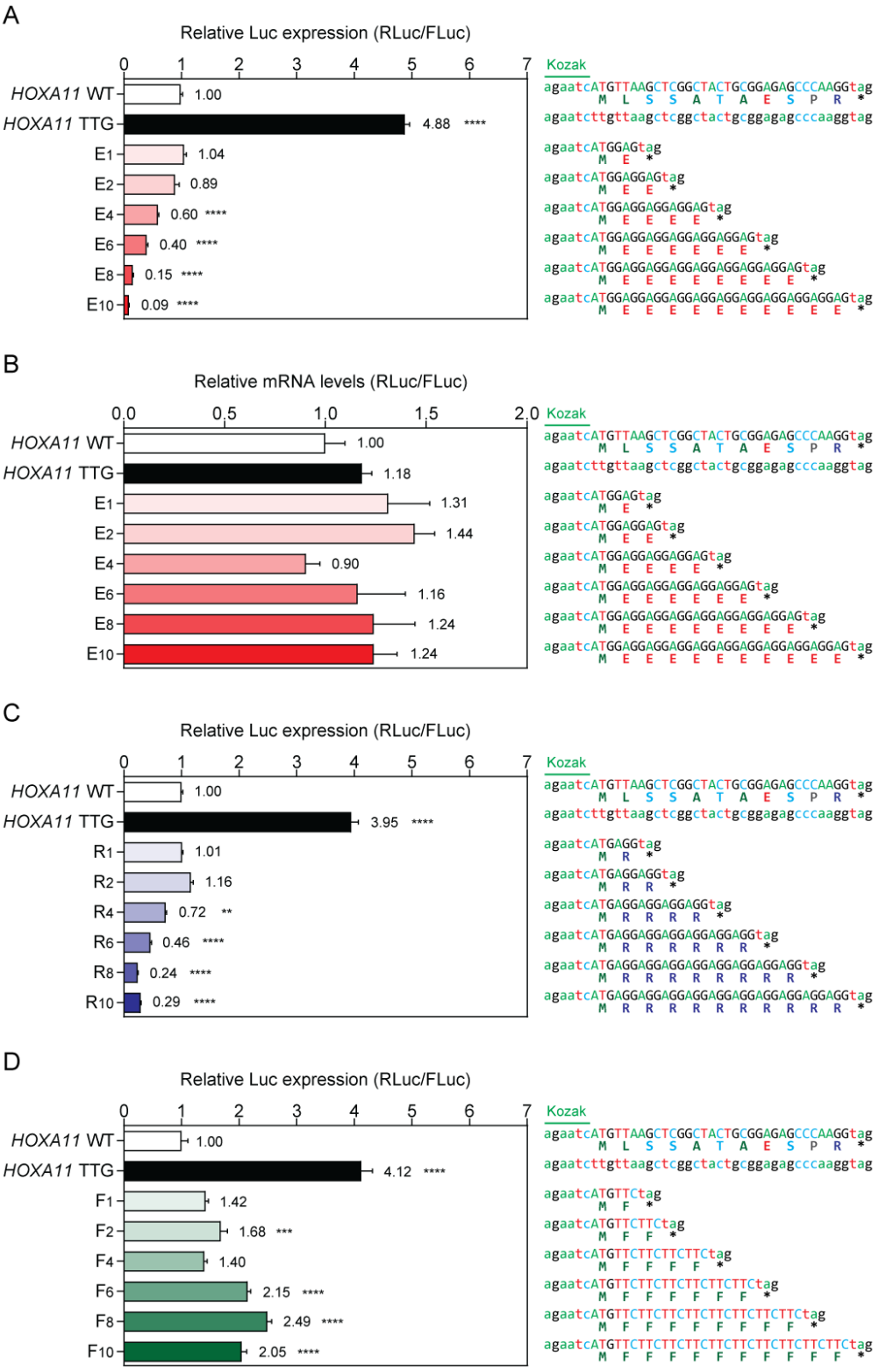

A
Relative Luc expression (RLuc/FLuc)
0 1 2 3 4 5 6 7
HOXA11 WT 1.00
HOXA11 TTG 4.88 ****
E1 1.04
E2 0.89
E4 0.60 ****
E6 0.40 ****
E8 0.15 ****
E10 0.09 ****
Kozak
agaatcATGTTAAGCTCGGCTACTGCGGAGAGCCCAAGGtag
M L S S A T A E S P R *
agaatcttgttaagctcggctactgcggagagcccaaggtag
agaatcATGGAGtag
M E *
agaatcATGGAGGAGtag
M E E *
agaatcATGGAGGAGGAGGAGtag
M E E E E *
agaatcATGGAGGAGGAGGAGGAGGAGtag
M E E E E E E *
agaatcATGGAGGAGGAGGAGGAGGAGGAGGAGtag
M E E E E E E E E *
agaatcATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGtag
M E E E E E E E E E E *
B
Relative mRNA levels (RLuc/FLuc)
0.0 0.5 1.0 1.5 2.0
HOXA11 WT 1.00
HOXA11 TTG 1.18
E1 1.31
E2 1.44
E4 0.90
E6 1.16
E8 1.24
E10 1.24
Kozak
C
Relative Luc expression (RLuc/FLuc)
0 1 2 3 4 5 6 7
HOXA11 WT 1.00
HOXA11 TTG 3.95 ****
R1 1.01
R2 1.16
R4 0.72 **
R6 0.46 ****
R8 0.24 ****
R10 0.29 ****
Kozak
agaatcATGAGGtag
M R *
agaatcATGAGGAGGtag
M R R *
agaatcATGAGGAGGAGGAGGtag
M R R R R *
agaatcATGAGGAGGAGGAGGAGGAGGtag
M R R R R R R *
agaatcATGAGGAGGAGGAGGAGGAGGAGGAGGtag
M R R R R R R R R *
agaatcATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGtag
M R R R R R R R R R R *
D
Relative Luc expression (RLuc/FLuc)
0 1 2 3 4 5 6 7
HOXA11 WT 1.00
HOXA11 TTG 4.12 ****
F1 1.42
F2 1.68 ***
F4 1.40
F6 2.15 ****
F8 2.49 ****
F10 2.05 ****
Kozak
agaatcATGTTCtag
M F *
agaatcATGTTCTTCtag
M F F *
agaatcATGTTCTTCTTCTTCtag
M F F F F *
agaatcATGTTCTTCTTCTTCTTCTTCtag
M F F F F F F *
agaatcATGTTCTTCTTCTTCTTCTTCTTCTTCtag
M F F F F F F F F *
agaatcATGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCtag
M F F F F F F F F F F *

Figure 5.12 The effect of truncating homopolymeric amino acid chains on uORF activity. HEK293T cells were transfected with pDLR *HOXA11* 5′ UTR mutant vectors encoding methionine followed by truncations of the acidic (Poly-E), basic (Poly-R), and hydrophobic (Poly-F) homopolymeric chain mutants (i.e. with 1, 2, 4, 6, 8, and 10 amino acid residues). Dual-luciferase assay measuring RLuc expression, normalised to FLuc. (**B**) RT-qPCR of RLuc mRNA transcripts for the Poly-E truncation mutants, normalised to FLuc. Values are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the *HOXA11* WT control. *** = $P < 0.001$, **** = $P < 0.0001$.

Several studies have demonstrated that there is an increased probability for the 40S ribosomal subunit (and associated initiation factors) to remain bound to the mRNA and reinitiate translation at downstream ORFs where the preceding ORF is short [376,376,567,592]. This, however, cannot entirely account for the results observed here where decreasing the length of the poly-acidic or -basic uORF chain decreased the repressive effect of the full-length mutant, but only to WT levels of repression (and not those exhibited by the TTG uORF knockout mutant). Furthermore, the opposite direction of effect on reporter translation was observed with the hydrophobic chain truncations where the shorter uORF mutants (i.e. F_1-F_4) were more repressive than the longer mutants (i.e. F_6-F_{10}). These data strongly suggest that the biophysical properties of particular amino acids have the potential to influence uORF activity, and increasing the number of such residues within the chain enhances the effect accordingly. In addition to uORF length, the speed of uORF translation has also been shown to influence the probability of reinitiation at the pORF start [592]. A possible mechanism could be that the biophysical properties of the uORF homopolymeric chains promote, or reduce, charge-charge interactions between the amino acid side chains and the wall of the nascent peptide tunnel and/or water molecules therein. Stronger interactions, such as with the highly charged poly-acidic and poly-basic chains, may slow ribosomal progression, while weaker interactions, such as with the hydrophobic amino acid residues, may enhance the speed of uORF translation, thereby regulating the possibility for reinitiation of translation at the pORF start.

A panel of further synthetic uORFs with more complex combinations of uORF peptide sequences were created and tested in order to further investigate the influence of amino acid composition on poly-acidic uORF activity. Increasing the length of the E_{10} mutant to 20 residues (i.e. E_{20}) increased the repressive activity from 93% to 97% ($P < 0.0001$) relative to the WT control (**Figure 5.13**). Enhancing the Kozak consensus of the poly-acidic uORF

maintained the same highly repressive effect, consistent with the earlier finding that altering the WT *HOXA11* Kozak sequence context played a minor role in uORF activity (**Figure 5.8**).

The remaining amino acid composition uORF mutant designs were based on the E_6 construct as this resulted in intermediate levels of repression (**Figure 5.8**), allowing for a larger dynamic range for assessing changes in luciferase expression. Interspersing alanine residues with doublets of glutamic acid residues (i.e. $(AE_2)_3A$) resulted in a similar level of repression to the E_6 mutant. Conversely, alternating acidic and hydrophobic amino acid residues (i.e. $(EF)_5$) resulted in activity levels similar to the *HOXA11* WT uORF, suggesting that the opposing effects of acidic and hydrophobic residues compensated for one another. Similarly, alternating acidic and basic residues (i.e. $(ER)_5$) resulted in a balancing of the individual repressive effects. As such, the effect of uORF amino acid composition on uORF activity is more complex than purely being additive or subtractive. All of the completed uORF peptides assessed in the present study are expected to occupy a space entirely within the ribosomal tunnel which can accommodate up to 40 amino acid residues of an elongating nascent peptide chain [593]. Interactions between the side chains of the neighbouring amino acids that modulate the biophysical properties of the uORF peptide (e.g. net charge, hydrophobicity) may contribute towards interactions with the ribosomal tunnel, thereby influencing the processivity of the ribosome along the transcript.

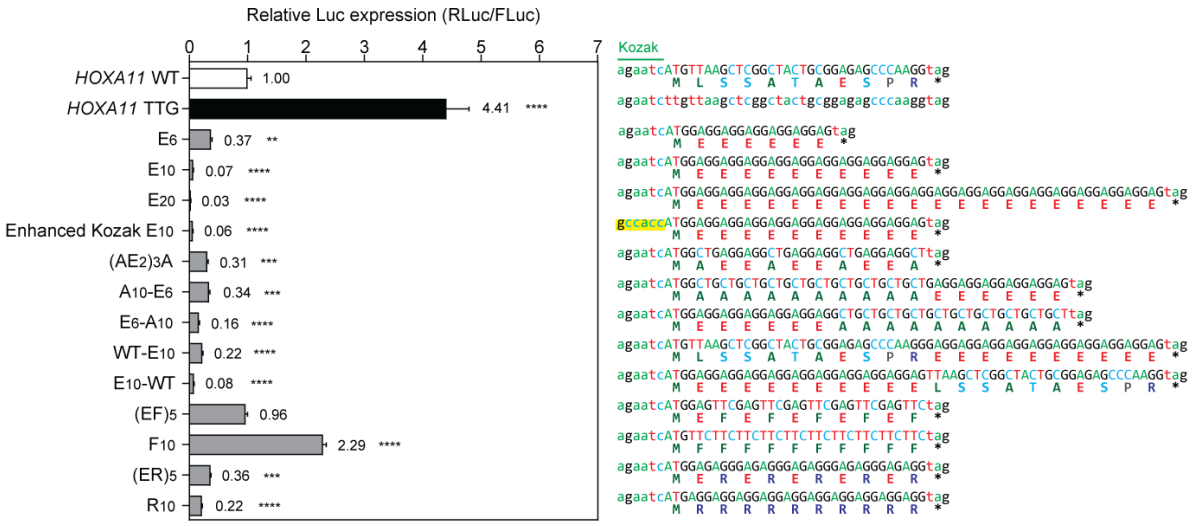


Figure 5.13 The effect of complex *HOXA11* uORF functional mutants on activity.

HEK293T cells were transfected with pDLR *HOXA11 5′* UTR mutant vectors encoding various advanced functional mutants in the *HOXA11 5′* UTR background. A dual-luciferase assay was carried out measuring RLuc expression, normalised to FLuc. Values are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the *HOXA11* WT control. ** = $P < 0.01$, *** = $P < 0.001$, and **** = $P < 0.0001$.

In order to determine whether the translational repressive effect of the poly-acidic amino acid residues is dependent on the order in which they are translated, the E_6 tract was placed either at the N- or the C-terminus of the nascent chain, and ten alanine residues at the corresponding C- or N-terminus respectively. Translation of the E_6 tract immediately following initiation (i.e. with the E_6-A_{10} mutant) resulted in a higher level of translational repression relative to the reverse order mutant (i.e. A_{10}-E_6). Similarly, where ten glutamic acid residues were placed ahead of the *HOXA11* WT uORF sequence, the level of translational repression was enhanced relative to the reverse order of this mutant (i.e. WT-E_{10}).

These data demonstrate that the number of acidic residues contributes strongly towards synthetic uORF activity, and that the biophysical properties (specifically the charge) of the neighbouring amino acids in the nascent polypeptide chain play an important role in uORF activity. Furthermore, the amino acids positioned nearer the N-terminus of the elongating micropeptide exert a greater effect on uORF activity, at least for the poly-acidic tract. It should be noted that runs of homopolymeric residues are typically not observed for native human uORFs, and that these synthetic mutants are primarily useful for identifying the hierarchy of uORF features that contribute towards uORF activity.

5.4.3.3 The effect of uORF nucleic acid sequence and structure on uORF activity

A number of possible confounding factors relating to the uORF nucleic acid sequence and secondary structure could account for the enhanced repressive activity of the poly-acidic amino acid-containing uORFs. These include the repetitive nature, low sequence complexity, and high G/C content of the RNA sequences encoding the synthetic uORFs. To assess whether any of these factors influence uORF activity, a number of additional mutants were created. Sequence complexity of the highly repressive acidic homopolymeric uORF mutants was increased with two codon optimised versions, namely E_{10} v2 which included alternating GAG and GAA codons, and $(ED)_5$ with alternating glutamic acid and aspartic acid residues (i.e. the use of all four possible acidic amino acid codons and thus rare codons were used in this instance). The repressive effect of these additional mutants was not significantly different to the original E_{10} and D_{10} mutants (**Figure 5.14A**). As codon optimisation resulted in the introduction of two novel uATGs (and by association the possibility for additional ribosomal initiation sites) in the $(ED)_5$ mutant, a version lacking the uATGs (i.e. $(ED)_5$ v2) was also tested. This maintained the same high level of uORF-mediated translational repression ($P < 0.0001$) observed with the poly-E and poly-D mutants (**Figure 5.14B)**. Knocking out the $(ED)_5$ v2 uORF (i.e. $(ED)_5$ v2 TTG) restored luciferase reporter activity to levels similar to those of the *HOXA11* TTG mutant, demonstrating that the $(ED)_5$ uORF represses its downstream pORF by up to 99%. These results confirm that the repressive nature of the poly-acidic synthetic uORFs is not RNA-dependent.

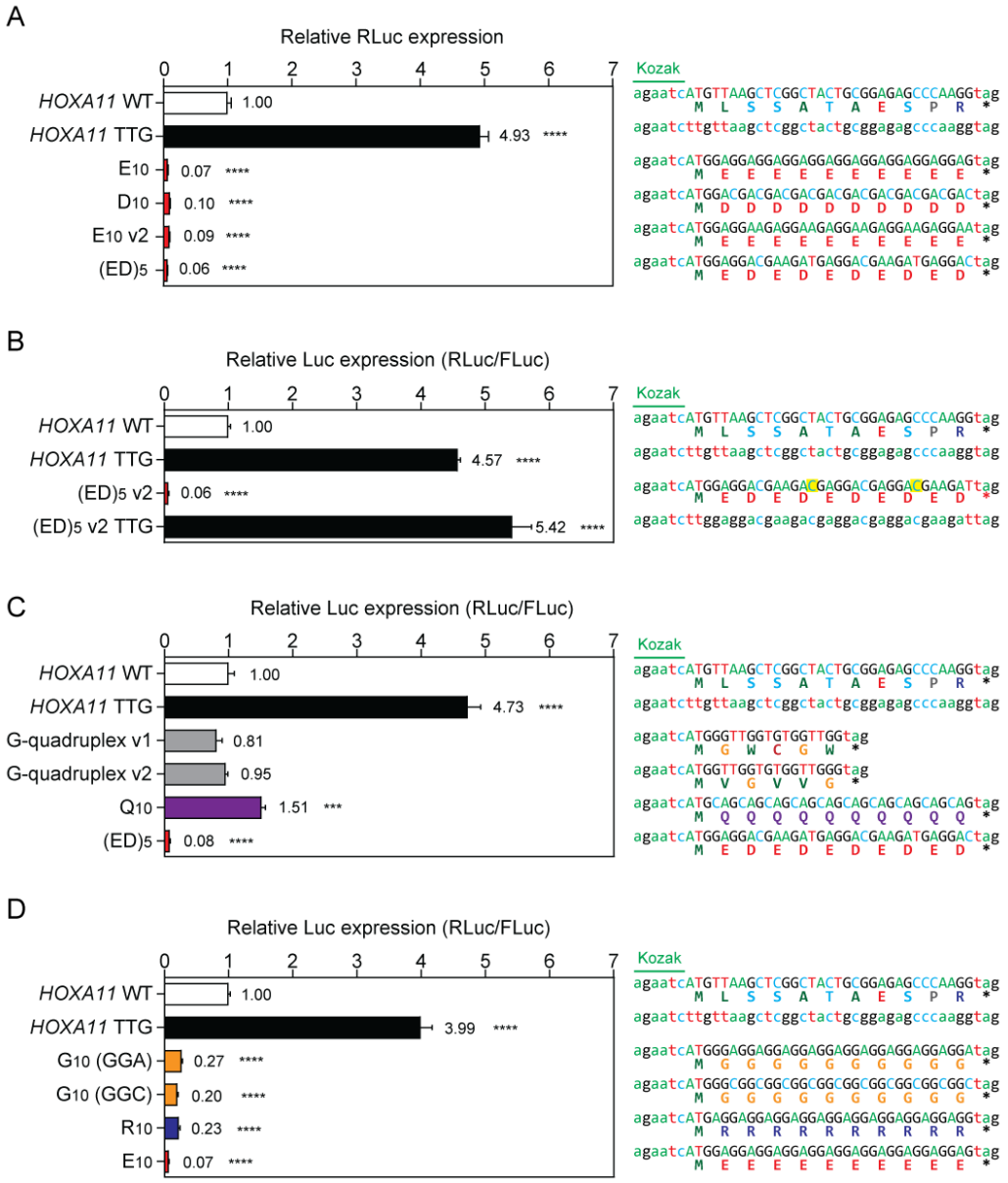

A
Relative RLuc expression
0 1 2 3 4 5 6 7
HOXA11 WT 1.00
HOXA11 TTG 4.93 ****
E10 0.07 ****
D10 0.10 ****
E10 v2 0.09 ****
(ED)5 0.06 ****
Kozak
B
Relative Luc expression (RLuc/FLuc)
0 1 2 3 4 5 6 7
HOXA11 WT 1.00
HOXA11 TTG 4.57 ****
(ED)5 v2 0.06 ****
(ED)5 v2 TTG 5.42 ****
Kozak
C
Relative Luc expression (RLuc/FLuc)
0 1 2 3 4 5 6 7
HOXA11 WT 1.00
HOXA11 TTG 4.73 ****
G-quadruplex v1 0.81
G-quadruplex v2 0.95
Q10 1.51 ***
(ED)5 0.08 ****
Kozak
D
Relative Luc expression (RLuc/FLuc)
0 1 2 3 4 5 6 7
HOXA11 WT 1.00
HOXA11 TTG 3.99 ****
G10 (GGA) 0.27 ****
G10 (GGC) 0.20 ****
R10 0.23 ****
E10 0.07 ****
Kozak

Figure 5.14 The effect of uORF nucleic acid sequence and structure on uORF activity. Codon optimised versions of the highly repressive acidic uORFs (i.e. E_{10} v2 and $(ED)_5$) were created to reduce sequence complexity. This was further reduced by removal of internal uATGs in the $(ED)_5$ construct (i.e. $(ED)_5$ v2). Additional *HOXA11* uORF mutants with complex secondary structure (i.e. G-quadruplex v1, and Q_{10}/CAG-repeat sequence) or high G/C content (i.e. G_{10} with GGA or GGC codon repeats). G-quadruplex v2 is frame-shifted by a single nucleotide relative to v1. (**A**-**D**) HEK293T cells were transfected with pDLR *HOXA11* 5′ UTR mutant vectors, and a dual-luciferase assay carried out measuring RLuc expression, normalised to FLuc. Values are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the *HOXA11* WT control. *** = $P < 0.001$, and **** = $P < 0.0001$.

The effect of nucleic acid structure on uORF activity was assessed with uORF-encoding G-quadruplexes in two reading frames (i.e. v1 and v2 which encode different uORF amino acid sequences, but are expected to generate the same RNA secondary structures). The G-quadruplex sequence is a thrombin-binding DNA aptamer for which the secondary structure has been well-characterised [594]. These mutants exhibited minor reductions in luciferase expression, which were not significantly different to the WT control (**Figure 5.14C**). Furthermore, a CAG-repeat uORF encoding poly-glutamine (i.e. Q_{10}) which is known to cause the formation of RNA hairpin structures in nature [595], had an opposite effect on luciferase expression, partially alleviating translational repression (**Figure 5.14C**). These data suggest that RNA secondary structures are unlikely to be responsible for the highly repressive activity of the synthetic poly-acidic uORFs.

To determine whether the high G/C content of the glutamic and aspartic acid uORF RNA sequences is a major determinant of their highly repressive activity (**Figure 5.11A**), two homopolymeric glycine (i.e. G_{10}) mutants, with GGA or GGC codon repeats, were created. The G_{10} (GGA) and G_{10} (GGC) constructs resulted in 73% and 80% reduction in reporter expression respectively, relative to the WT control (both $P < 0.0001$) (**Figure 5.14D**). This was comparable to the poly-basic uORF mutant (i.e. R_{10} which has AGG codon repeats), but was less repressive than the poly-acidic mutant (E_{10}) with GAG repeats and which diminished reporter expression by 93% ($P < 0.0001$). This suggests that G/C content alone cannot account for the high levels of repression observed with the acidic synthetic uORF sequence.

Taken together, these additional controls demonstrate that the RNA sequence and structure are far less important attributes of the uORF contributing towards activity, and instead the hyper activity of the poly-acidic amino acid chains is likely due to the specific biophysical properties of the amino acid residues within the uORF micropeptide sequence.

5.4.3.4 The effect of uORF position within the 5′ UTR on uORF activity

In order to assess the effect of uORF position within the 5′ UTR sequence on translational repressive activity, the uORF-naïve 5′ UTR of the *BRCA2* gene (a gene that was not identified in the uORF prediction database) was selected (**Figure 5.1**). The *HOXA11* WT, TTG, and hyper-functional $(ED)_5$ uORFs (all with the natural *HOXA11* Kozak context sequence) were transplanted into proximal, medial, and distal positions relative to the start of the luciferase reporter pORF in the pDLR vector, and the effect on luciferase reporter expression assessed (**Figure 5.15A**). *HOXA11* WT uORF repressive activity was strongest in the proximal position, intermediate in the medial position, and almost completely relieved at the distal position. The $(ED)_5$ mutant resulted in 98% and 97% repression in the proximal and medial positions respectively, and remained highly active in the distal position, resulting in 91% repression. No translation repressive effect was observed with the TTG mutant control. Furthermore, the translational repressive effects could not be explained by changes in reporter mRNA transcript levels (**Figure 5.15B**). Importantly, the small increase in luciferase transcript levels observed with the $(ED)_5$ proximal and distal mutants is opposite to the direction of effect observed at the protein level. The results from this study suggest that there is a hierarchy of rules governing uORF activity, where the biophysical properties of the uORF are dominant over the position of the uORF within the 5′ UTR.

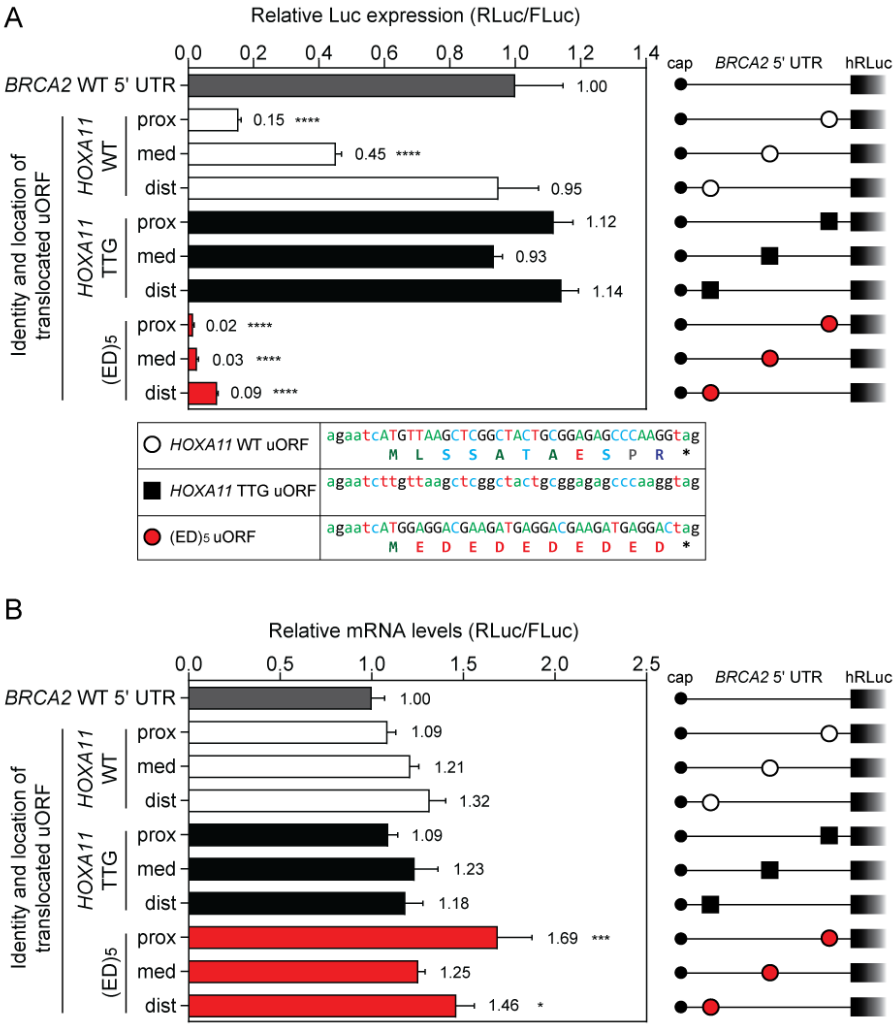


Figure 5.15 The effect of position on uORF activity within a uORF-naïve 5′ UTR. *HOXA11* WT, TTG, and $(ED)_5$ uORFs were cloned into positions proximal (prox), medial (med), or distal (dist) to the RLuc start codon within a uORF-naïve *BRCA2* 5′ UTR background (233 nt in length). The *BRCA2* WT 5′ UTR was included as a control. (**A**) Dual-luciferase assay measuring RLuc expression, normalised to FLuc. (**B**) RT-qPCR of RLuc mRNA transcripts, normalised to FLuc. Values are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the *BRCA2* WT 5′ UTR control. * = $P < 0.05$, *** = $P < 0.001$, and **** = $P < 0.0001$.

To determine whether non-functional uORFs are able to exhibit activity when placed in a different (and possibly more favourable) 5′ UTR context, a series of combination mutants of the *SMN2* and *HOXA11* uORFs were created and assessed using the pDLR system (**Figure 5.16A and B**). One set of mutants involved shifting the *SMN2* uORF from the distal (cap-proximal) to medial and proximal positions relative to the pORF within the *SMN2* 5′ UTR. Notably, in the two additional positional mutants, the sequence of the original *SMN2* uORF was retained but any possible translation activity disrupted by an ATG > TTG start codon mutation so as to introduce minimal additional changes into the UTR sequence. The *SMN2* uORF was active when shifted to the medial position, resulting in ~2-fold increase ($P < 0.0001$) in reporter activity when knocked out (**Figure 5.16B**). Interestingly, the *SMN2* uORF was found to be non-functional when placed in the proximal position (**Figure 5.16B**), contrary to the previous findings of progressively increased *HOXA11* uORF activity with increased proximity to the pORF within the *BRCA2* 5′ UTR context (**Figure 5.15**). These findings suggest that the *SMN2* uORF has the potential to be translationally repressive when located in a favourable position within the 5′ UTR.

Additional series of mutants were created by combining the features of the *SMN2* 5′ UTR and uORF with those of the *HOXA11* gene. This was carried out as the *HOXA11* uORF is functional, and the *SMN2* and *HOXA11* uORFs are similar in length (i.e. 8 and 11 codons respectively, **Figure 5.16A**). The major difference is that the *SMN2* uORF is positioned close to the 5′ cap, while the *HOXA11* uORF is proximal to the pORF start. Furthermore, the Kozak context sequence of the *SMN2* uORF is weaker than that of *HOXA11*. The repressive activity of the *SMN2* uORF was conferred when transplanted into the *HOXA11* 5′ UTR context, and was further enhanced by interchanging the weaker *SMN2* Kozak context for the stronger *HOXA11* uORF-specific Kozak consensus (**Figure 5.16B**). Furthermore, levels of activity of the *SMN2* uORF in the *HOXA11* 5′ UTR (combined with the *HOXA11* Kozak context) were

~8-fold higher for the TTG mutant relative to the WT version. Interestingly, this variant exhibits double the level of translational repression relative to the native *HOXA11* uORF. Finally, the activity of the *HOXA11* uORF is lost when placed in the *SMN2* context, regardless of the strength of the surrounding Kozak sequence context (**Figure 5.16B**). Other studies conducted using simian virus [565,581], yeast [596], and *in vitro* [582] models of translation have reported an inhibitory effect of uAUGs on downstream protein expression when in close proximity to the m^7G cap. The results reported in the present study provide evidence that uORF proximity to the cap is inhibitory in a human cell model context. Furthermore, the data justifies the hypothesis that a hierarchy of uORF features exists where, for the activity of native uORF sequences, Kozak context sequence as well as uORF position within the 5′ UTR were found to be important. Kozak consensus can therefore be a major determinant of uORF activity when the uORF is also in a favourable location, however, a strong kozak consensus cannot compensate for a uORF that is present in an unfavourable location (i.e. in proximity to the m^7G cap).

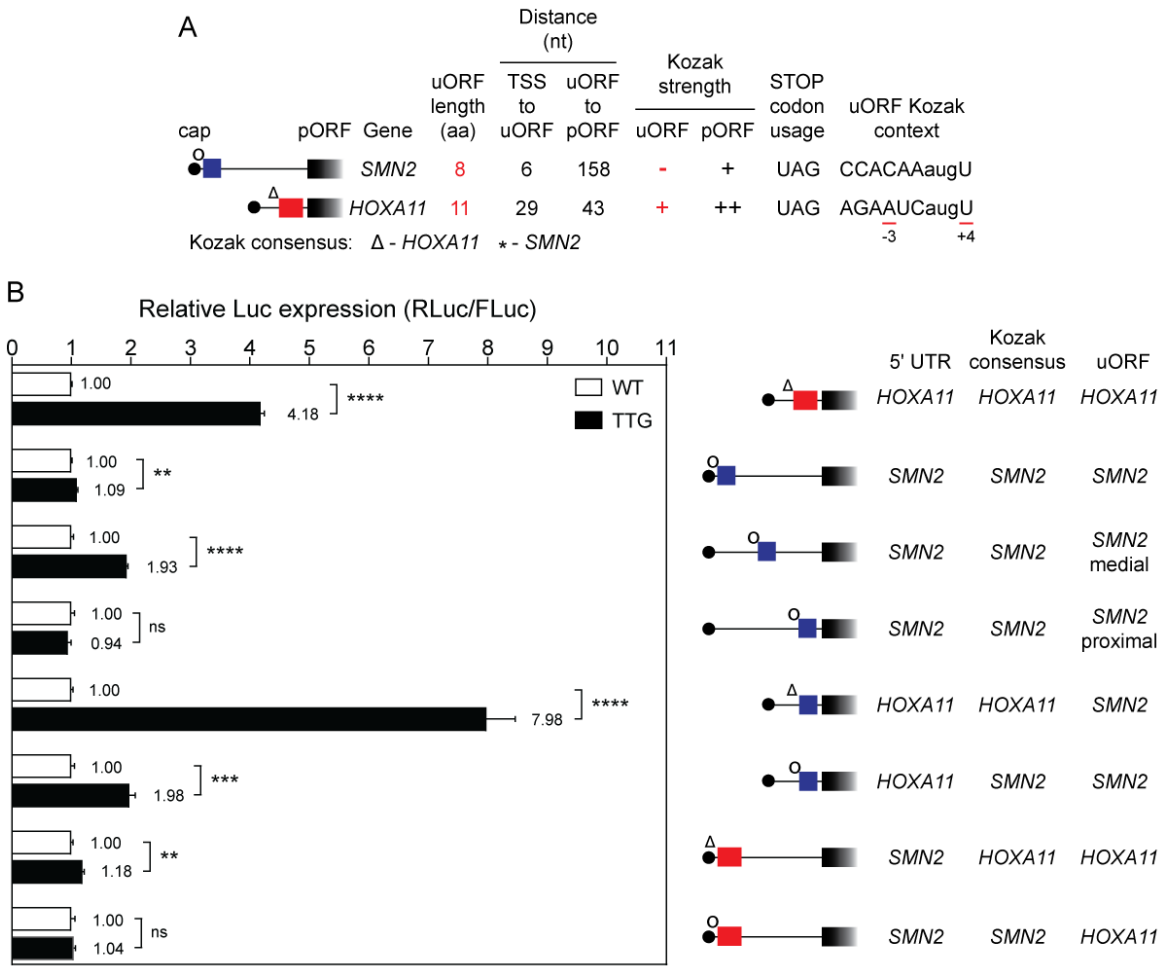


Figure 5.16 Activity of the *SMN2* uORF depends on surrounding sequence context.

The functionality of the *SMN2* uORF was assessed in the WT 5′ UTR context, as well as when it was shifted to the medial and proximal positions, and with different combinations of the *SMN2* or *HOXA11* 5′ UTR/Kozak consensus/uORF using the pDLR system. The *SMN2* and *HOXA11* uORFs are depicted with a blue and red box respectively. (**A**) Features of the *HOXA11* and *SMN2* 5′ UTRs and uORFs. Kozak consensus sequences: Δ = *HOXA11*; O = *SMN2*. (**B**) DLR assay measuring the WT (uORF-containing) and TTG (uORF-knockout) versions of the combination mutants. WT and TTG versions of the *HOXA11* and *SMN2* constructs were included as controls. Values are means ± SEM, and were scaled such that the mean of the corresponding WT control group was returned to a value of 1. Statistical significance was determined by a Student's *t*-test relative to the corresponding WT control. ** = $P < 0.01$, *** = $P < 0.001$, **** = $P < 0.0001$, ns = not significant. TSS, transcription start site.

5.4.4 The *HOXA11* uORF acts only in *cis*

To test whether the *HOXA11* uORF sequence is also able to mediate translational repressive effects on the luciferase reporter in *trans*, the hyper-functional synthetic uORFs (i.e. E_{10} and $(ED)_5$) within the *HOXA11* 5′ UTR context were cloned into RNA Pol II expression vectors (driven by the CMV promoter and enhancer) lacking a reporter gene. These were then co-transfected with the *HOXA11* WT or TTG mutant dual-luciferase reporter into cells and, as such, the effect of the micropeptide produced from the poly-acidic uORFs on the reporter gene construct could be measured (**Figure 5.17A**). Neither of the hyper-functional uORFs influenced the activity of the *HOXA11* WT reporter construct when supplied in *trans* (**Figure 5.17B**). These findings, taken together with the relatively low stability of the HiBiT-tagged uORF peptide (**Figure 5.9**), suggest that the *HOXA11* uORF generates a lowly abundant peptide which is rapidly degraded by the proteasome, and that the *HOXA11* uORF regulates its downstream gene in *cis*. This does not exclude the possibility that other independent *trans*-acting functions of the *HOXA11* uORF micropeptide may exist, as has been suggested of several human uORF-derived micropeptides in a recent study [363].

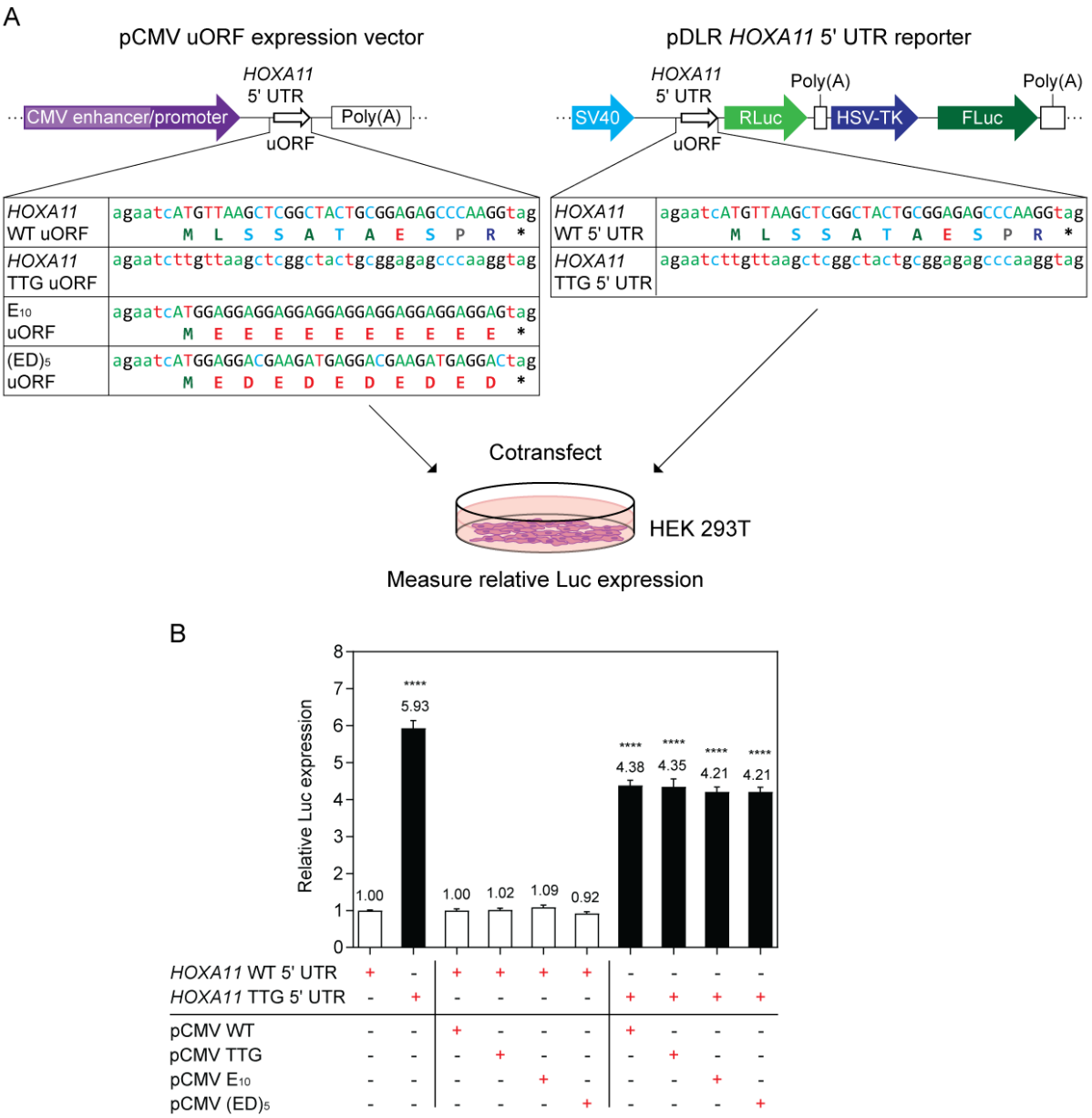


Figure 5.17 Translation initiation at a uORF acts on pORF in *cis*.

(**A**) Plasmids encoding the *HOXA11* WT, TTG, E_{10}, or $(ED)_5$ uORF driven by the cytomegalovirus (CMV) promoter were co-transfected into HEK293T cells with either the *HOXA11* WT or TTG pDLR vector to assess whether the uORF acts in *cis* or in *trans*. (**B**) Dual-luciferase assay measuring RLuc expression, normalised to FLuc. Values are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the *HOXA11* WT only control. **** = $P < 0.0001$.

5.4.5 Applications of uORFs

5.4.5.1 uORF-mediated transgene attenuation

Given the observed differences in levels of luciferase protein expression using various synthetic uORFs, the idea that this technology could be used as a means of attenuating transgene expression was explored. For example, this strategy could be applied to prevent the toxic overexpression of a gene therapy transgene as typically promoter optimisation has been focused on enhancing strength, rather than attenuating activity towards more physiologically relevant levels of expression [597]. The Poly-E mutant exhibited attenuation of the reporter protein expression in a manner proportional to the number of glutamic acid residues in the uORF sequence. As such, a series of truncations of the $(ED)_5$ hyper-functional mutant (i.e. the codon optimised version of Poly-E) within the *HOXA11* 5′ UTR background were cloned into a near infrared reporter protein (iRFP) vector, downstream of the CMV promoter and upstream of the iRFP gene (**Figure 5.18A**). These included $(ED)_5$, $(ED)_3$, and $(ED)_1$, and the $(ED)_5$ TTG control. Detection of iRFP expression as a result of uORF-mediated control was possible by direct measurement of iRFP fluorescence intensity. The transgene resides in an AAV-packagable vector to enable further *in vivo* experimentation. Specifically, the iRFP cassette is flanked by the inverted terminal repeats (ITRs) encoding a self-complementary AAV (scAAV) vector, whereby a mutation in the terminal resolution site (Δ*trs*) within the 3′ ITR facilitates hairpin formation during AAV packaging [598]. Consequently, this bypasses the need for the rate-limiting second-strand synthesis step, enabling more rapid transgene expression following transduction [598]. The aim of this experiment is to eventually assess whether defined lengths of the hyper-functional uORF mutant can control transgene attenuation in a predictable manner *in vivo*.

The system was first assessed in cell culture experiments, where the fluorescence intensity of iRFP was partially diminished by the presence of the full- and intermediate-length

hyper-functional uORFs by 70% ($P < 0.01$) and 40% respectively, but not for the shortest uORF micropeptide with only two acidic residues (i.e. $(ED)_1$), relative to the TTG uORF knockout control (**Figure 5.18B**). The pattern of uORF-mediated reporter translation repression was similar to that observed in the dual-luciferase system where the E_{10} and E_6 mutants resulted in 90% and 60% luciferase translation repression respectively (**Figure 5.12A**). While the effects were slightly weaker in the iRFP system, this confirms that the synthetic uORF can be employed for the proportional attenuation of transgenes in at least two different reporter gene contexts. Future experiments will involve testing the scAAV-iRFP uORF systems in mice using IVIS (*in vivo* imaging system), to assess whether uORF-mediated translation repression can be achieved in a controllable manner *in vivo*.

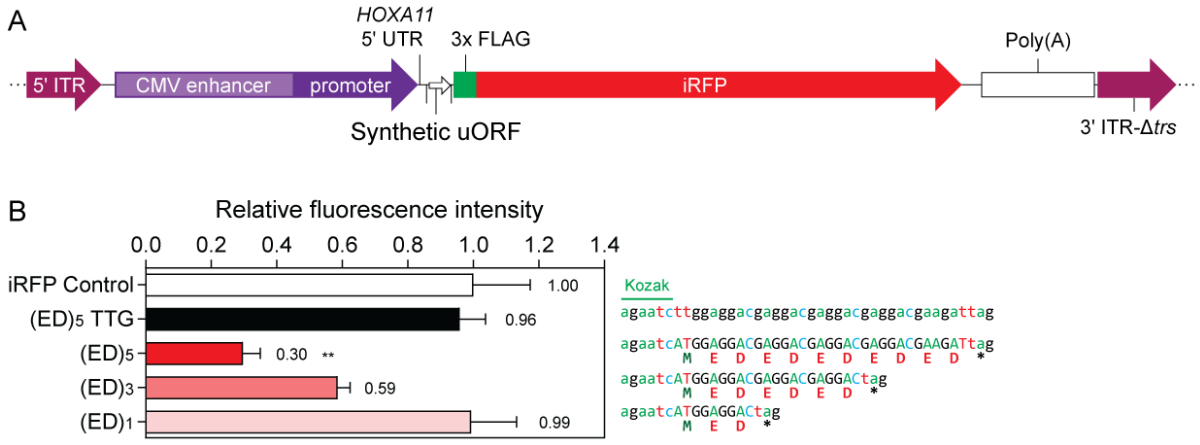


Figure 5.18 Transgene attenuation with truncations of hyper-functional uORFs.

(**A**) The 5′ leader sequence of the near infrared protein (iRFP) gene was replaced with the 5′ UTR of the *HOXA11* gene with four different versions of uORF (i.e. $(ED)_5$ TTG, $(ED)_5$, $(ED)_3$, and $(ED)_1$). The vector was additionally modified for expression of an N-terminal 3× FLAG-tagged protein. iRFP expression was driven by the CMV enhancer/promoter, and contained within a self-complementary AAV (scAAV) vector. The scAAV vector has a WT 5′ inverted terminal repeat (ITR), and a mutation within the terminal resolution site (Δ*trs*) 3′ ITR to promote hairpin formation of the AAV genome during packaging. (**B**) HEK293T cells were transfected with the iRFP vectors, and fluorescence intensity measured 48 hrs later. The uORF DNA and micropeptide sequences are depicted on the right. Values are means ± SEM, and were scaled such that the mean of the iRFP control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the iRFP control. ** = $P < 0.01$.

5.4.5.2 uORF knockout by CRISPR/ADAR-mediated RNA editing

The focus of this chapter thus far has been on elucidating the functional role of uORFs using synthetic reporter assay systems. The disruption or creation of uORFs has been proposed to be involved in a number of human diseases [359], and so the ability to manipuate endogenous uORF sequences *in situ* is highly attractive for therapeutic development of novel strategies that result in translational control of a particular GOI. One such strategy is to employ ASOs to mask uAUG sites from ribosomal recognition [392,599], thereby upregulating downstream protein expression. Work to this effect is ongoing in our lab. An alternative strategy involves RNA editing to mutate the uORF start codon. This can be achieved using engineered ADAR-recruiting RNA sequences [230,600–602], or CRISPR proteins fused to ADAR [222], for the conversion of uAUG to uIUG, where inosine (I) is read as G by the translation machinery. The CRISPR/ADAR system employs a Cas13b protein derived from *Prevotella sp.* (PspCas13b) translationally fused to ADAR, and a crispr RNA (crRNA) which directs the editing complex to the target A within the mRNA sequence. A mismatched C at the target A is required to achieve A>I editing.

The effect of CRISPR/ADAR-mediated uORF knockout on downstream gene expression was first assessed by targeting the *HOXA11* uORF within the luciferase reporter system. A panel of 15 crRNAs were designed to tile the sequence surrounding the uAUG, with a mismatched C at the target A, and screened for activity using the luciferase reporter assay (**Figure 5.19A**). The maximum mismatch distance (MD, i.e. the distance between the 3′ end of the crRNA target seuqence and the mismatched nucleotide corresponding to the target A site [222]) was limited to 30 nt as this corresponds to the start of the transcript. Cells were transfected with the dPspCas13b-ADAR, crRNA and pDLR *HOXA11* WT vectors, and the effect of uORF start site editing on uORF activity assessed by measuring luciferase reporter expression (**Figure 5.19B**). Treatment with the *HOXA11* WT or TTG vectors in combination with a non-

targeting crRNA were included as controls for maximal and minimal uORF-mediated repression respectively. Luciferase reporter expression was partially relieved by treatement with crRNA2 and crRNAs 9-13, relative to the *HOXA11* WT uORF non-targeting crRNA control. These levels of alleviation of translational repression ranged from 1.29-1.5-fold, where the target level of de-repression represented by the= uORF knockout control was 3-fold higher than the WT control. The MD of the cluster of functional crRNAs (i.e. crRNAs 9-13) was 23-24 nt, and some activity was also observed with a MD of 29 nt (i.e. with crRNA2). Future experiments are required to confirm the introduction of A>G mutations at the AUG start of the *HOXA11* uORF. Thereafter, the editing capacity of the CRISPR-ADAR system can be assessed on endogenous targets.

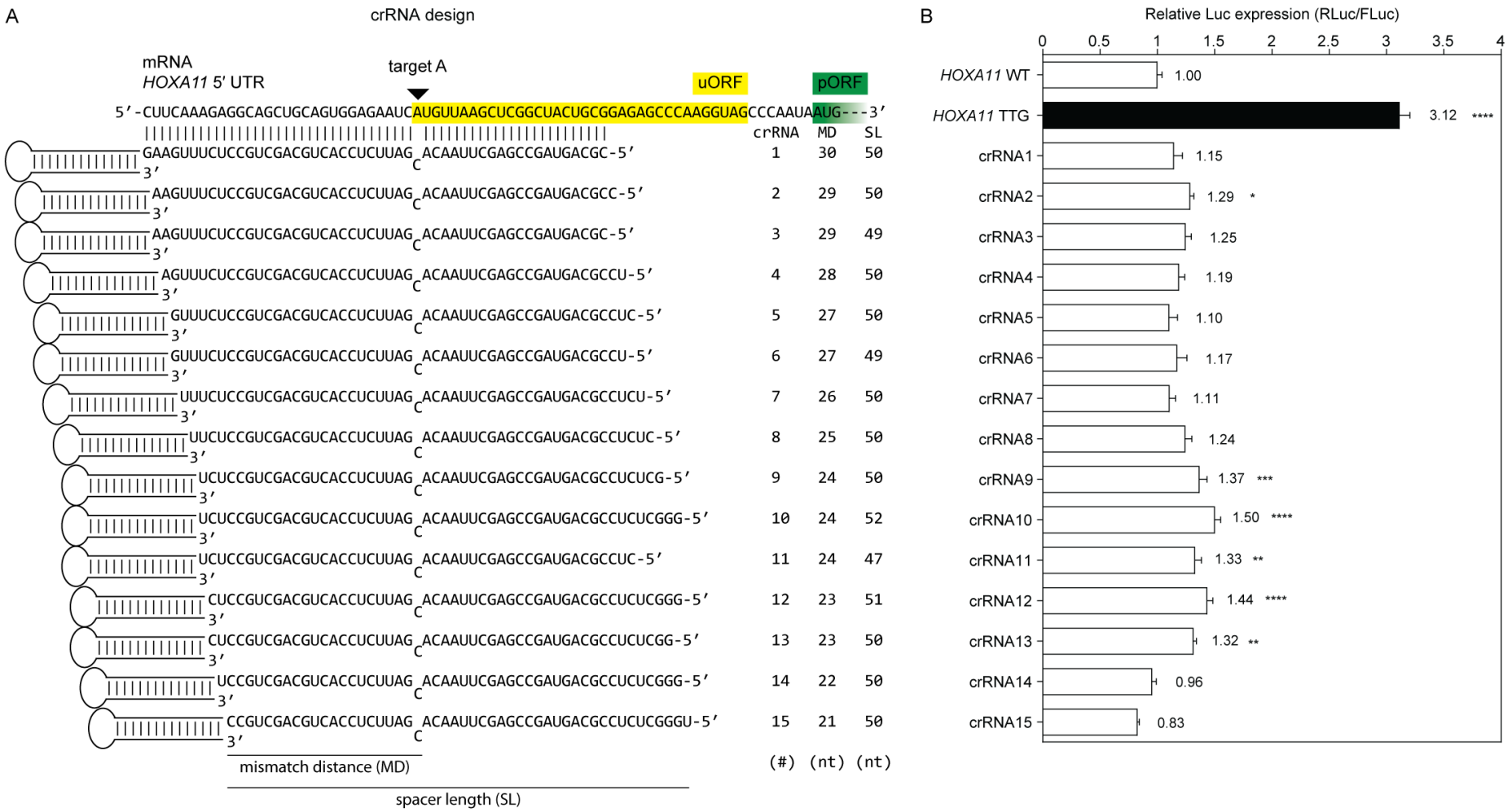

A
crRNA design
mRNA
HOXA11 5′ UTR
target A
uORF
pORF
5′-CUUCAAAGAGGCAGCUGCAGUGGAGAAUCAUGUUAAGCUCGGCUACUGCGGAGAGCCCAAGGUAGCCCAAUAAUG---3′
crRNA MD SL
GAAGUUUCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGC-5′ 1 30 50
AAGUUUCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCC-5′ 2 29 50
AAGUUUCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGC-5′ 3 29 49
AGUUUCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCU-5′ 4 28 50
GUUUCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCUC-5′ 5 27 50
GUUUCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCU-5′ 6 27 49
UUUCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCUCU-5′ 7 26 50
UUCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCUCUC-5′ 8 25 50
UCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCUCUCG-5′ 9 24 50
UCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCUCUCGGG-5′ 10 24 52
UCUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCUC-5′ 11 24 47
CUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCUCUCGGG-5′ 12 23 51
CUCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCUCUCGG-5′ 13 23 50
UCCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCUCUCGGG-5′ 14 22 50
CCGUCGACGUCACCUCUUAG C ACAAUUCGAGCCGAUGACGCCUCUCGGGU-5′ 15 21 50
(#) (nt) (nt)
mismatch distance (MD)
spacer length (SL)
B
Relative Luc expression (RLuc/FLuc)
0 0.5 1 1.5 2 2.5 3 3.5 4
HOXA11 WT 1.00
HOXA11 TTG 3.12 ****
crRNA1 1.15
crRNA2 1.29 *
crRNA3 1.25
crRNA4 1.19
crRNA5 1.10
crRNA6 1.17
crRNA7 1.11
crRNA8 1.24
crRNA9 1.37 ***
crRNA10 1.50 ****
crRNA11 1.33 **
crRNA12 1.44 ****
crRNA13 1.32 **
crRNA14 0.96
crRNA15 0.83

Figure 5.19 Design and screening of a panel of crRNAs for RNA editing of the *HOXA11* uORF start codon.

(**A**) A panel of 15 crRNAs were designed for targeting the *HOXA11* 5′ UTR with a mismatched cytosine (C) to facilitate adenine to guanine (A to G) editing in the mRNA transcript at the uORF AUG using a CRISPR/ADAR system derived from *Prevotella sp.* Cas13b (PspCas13b). The mismatch distance (MD) was optimised in the crRNA design, and the spacer length (SL) kept at ~ 50 nt. The crRNA hairpin structure required for recognition by the PspCas13b-ADAR fusion protein is shown. (**B**) HEK293T cells were transiently transfected with the CRISPR/ADAR and crRNA plasmids alongside the *HOXA11* WT pDLR vector to assess the effect on uORF activity using a dual-luciferase assay measuring RLuc expression, normalised to FLuc. *HOXA11* WT and TTG vectors combined with the CRISPR/ADAR and an empty crRNA plasmid were included as controls. Values are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the *HOXA11* WT + empty crRNA control. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$, and **** = $P < 0.0001$.

CRISPR/ADAR-mediated RNA editing results in an AUG to GUG modification. While the canonical start codon is known to be AUG, it has also been shown that alternative start codons can be recognised by the scanning ribosome to initiate translation [317,603]. A comparison between the levels of activity of luciferase reporters driven by the previously used TTG start codon uORF mutant, and a GTG uORF was performed as the CRISPR-ADAR-mediated RNA editing approach results in a GUG conversion within the mRNA sequence. No significant difference in activity of the *HOXA11* TTG and GTG uORF mutants was observed, confirming that the approach taken with CRISPR/ADAR-mediated RNA editing has the potential to achieve the desired outcome of uORF knockout (**Figure 5.20**). Interestingly, the hyper-functional $(ED)_5$ uORF is sensitive to alternative start codon usage, where the GTG and CTG mutants did not fully alleviate uORF repression to levels similar to the TTG mutant ($P < 0.0001$ for both), whereas the GGG, AAG, and AGG start codon mutations did. These findings are consistent with reports of the frequency of usage of non-AUG start codons in HEK293T [604] and mouse embryonic stem cells [560] where CTG and GTG were found to be the next most frequently used start codons after ATG. The apparent 'leaky' $(ED)_5$ uORF expression at CUG and GUG initiation sites is likely apparent as a result of the high levels of uORF-mediated translational repressive activity observed with these synthetic uORFs. Furthermore, the $(ED)_5$ mutant is in an optimal sequence context (i.e. a strong Kozak consensus sequence) which has also been shown to promote ribosomal initation at non-canonical start sites [605,606].

Future experiments will involve the application of the optimised CRISPR/ADAR editing for endogenous *HOXA11* uORF editing, and thereafter on alternative gene candidates where disruption of uORF(s) is expected to be therapeutically beneficial.

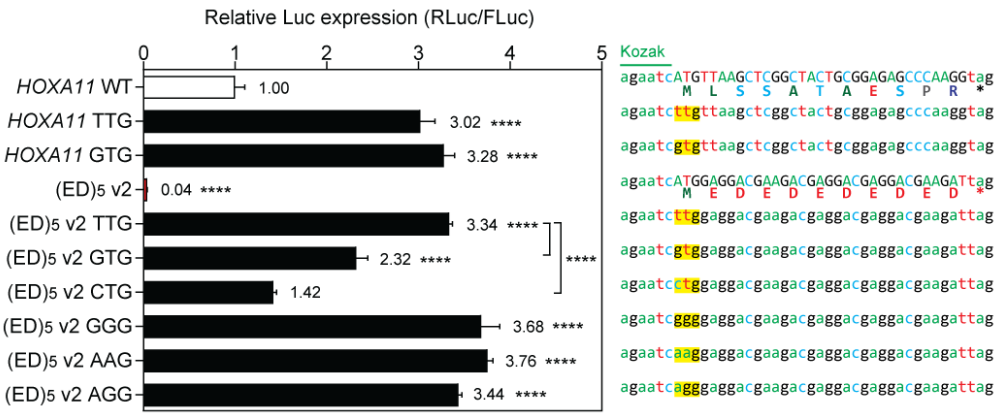


Figure 5.20 Non-canonical start codon mutants of the *HOXA11* WT and hyper-functional uORFs.

A *HOXA11* GTG start codon mutant, as well as a panel of $(ED)_5$ start codon mutants were created to assess whether uORF translation could be initiated at other non-canonical start sites. The *HOXA11* GTG mutant is akin to the CRISPR/ADAR-mediated RNA editing approach which would convert the AUG to GUG in the mRNA sequence. HEK293T cells were transiently transfected alongside the *HOXA11* WT and TTG control pDLR vectors to assess the effect of start codon mutations on uORF activity using a dual-luciferase assay measuring RLuc expression, normalised to FLuc. Values are means ± SEM, and were scaled such that the mean of the WT control group was returned to a value of 1. Statistical significance was determined by one-way ANOVA with Bonferroni *post hoc* test, and is shown relative to the *HOXA11* WT control. **** = $P < 0.0001$.

5.5 Discussion

5.5.1 Putative uORF functionality

The computational analyses performed on the human genome in this study identified putative uORFs in over half of human transcripts. This suggests that a large proportion of the genome may be held in a semi-repressed state. It remains to be determined what proportion of these putative uORFs are indeed functional, and whether the functionality is conserved between cell types and/or under different environmental conditions (e.g. the integrated stress response [607], or amino acid starvation [559]). It is possible to gain an insight into this using ribosome profiling and Ribo-Seq data, and work to this effect is ongoing in our lab. Previous methods of ribosome profiling involved the use of small molecules that arrest ribosome procession ~12-13 nt downstream of the TIS [386], however, this method would exclude smaller uORFs which are abundant in our bioinformatics pipeline and could be functional. Novel methods developed for single nucleotide resolution mapping of translational start sites, carried out in HEK293T [572,608] and THP-1 [609] cell lines, will be employed in future studies for ribosome profiling, including of minimal uORF sequences (i.e. ≥1 codon).

Results from this thesis suggest that the translational repressive effects of functional uORFs operate in *cis* with the pORF, and that the micropeptide is expressed at a low abundance with rapid turn-over. The micropeptides expressed from uORF sequences are thus unlikely to mediate other *trans* effects, although there may be rare exceptions with highly specific functional significance. For example, a class of micropeptide hormones, including myomaker (a 221 amino acid peptide) and myomixer (also called minion or myomerger, an 84 amino acid peptide), are small proteins that play pivotal roles in the process of muscle cell fusion [348,349,610,611]. The size of these micropeptide hormones is, however, larger than the length of approximately 75% of all predicted uORF-encoded sequences (i.e. ~1-30 amino acids). Interestingly, 26 amino acid synthetic micropeptides consisting of only leucine and isoleucine

(which were appropriately called LIL peptides) have been shown to promote cell growth [612], suggesting that certain micropeptides encoded by uORFs in the human genome could possess dual-functions, one being the *cis* repression of pORF translation and the other at a distant target. The development of novel synthetic micropeptides that promote cell growth may be beneficial in the context of muscle wasting diseases, however, uncontrollable cell growth also has the potential to result in tumorigenesis. Furthermore, any indirect biological functions of synthetic uORF-derived micropeptides should be carefully considered where introduced into cells to serve the canonical function of translational repression in *cis*.

Several technical challenges exist as a consequence of the small size of micropeptides when applying previously explored proteomic approaches for characterising global protein expression. Oyama and colleagues described the detection of novel micropeptides in the range of 40-100 amino acids, and alternative open reading frame proteins consisting of >29 amino acids have been detected by a combination of ribosome profiling and mass spectrometry [613]. Reasons for difficulties in isolating small peptides include that they may be physically removed during sample clean-up steps, uORF peptides that do not contain trypsin-sensitive amino acids (i.e. arginine and/or lysine) will be excluded from downstream processing, and mapping of very short trypsin-digested fragments is likely to be ambiguous. Furthermore, the data presented here suggests that the stability and abundance of uORF-derived micropeptides is relatively low. Modifications to current proteomics workflows are required for future micropeptidomic studies, including the generation of customised peptide spectradatabases to broaden the range of detectable uORF-derived micropeptides.

5.5.2 A hierarchy of uORF feature effects

The ability to isolate and manipulate individual uORF characteristics is paramount to being able to elucidate the relative contribution of each towards uORF functionality. Although fully

controlling for this is technically difficult (e.g. uORF composition and length are unavoidably linked), a hierarchy of effects was established by the generation and testing of various synthetic uORF mutants where the surrounding sequence context was kept constant (i.e. with the use of the *HOXA11* 5′ UTR). Features which were found to influence uORF activity will be elaborated on below in the order of importance as determined by the experiments conducted in this thesis. Such features are proposed to exert control on various stages of uORF translation, including initiation, stalling during elongation by the interaction between the nascent micropeptide biophysical properties and the ribosome tunnel. Furthermore, proximity to the m^7G cap, and distance from the pORF start would be expected to affect the time allowed for the ribosome to associate with relevant co-factors pre- and/or post-translation of the uORF micropeptide. As many of these features were found to influence uORF activity, this suggests that multiple mechanisms of uORF-mediated translational repression exist, independent of the NMD pathway.

5.5.2.1 Initiation of uORF translation

Initiation of translation at the uORF start site was found to be the most important determinant of uORF activity. Mutating the DNA start codon from ATG to TTG provided a useful means to validate and quantify the activity of the putative uORFs in a panel of genes identified through bioinformatic prediction. The levels of activity ranged from minimal to highly functional, and not all uORFs were found to be functional in their native context and under the cellular conditions used. Examples of such non-functional uORFs include those of *SMN2* and *UTRN* where, in the context of neuromuscular disease, upregulation of SMN2 or utrophin would be therapeutically beneficial for SMA and DMD respectively. While knocking out the uORFs harboured in the *SMN2* and *UTRN* 5′ UTRs is not therapeutically beneficial, such a strategy may be applied to other hapoinsufficiency diseases where uORF(s) exist, and naturally repress the translation of a protein that is required for disease alleviation.

The bioinformatic prediction of putative uORFs reported in this thesis excluded those with non-canonical start codons, as under normal physiological conditions, the majority of functional uORFs would be expected to initiate most efficiently at the universal AUG start site. However, it has been known for decades that ribosomes can also initiate translation by recognition of near-cognate codons (i.e. CUG, GUG, and UUG) in human cell lines, albeit at lower frequencies (**Figure 5.20**) [614]. Furthermore, ribosome profiling experiments suggest that functional uORFs are present at non-canonical start codons [560,608,615–617]. The direct detection of uORF micropeptide expression at non-canonical start sites has also been reported in a recent study where mass spectrometry was carried out on N-terminal enriched proteins [605]. The approach has been termed TAILS N-terminomics, and improves the detection of smaller ORF translation products by enriching for their N-termini. Other studies have applied this N-terminomic approach for the detection of non-canonical start uORF peptides in human [618], mouse [606,618], and plant [619] genomes, albeit with limited success, most likely due to various technical limitations discussed earlier [606,618]. Interestingly, the studies carried out on mammalian genomes reported a preference for a strong Kozak context (i.e. a purine in the -3 position, and a G at +4), when uORF translation was initiated at non-canonical start codons [605,606]. This may be an important contributing factor towards the repressive effects observed with the hyper functional mutant (i.e. $(ED)_5$) at CUG and GUG non-canonical start codons, as the glutamic acid codon introduced a G at the +4 position, thereby enhancing the strength of the native *HOXA11* uORF Kozak sequence (naturally containing an A in the -3 position). In contrast, the GTG *HOXA11* uORF start codon mutant was not functional, consistent with reports that the probability of initiation at this non-canonical start codon is <10% of AUG in HEK293T cells [603], which is further reduced by the sub-optimal Kozak sequence context used here. A low probability of translation initiation at the GUG near-cognate start codon is important as CRISPR/ADAR-mediated RNA editing which will be utilised for knocking out

endogenous uORFs in future studies (discussed in more detail below) facilitates the conversion of AUG to GUG, and is thus expected to abolish the activity of targeted native uORFs.

5.5.2.2 uORF amino acid identity and length

The ribosome has been shown to function as a metabolite sensor where, in the presence/absence of a particular intracellular molecule, translation is stalled during uORF peptide elongation by particular amino acid motifs or sequences [620]. For example, uORF-mediated translational repression has been shown to be enhanced by increases in cellular levels of arginine in fungi [621], sucrose in plants [622–624], polyamines in mammalian cells and yeast, [625], and magnesium in mammalian cells [394]. Takahashi and colleagues identified conserved uORFs across at least two orders of animals, and proposed that several of these were controlled by ribosomal metabolite sensing in a manner dependent on the uORF amino acid sequence [563].

uORF amino acid composition was determined to have a profound effect on translational activity. The uORF activity of the poly-acidic and -basic amino acid chains enabled suprarepression of reporter gene translation, while poly-hydrophobic tracts surprisingly had the opposite effect, partially alleviating the WT levels of translational repression. Codon optimisation of the hyper-functional poly-acidic uORF mutant did not alter the highly repressive activity, and uORF nucleic acid sequence and secondary structure were not found to be important for uORF activity, at least in the context of the *HOXA11* 5′ UTR background. Interestingly, the uORF mutants that contained alternating residues with different side chain properties compensated for each other, which suggests an important interaction between neighbouring residues and/or of the biophysical properties of the micropeptide. Although the homopolymeric amino acid sequences tested in these experiments are not typically found in nature and were not identified in our custom uORF database, the information gleaned from such experiments contributes towards the mechanistic understanding of uORF

activity. Furthermore, this may be useful for building computational models to predict uORF activity through the identification of runs of pro-repressive residues, or instances of individual amino acids which contribute towards uORF activity which data from the alanine scan experiment suggests may be relevant (**Figure 5.10**).

The activity of ribosomal peptidyl transferase is sensitive to pH levels [626], as well as to the biophysical properties of the side chains of amino acids contained in the 80S ribosomal P-site [627]. Furthermore, arrest peptides with particular sequences that interact with the ribosomal exit tunnel, usually in response to an environmental cue, have been described [628]. A local change in pH as a result of the side chain properties of the repeating amino acid residues in the homopolymeric chains may have had an impact on the functionality of peptidyl transferase, or interacted with the ribosomal exit tunnel, in both cases either promoting or relieving ribosomal stalling during elongation. Polysome (i.e. multiple ribosomes per transcript) and ribosome profiling experiments [559,560,629] enable separation of translated mRNA species based on the number of associated ribosomes from fragments of a defined size (i.e. ~30nt per ribosome) which are then deep sequenced to determine the level of translation at a particular locus. Combining ribosome and polysome profiling methods could provide a means to elucidating the mechanism of translational control conferred by the homopolymeric uORF mutants reported in this thesis. We propose that prolonged instances of ribosome stalling as a result of highly repressive uORFs would be expected to result in ribosomal queuing, whereby upstream scanning ribosomes are impeded by the uORF-stalled complex, and become stacked along the mRNA transcript [630]. In this way, sequenced polysome-protected mRNA fragments could be ranked based on the number of ribosomes at both the uORF and cognate pORF sites, and the level of translational repression determined.

While alanine scanning revealed that the individual contribution of the amino acids in the *HOXA11* uORF is minimal, several studies have suggested an influential role of particular

amino acids on uORF activity [563,631,632]. For instance, incorporation of a proline residue near the C-terminal end of the uORF peptide sequence has been found to result in ribosomal stalling, thereby impeding translation termination and the probability of reinitiation at downstream ORF starts [631,632]. Notably, the *HOXA11* uORF encodes a proline at the penultimate codon, but this was not found to be integral to uORF activity according to the alanine scanning experiment conducted. An interesting relationship exists between nucleic acid sequence and amino acid identity, which we did not consider in this thesis, involving translational pauses brought about by particular codon pairs/combinations in the A- and P-sites of the 80S ribosome [560,633]. The contribution of inhibitory codon pairs to global uORF activity warrants deeper investigation, as in certain instances, this may hypothetically override the effects of other uORF features such as length. This mechanism of translational repression may be similar to that observed with the single codon mutant (i.e. Met-Stop, the 'minimal uORF') through unusual occupation of codons in the active sites of the 80S ribosome. The *MAP3K1* gene is an example of a gene harbouring a single codon uORF, however, this was found to be non-functional (**Figure 5.4**). Nonetheless, a large proportion of the predicted uORFs were encoded by a single amino acid, and some of these minimal uORFs may indeed be active. Further validation experiments are required to determine the extent to which this is true.

The propensity for ribosome recycling post-termination is dependent on the speed of translation, whereby higher rates of ribosomal procession result in increased 'leaky' scanning [376]. Similarly, shorter uORFs have been shown to increase the probability for the 40S ribosomal subunit to remain bound to the transcript, thereby increasing the chances of reinitiation [376,377]. uORF length, however, is inextricably linked to uORF amino acid sequence identity. In the context of the *HOXA11* uORF, truncations of the uORF sequence exhibited predictable alleviating effects, with the exception of the single codon mutant (i.e. Met-Stop, 'minimal uORF') which functioned with similar activity to the WT uORF (discussed in more detail

above). Repetition of the repressive uORF sequence enhanced the repressive activity, however, adding a HiBiT or FLAG tag to the C-terminal end of the micropeptide sequence did not significantly alter uORF activity. The repressive effect of the E_{10} synthetic mutant was further enhanced by doubling the number of residues, suggesting that uORF length influences activity, but that the biophysical properties of the residues within the chain may exert a greater effect.

5.5.2.3 uORF position and Kozak sequence context

In addition to the strict requirement for the presence of a start codon for uORF activity, the probability for recognition and initiation by the scanning ribosome is increased by placement of the uORF in an optimal position and sequence context within the 5′ UTR. Short leader sequences are known to be inhibitory to translation as a result of reduced time allowed for the scanning 40S ribosomal subunit to accumulate the co-factors required for initiation prior to encountering the start codon [565,582]. Indeed, the *SMN2* uORF is positioned proximal to the m^7G cap and was found to be non-functional. Alternatively, translocating the *SMN2* uORF to the pORF-proximal position within the *HOXA11* 5′ UTR conferred translational repressive activity. Cap-proximal inhibitory effects were equally exerted on the WT *HOXA11* uORF when translocated into the distal position of the uORF-naïve (*BRCA2*) 5′ UTR. The activity of the $(ED)_5$ uORF increased with proximity to the pORF start in this context, however, relatively high levels of repression (i.e. 91%) were exhibited at the distal position, suggesting that the biophysical properties of the hyper-functional micropeptide override positional effects within the UTR.

Translational repression of the WT *HOXA11* uORF was greatest when in a proximal position relative to the luciferase reporter start site, and was progressively diminished with increasing distance from the primary coding sequence. The influence of uORF sequence on ribosome elongation is expected to be similar regardless of position within the UTR, however,

ribosomes that undergo 'leaky scanning' from uORFs that are proximal to the pORF start will be temporally restricted from regaining the appropriate factors required for reinitiation thereby conferring greater activity to the uORF. Furthermore, reduced sequence space between the uORF and the pORF decreases the possibility for internal ribosome entry, although the occurrence and significance of this phenomenon warrants further investigation. While in this chapter only 5′ UTR sequences with a single uORF that does not span the TIS were studied, a large proportion of transcripts were predicted to harbour >1 uORF and, in some cases, uORFs that span the TIS of the pORF. Multiplicity of uORFs are expected to further increase translational repression, and the uORFs that span the TIS may be equally, or more, repressive than the proximal uORF sequences. The contribution of these features to uORF activity will be investigated in future studies.

The bioinformatics analyses of putative uORFs in this thesis determined that uORFs, in general, are accompanied by weaker Kozak sequences, relative to primary coding sequences (**Figure 5.4**). Translation of the proteome has been found to be governed by the strength of Kozak consensus sequences, whereby ORFs preceded by weak Kozak contexts are more sensitive to drops in global initiation rates compared to those with stronger consensus sequences [634,635]. It is possible that uORFs with weak Kozak consensus contexts may be similarly regulated, thereby allowing for the upregulation of certain families of genes in response to changes in cellular conditions and global rates of translation [607]. The data described in this thesis demonstrate the complex nature of the effect of Kozak consensus sequence on uORF activity. Replacing the weaker *HOXA11* Kozak consensus sequence with an enhanced version resulted in increased activity, but this was not found to be statistically significant. Conversely, exchanging the weaker *SMN2* Kozak sequence with that of *HOXA11*, combined with translocation of the *SMN2* uORF into the *HOXA11* locus, generated a recombinant uORF with even greater activity than the native *HOXA11* uORF. Notably, the *SMN2* uORF sequence

encodes a Glu-Gly-Asp (i.e. EGD) motif, of which homopolymeric chains of each of these residues exhibited highly pro-repressive effects on reporter protein expression. The extreme range of functionality of the *SMN2* uORF as a result of its position relative to the pORF start, the sequence from which the 5′ UTR was derived, as well as the Kozak context highlights the interplay and complexities of these three key uORF features and the resulting effect on uORF activity. A recent study has identified over 80 different uORF features that could contribute towards activity [386], and while not all of these are easily experimentally tested, attributes other than those described in this thesis may be identified as important for determining uORF activity in future studies.

5.5.3 Future prospects of uORF research

5.5.3.1 Mechanism of uORF-mediated translational repression

The data reported in this thesis highlight the fact that various uORF features contribute towards translational repressive activity, but that the intricacies of the relative influence of combinations of these characteristics are highly complex. Global uORF validation efforts, and high throughput functional mutant studies are required to be able to fully elucidate the hierarchy of effects governing uORF activity, and to be able to build a reliable predictive model of uORF activity. Furthermore, it is apparent that several different mechanisms of translational repression are involved in uORF activity, which contributes towards our fundamental understanding of ribosome function. The major focus of future studies will involve trying to uncover the mechanism of uORF-mediated translational repression under different influences, such as for the highly repressive poly-acidic chain, and the two codon 'minimal' uORF. It will be important to determine whether ribosomes stall along the transcript, if the rate of translation is slowed significantly, or if some other unexpected event arises (such as premature ribosome recycling) to uncover the effects of pro-repressive amino acids on uORF activity. A method that could be employed to elucidate this phenomenon is that which has been previously

described [560], whereby cells are treated with harringtonine to inhibit initiating ribosomes immediately downstream of the start site, and thereafter treated at various timepoints with cycloheximide to immobilise the active ribosomes along the remainder of the uORF (i.e. those unaffected by harringtonine treatment). This is expected to provide information about the rate of translation, and the extent of ribosome queuing along the transcript [560]. Modifications to existing protocols will be required to capture ribosomal activity along uORFs, however, as the pool of synthetic uORFs used in this thesis was ~10-20 amino acids, and the average rate of translation in mammalian cells was reported to be ~5-6 amino acids per second [560]. Alternatively, longer synthetic uORFs could be designed to navigate these technical challenges (e.g. by the inclusion of peptide tags).

5.5.3.2 Biotechnological applications of uORFs for transgene expression

To date, available approaches for the control of transgene expression in classical gene therapy are relatively limited. Physiological and safe levels of transgene expression will vary depending on the target cell- or tissue-type, as well as the gene-of-interest. A recent publication highlights an important consideration for achieving long-term therapeutic levels of transgene expression, describing possible toxic gain of function effects in the motor neurons *in vivo* resulting from sustained AAV9-mediated SMN protein expression [636]. There is therefore a need to engineer systems that have the capacity to regulate transgene expression (i) in a cell- or tissue-specific manner, (ii) such that overall levels are not supraphysiological, and (iii) to enable inducible or for tightly regulated expression. Current approaches to cell- and tissue-specific control involve miRNA restriction [637,638], selective promoter usage, and viral vector tropism. A new generation of promoters inducible by inflammatory cytokines have also been developed to control transgene expression in response to environmental cues [639]. Strategies enabling fine-tuning of the levels of transgene expressed are largely limited to vector dosing, and/or the use of promoters with generally high or low levels of activity. Current state-of-the-art inducible

expression systems involve the use of spatial and temporal control through optogenetics [640,641], and ultrasound [642], as well as using small molecules (e.g. tetracycline) to induce promoter [643] or riboswitch [644,645] activity. The application of uORFs has the potential to address each of these challenges simultaneously, enabling rapid and dynamically regulated transgene expression.

Features that can be introduced upstream of transgenes facilitating inducible expression include riboswitches. These are structured RNA sequences (typically <100 nt in length) containing aptamers that function to control downstream gene expression in response to ligand binding [646]. The inclusion of riboswitches in the 5′ UTR of a transgene is attractive as the genetic footprint is small enough to easily fit within the limited packaging capacity of AAV vectors. Furthermore, ribsoswitches have been developed that can be induced/repressed by treatment with safe, FDA-approved small molecules such as tetracycline and theophylline [647]. Hypothetically, the use of riboswitch sequences that mask or expose splice signal sequences of exons containing uORFs as a result of ligand binding provides a means to create an inducible and reversible uORF on-off switch. Alternatively to the use of small molecules, ASOs could be employed to modulate riboswitch (and consequently uORF) activity. Furthermore, specification of the length of hyper-repressive uORF sequences, such as the $(ED)_5$ mutant, enables user-defined tailoring of transgene expression levels. Stratified transgene expression that does not rely on repetitive, unnatural homopolymeric runs of amino acid residues may be achieved through uORF multiplicity, or modular design of various uORF features such as Kozak context, and position within the UTR.

5.5.3.3 Disruption of uORFs in human disease

The abundance of uORF predictions in the human genome has opened up the possibility for uORF engineering to regulate the expression of genes with relevance to a plethora of human

diseases. This is a novel approach to translational control which can be carried out in a permanent manner by editing of the genomic DNA, or transiently through mRNA editing. Several studies have reported uORF knockout using classic single guide genomic DNA targeting CRISPR strategies in human cells [394] and in plants [395,396]. The recently-developed prime editing CRISPR tool enables the highly specific introduction of user-defined changes into the target site, and could be useful for complex manipulation of uORF features, such as the Kozak context strength, length, and amino acid composition. The DNA-targeting CRISPR approaches result in permanent changes to the genetic sequence which may be beneficial for haploinsufficiency diseases where the long-term restoration of protein expression is required for disease alleviation. Alternatively, where therapeutic activity is required within a narrow time-frame, CRISPR/ADAR RNA editing approaches, such as that employed in this thesis, offer a transient means to achieving uORF start codon knockout. Furthermore, the extent and long-term implications of off target gene editing are not fully known.

Riboswitches have recently been developed as rapid and reversible sensors to control CRISPR gene editing activity by the discovery of ligands that bind to the guide RNA scaffold sequence, thereby altering the secondary structure. In the presence of a ligand, the guide RNA sequence becomes unrecognisable by the CRISPR-associated protein, removing the possibility for gene editing. The use of a riboswitch to control CRISPR RNA editing activity that is directed towards uORF knockout could provide a safe and reversible therapeutic approach.

5.5.4 Conclusions

The research described in this chapter contributes towards elucidating the functional role of uORFs in *cis*-mediated translational attenuation of the associated downstream pORF. The activity of several candidates selected from an extensive database of predicted uORFs within the human transcriptome were successfully validated using a luciferase reporter system. Not

all of the putative uORFs were found to be functional in their native sequence context, however, particularly those that exist in close proximity to the m^7G cap of the mRNA transcript. Interestingly, activity of the non-functional *SMN2* uORF was achieved by altering its position and surrounding sequence context. A relatively complex hierarchy of effects based on several uORF properties emerged from studies based on an array of synthetic uORFs in the context of the *HOXA11* gene, placing micropeptide biophysical properties, position within the 5′ UTR, and Kozak context sequence higher on the order of importance compared to the uORF length, individual amino acid identities, nucleic acid sequence secondary structures, and G/C content. Deeper investigation on a global scale is required to further understand the relative role of various uORF properties in functionality. The *in situ* manipulation of uORFs using a CRISPR-ADAR system has the potential to upregulate protein expression in a highly targeted manner. This research provides a foundation for therapeutic development of novel molecular medicine approaches targeting uORFs, as well as for synthetic biology applications of uORFs.

6 Discussion and conclusions

This thesis has explored several potential molecular therapeutic approaches with a focus on neuromuscular disease (and DMD in particular). CRISPR-AAV-mediated genetic correction of the *Dmd* gene in the severely-affected dKO mouse restored dystrophin protein levels in the heart and diaphragm up to 4% of WT levels, although this was insufficient to extend the lifespan of these severely affected mice. Furthermore, within-fibre patchiness of dystrophin expression was observed along the length of treated dKO TA myofibres. Incomplete CRISPR/Cas9-mediated editing of myonuclei contributes further valuable insights into the field of understanding the myonuclear domain structure of myofibres. Important concerns are also raised by these findings, however, including whether total dystrophin levels are predictive of functional correction in myofibres that contain dystrophin-positive as well as dystrophin-negative sarcolemmal regions. As this within-fibre patchiness is not evident in PPMO-treated muscle (even at similarly low levels of dystrophin restoration) [456], understanding the diffusive potential of PPMOs versus CRISPR/Cas9 gene editing complexes and AAV delivery vectors will be highly informative for determining the therapy-specific target thresholds required for productive genetic correction in dystrophic myofibres. A combination therapeutic approach may help to mitigate lack of uniformity effects. For example, a 'priming' strategy could involve pre-treatment of patients with ASOs in order to transiently stabilise dystrophic muscle turnover, and thereby enhance the efficacy of an AAV-CRISPR therapy injected some time later by minimising vector genome loss [163]. A similar 'priming' approach was assessed by Forand and colleagues whereby dKO mice were treated with three injections of PPMO, as well as a low dose of AAV9-delivered U7 snRNA which, as a monotherapy, was far less effective [648]. Alternatively, should AAV initial treatment prove to be insufficient to alleviate disease pathology, it may be necessary to initiate treatment with ASOs later in life. Such a 'top up' style treatment has been proposed in the case of SMA, whereby initial treatment with

zolgensma (FDA-approved AAV-delivered gene therapy) may be supplemented with nusinersen (FDA-approved splice-correcting ASO) [649]. Biogen is currently leading a Phase IV clinical trial (NCT04488133) to evaluate the clinical benefit, as well as safety and tolerability, of nusinersen treatment in patients who have previously received zolgensma gene therapy.

AAV-based gene replacement therapies are in the late stages of clinical trials for DMD via the transfer of micro-dystrophin [650], and for X-linked myotubular myopathy via the transfer of the full-length cDNA sequence of Myotubularin 1 (*MTM1*, NCT03199469). A limitation of the gene replacement therapeutic approach is that there is a risk of ectopic overexpression of the transgene which may result in excessive levels of protein production, recently highlighted as a potential area of concern for zolgensma therapy (based on neuronal toxicity observed in mice) [636]. Conversely, promoter silencing by host cell-mediated epigenetic methylation is another ongoing challenge to long-term transgene expression in preclinical studies [651], although the persistence of transgene expression in patients is unknown as lifelong data is unavailable as of yet. Alternatively, genetic intervention using CRISPR/Cas9-based genetic therapies only requires initial productive editing for sustained correction. Commercial development of CRISPR therapies for DMD is currently in the early stages at a number of companies, including Editas Medicine (Cambridge, MA, USA), Vertex Pharmaceuticals (Boston, MA, USA), CRISPR Therapeutics (Zug, Switzerland), Eli Lilly (Indianapolis, IN, USA), and Intellia Therapeutics (Cambridge, MA, USA). The majority of research is still at the preclinical stage, however, Intellia is investigating the efficacy of the first systemically delivered CRISPR/Cas9 system (using non-viral lipid nanoparticles) for the genetic correction of hereditary transthyretin amyloidosis in patients (NCT04601051). The use of nanoparticles for genetic therapy delivery faces several challenges over viral vectors, however, including stability in the blood (i.e. protection from nucleases and proteases), avoiding sequestration and clearance by the reticuloendothelial system in the liver, extravasation out of the blood stream into target

tissues, and escape from endosomes to avoid degradation within the target cell [652]. Luxturna (voretigene neparvovec), a treatment for inherited retinal dystrophies (such as retinitis pigmentosa and Leber congenital amaurosis, LCA), was the first FDA-approved AAV-delivered gene therapy in 2017 [653]. EDIT-101 is an alternative genetic therapy for LCA type 10 that is an AAV-delivered CRISPR-based therapy that employs a dual-cut strategy for the removal of an aberrant splice donor sequence within intron 26 of the *CEP290* gene, and is currently in Phase II clinical trials (NCT03872479) [654]. While the eye is easily accessible for the delivery of genetic therapies, the large relative mass of muscle tissue presents a significant technical challenge for drug delivery to patients with DMD. Furthermore, improvements to current delivery modalities that enable satellite cell targeting is of paramount importance as these are the populations of cells that replenish the muscle tissue post-injury. A major advantage of AAV-based CRISPR/Cas9 gene editing therapies is the possibility for a single intervention, however, high doses of viral vectors will likely be required to enable editing in all muscles, and sufficient numbers of myonuclei in patients with neuromuscular disease. High dose AAV poses potential risks to patient health, and the resources required to generate the required amount of virus to treat all patients may potentially be prohibitive.

EVs that reduce fibrosis and inflammation, and that promote muscle regeneration, could be useful for the augmentation of other therapies (e.g. ASO-mediated exon skipping and CRISPR-mediated gene editing) by potentially improving the quality of the muscle architecture. Additionally, it is possible to load EVs with CRISPR/Cas9 gene editing components [655] which could be targeted for the correction of the genetic cause of a particular neuromuscular disease. Investigation of the pro-myogenic potential of the myotube-derived secretome in this thesis revealed that the vesicular fraction possesses transferrable pro-myogenic properties at lower, rather than higher, doses. A simple explanation for this could be that EVs are a complex mix of different biomolecules whereby threshold effects may exist at

particular concentrations and therefore manifest at different doses. A combination therapy that involves administration of *DMD*-corrective CRISPR-AAVs, as well as EVs that possess inherent pro-myogenic properties offers an alternative 'priming' or 'top up' strategy for DMD therapy development.

Pre- and/or post-production bioengineering of EVs, specifically in the context of skeletal muscle targeting, will be necessary to customise loading of pro-myogenic cargo, and to dictate tissue targeting and biodistribution of the regenerative EVs. At present, the dose of EVs required to effectively treat human patients is unknown. It remains to be determined whether doses used in animal studies will translate for use in humans, especially considering the immunological and anatomical differences between organisms. Nevertheless, clinical trials are underway to investigate the safety and efficacy of EV and exosome-based therapies [656]. The majority of these trials are focused on the therapeutic potential of EVs derived from mesenchymal stem cells (MSCs) which are known to transfer inherent anti-inflammatory and pro-regenerative signals via EVs [302]. The current global COVID-19 pandemic has also sparked interest in the therapeutic potential of MSC-derived EVs (administered via inhalation) to mitigate the cytokine storm experienced by some patients with novel coronavirus pneumonia as a result of SARS-CoV2 infection (NCT04276987, NCT04491240, and NCT04798716). Alternatively, MSC-derived EVs loaded with miR-124 (a miRNA expressed in the central nervous system that promotes proliferation and differentiation of neural progenitor cells) are being tested as an intervention for ischaemic stroke (NCT03384433), and those loaded with an anti-KRAS miRNA are being tested as a therapy (administered via the intravenous route) for pancreatic cancer (NCT03608631) [657]. Initially the hypothesis of chapter 2 of this thesis was that myomiRs within myotube-derived EVs are functionally transferrable, however, as both myotube- and myoblast-derived EVs were found to possess transferrable pro-myogenic effects, it suggests that some other signals may be involved and further research is required to uncover

potential candidates. EV-based therapies are yet to achieve FDA approval, however, preliminary findings from clinical trials report only mild to moderate side effects, indicating that the administration of EVs to patients can generally be considered safe [658]. The use of EVs as delivery vehicles is being developed by EVOX Therapeutics (Oxford, UK) and Kodiak Biosciences (Cambridge, MA, USA).

The major benefits of employing EVs as therapeutic modalities are that exosome release is regulated, production can be easily scaled-up, the encapsulation of biomolecular cargo is protective, EVs are readily internalised by target cells, and the components are released in a functional manner. Recent years have seen rapid advancements in the development of methods for EV bioengineering that enable specific therapeutic cargo loading, as well as tissue targeting [656,659]. The production of potent EV-based drugs will, however, also be dependent on reproducible methods for high quality, high yield isolation. An important consideration highlighted by the results in this thesis are that the co-purification of culture media components may contaminate the final preparation of EVs. This is of particular importance as this may lead to the misattribution of paracrine effects to EVs. Strategies that enhance EV production without necessitating the use of proportional increases in producer cell numbers and thereby mitigating the current limitations to manufacturing capacities and the associated costs will be valuable for the scalable manufacture of EVs for clinical applications. However, the use of culture media that promotes cell growth and/or proliferation to achieve this end should be carefully considered. The data from this thesis supports the use of methods such as UF-SEC for EV isolation, and highlights that some media additives should be avoided.

A large number of human diseases arise from a loss or reduction in function of a single gene and, as such, strategies for the targeted upregulation of protein expression are highly attractive for the development of therapeutics for haploinsufficiency disease. A number of companies are developing technologies based on RNA-mediated activation of protein

expression. For example, MiNA Therapeutics (London, UK) is investigating the use of small activating RNAs to enhance gene expression at the level of transcription. Stoke Therapeutics (Bedford, MA, USA) has developed a technology called TANGO (targeted augmentation of nuclear gene output) that employs splice-corrective ASOs for the exclusion of naturally occurring 'poison exons' (i.e. exons containing a premature termination codon) in mRNA transcripts of genes involved in haploinsufficiency disease (e.g. Dravet's syndrome). TransSINE Technologies (Yokohoma, Japan), GeneTx Biotherapeutics (Sarasota, FL, USA), and CAMP4 Therapeutics (Cambridge, MA, USA) are all developing technologies that employ ASOs to knock down gene-specific regulatory non-coding RNAs, thereby increasing protein expression from the cognate mRNA target. To date there are no approved drugs in this research area, and the development of other genetic therapy approaches to gene activation at the translational level is highly desirable.

The *in situ* targeting of uORFs offers a direct means of enhancing protein expression in haploinsufficiency diseases by disrupting the naturally existing mechanisms of *cis* translational repression of the pORF from within the functional allele. Preliminary analysis from our uORF bioinformatics prediction approach identified many applicable monogenic haploinsufficiency genes containing uORFs. The obvious choices for upregulation targets for the treatment of neuromuscular diseases are the uORFs within *SMN2* and *UTRN* transcripts (to treat SMA and DMD respectively). Steric block oligonucleotides have been employed to mask uORFs and thereby upregulate pORF expression [393]. This approach was taken in a recent study to mask the uORF within the 5′ UTR of *SMN2* [660]. While SMN protein expression was found to be increased, this was attributed to mRNA transcript stabilisation by oligonucleotide hybridisation, rather than to the disruption of uORF activity. Importantly, the data from this thesis demonstrate that both the *SMN2* and *UTRN* uORFs are non-functional in their native genetic contexts, suggesting that these are unlikely to constitute attractive therapeutic targets.

The use of oligonucleotides with secondary structures that result in the recruitment of cellular ADARs have also been employed to effect site-specific RNA editing [661,662]. Such strategies could be used for uORF start codons, thereby disrupting uORF activity. In this thesis, RNA editing using a CRISPR/Cas13b system was investigated as an alternative method that could be developed in future studies as a means for endogenous disruption of uORF activity at a user-defined target. Hypothetically, the targeted disruption of uORF activity using steric block oligonucleotides, or RNA editing, could be applied to any gene where upregulation could be beneficial. For example, targeting of the uORFs in the *BDNF* 5′ UTR could be utilised to activate expression of this neuroprotective protein, which could be therapeutically beneficial in multiple neurological diseases. Conversely, the introduction of a uORF by AUA > AUG RNA editing may enable repression of a gene without abolishing full activity [661,662]. This could be useful in the context of aberrant increases in expression of proto-oncogenes which also serve important cellular functions.

The ribosome possesses an interesting non-canonical function as a metabolite sensor. Indeed the activity of a number of highly conserved uORFs are involved in feedback loops with the metabolic product of the associated pORF [563,663,664]. Some of these are of potential relevance to DMD where the absence of dystrophin results not only in structurally compromised myofibres, but also dysregulation in metabolic pathways and muscle homeostasis. Examples of conserved uORFs are *SLC6A8*, which is involved in creatine transportation and regeneration of ATP from ADP in cells [663], and *LRRC8B*, which is involved in intracellular Ca^{2+} signalling [664]. Identifying conserved dual-functional uORFs that are involved in feedback loops with the ribosome and protein product of the associated translationally repressed downstream pORF, could be an interesting novel avenue of exploration for the development of therapies for dystrophinopathies.

In conclusion, the utility of AAV-deliverable CRISPR/Cas9-mediated gene editing, EV-based pro-myogenic paracrine signalling, and protein upregulation through the modulation of uORF activity, were each investigated separately in this thesis as potential molecular medicine approaches for the development of therapies for neuromuscular disease. Valuable insights into the successes and limitations of the technological design and implementation of each approach were highlighted. The complexity of neuromuscular disease is such that a combination therapy approach that employs two or more genetic therapy strategies such as those described in this thesis could be therapeutically beneficial and should be considered in future iterations of therapeutic design.

7 References

1. Benarroch L, Bonne G, Rivier F, Hamroun D. The 2020 version of the gene table of neuromuscular disorders (nuclear genome). Neuromuscul Disord 2019; 29:980–1018.

2. Mercuri E, Bönnemann CG, Muntoni F. Muscular dystrophies. The Lancet 2019; 394:2025–38.

3. Hoffman EP, Brown RH, Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell 1987; 51:919–28.

4. Lee T, Takeshima Y, Kusunoki N, Awano H, Yagi M, Matsuo M, Iijima K. Differences in carrier frequency between mothers of Duchenne and Becker muscular dystrophy patients. J Hum Genet 2014; 59:46–50.

5. Janssen I, Heymsfield SB, Wang ZM, Ross R. Skeletal muscle mass and distribution in 468 men and women aged 18-88 yr. J Appl Physiol (1985) 2000; 89:81–8.

6. Pedersen BK. Muscle as a secretory organ. Comprehensive Physiology 2013; 3:1337–62.

7. Pedersen BK. Muscles and their myokines. J Exp Biol 2011; 214:337–46.

8. Pedersen BK, Steensberg A, Fischer C, Keller C, Keller P, Plomgaard P, Febbraio M, Saltin B. Searching for the exercise factor: is IL-6 a candidate? J Muscle Res Cell Motil 2003; 24:113–9.

9. Lee JH, Jun H-S. Role of Myokines in Regulating Skeletal Muscle Mass and Function. Front Physiol 2019; 10:42.

10. Kim JH, Jin P, Duan R, Chen EH. Mechanisms of myoblast fusion during muscle development. Current Opinion in Genetics & Development 2015; 32:162–70.

11. Schiaffino S, Reggiani C. Fiber types in mammalian skeletal muscles. Physiol Rev 2011; 91:1447–531.

12. Sasaki T, Giltay R, Talts U, Timpl R, Talts JF. Expression and distribution of laminin alpha1 and alpha2 chains in embryonic and adult mouse tissues: an immunochemical approach. Exp Cell Res 2002; 275:185–99.

13. Ehrig K, Leivo I, Argraves WS, Ruoslahti E, Engvall E. Merosin, a tissue-specific basement membrane protein, is a laminin-like protein. Proc Natl Acad Sci U S A 1990; 87:3264–8.

14. Colognato H, Yurchenco PD. Form and function: the laminin family of heterotrimers. Dev Dyn 2000; 218:213–34.

15. Patton BL, Connoll AM, Martin PT, Cunningham JM, Mehta S, Pestronk A, Miner JH, Sanes JR. Distribution of ten laminin chains in dystrophic and regenerating muscles. Neuromuscul Disord 1999; 9:423–33.

16. Patton BL, Miner JH, Chiu AY, Sanes JR. Distribution and function of laminins in the neuromuscular system of developing, adult, and mutant mice. J Cell Biol 1997; 139:1507–21.

17. Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 1987; 50:509–17.

18. Koenig M, Monaco AP, Kunkel LM. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. Cell 1988; 53:219–28.

19. van Ommen GJ, Bertelson C, Ginjaar HB, den Dunnen JT, Bakker E, Chelly J, Matton M, van Essen AJ, Bartley J, Kunkel LM. Long-range genomic map of the Duchenne muscular dystrophy (DMD) gene: isolation and use of J66 (DXS268), a distal intragenic marker. Genomics 1987; 1:329–36.

20. van Ommen GJ, Verkerk JM, Hofker MH, Monaco AP, Kunkel LM, Ray P, Worton R, Wieringa B, Bakker E, Pearson PL. A physical map of 4 million bp around the Duchenne muscular dystrophy gene on the human X-chromosome. Cell 1986; 47:499–504.

21. Burmeister M, Lehrach H. Long-range restriction map around the Duchenne muscular dystrophy gene. Nature 1986; 324:582–5.

22. Nudel U, Zuk D, Einat P, Zeelon E, Levy Z, Neuman S, Yaffe D. Duchenne muscular dystrophy gene product is not identical in muscle and brain. Nature 1989; 337:76–8.

23. Holder E, Maeda M, Bies RD. Expression and regulation of the dystrophin Purkinje promoter in human skeletal muscle, heart, and brain. Hum Genet 1996; 97:232–9.

24. Muntoni F, Torelli S, Ferlini A. Dystrophin and mutations: one gene, several proteins, multiple phenotypes. Lancet Neurol 2003; 2:731–40.

25. D'Souza VN, Nguyen TM, Morris GE, Karges W, Pillers DA, Ray PN. A novel dystrophin isoform is required for normal retinal electrophysiology. Hum Mol Genet 1995; 4:837–42.

26. Lidov HG, Selig S, Kunkel LM. Dp140: a novel 140 kDa CNS transcript from the dystrophin locus. Hum Mol Genet 1995; 4:329–35.

27. Tozawa T, Itoh K, Yaoi T, Tando S, Umekage M, Dai H, Hosoi H, Fushiki S. The shortest isoform of dystrophin (Dp40) interacts with a group of presynaptic proteins to form a presumptive novel complex in the mouse brain. Mol Neurobiol 2012; 45:287–97.

28. Martínez-Herrera A, Aragón J, Bermúdez-Cruz RM, Bazán ML, Soid-Raggi G, Ceja V, Santos Coy-Arechavaleta A, Alemán V, Depardón F, Montañez C. In silico analyses of dystrophin Dp40 cellular distribution, nuclear export signals and structure modeling. Data Brief 2015; 4:349–52.

29. Byers TJ, Lidov HG, Kunkel LM. An alternative dystrophin transcript specific to peripheral nerve. Nat Genet 1993; 4:77–81.

30. Hugnot JP, Gilgenkrantz H, Vincent N, Chafey P, Morris GE, Monaco AP, Berwald-Netter Y, Koulakoff A, Kaplan JC, Kahn A. Distal transcript of the dystrophin gene initiated from an alternative first exon and encoding a 75-kDa protein widely distributed in nonmuscle tissues. Proc Natl Acad Sci U S A 1992; 89:7506–10.

31. Tadayoni R, Rendon A, Soria-Jasso LE, Cisneros B. Dystrophin Dp71: the smallest but multifunctional product of the Duchenne muscular dystrophy gene. Mol Neurobiol 2012; 45:43–60.

32. Broderick MJF, Winder SJ. Spectrin, alpha-actinin, and dystrophin. Adv Protein Chem 2005; 70:203–46.

33. Prins KW, Humston JL, Mehta A, Tate V, Ralston E, Ervasti JM. Dystrophin is a microtubule-associated protein. J Cell Biol 2009; 186:363–9.

34. Belanto JJ, Mader TL, Eckhoff MD, Strandjord DM, Banks GB, Gardner MK, Lowe DA, Ervasti JM. Microtubule binding distinguishes dystrophin from utrophin. Proc Natl Acad Sci U S A 2014; 111:5723–8.

35. Rybakova IN, Amann KJ, Ervasti JM. A new model for the interaction of dystrophin with F-actin. J Cell Biol 1996; 135:661–72.

36. Koenig M, Kunkel LM. Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility. J Biol Chem 1990; 265:4560–6.

37. Ervasti JM, Sonnemann KJ. Biology of the striated muscle dystrophin-glycoprotein complex. Int Rev Cytol 2008; 265:191–225.

38. Ibraghimov-Beskrovnaya O, Ervasti JM, Leveille CJ, Slaughter CA, Sernett SW, Campbell KP. Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. Nature 1992; 355:696–702.

39. Ervasti JM, Campbell KP. A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin. J Cell Biol 1993; 122:809–23.

40. Ervasti JM, Campbell KP. Membrane organization of the dystrophin-glycoprotein complex. Cell 1991; 66:1121–31.

41. Rentschler S, Linn H, Deininger K, Bedford MT, Espanel X, Sudol M. The WW Domain of Dystrophin Requires EF-Hands Region to Interact with β-Dystroglycan. 1999; 380:431–42.

42. Brenman JE, Chao DS, Xia H, Aldape K, Bredt DS. Nitric oxide synthase complexed with dystrophin and absent from skeletal muscle sarcolemma in Duchenne muscular dystrophy. Cell 1995; 82:743–52.

43. Chang WJ, Iannaccone ST, Lau KS, Masters BS, McCabe TJ, McMillan K, Padre RC, Spencer MJ, Tidball JG, Stull JT. Neuronal nitric oxide synthase and dystrophin-deficient muscular dystrophy. Proc Natl Acad Sci U S A 1996; 93:9142–7.

44. Brenman JE, Chao DS, Gee SH, McGee AW, Craven SE, Santillano DR, Wu Z, Huang F, Xia H, Peters MF, et al. Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains. Cell 1996; 84:757–67.

45. Seale P, Rudnicki MA. A new look at the origin, function, and "stem-cell" status of muscle satellite cells. Dev Biol 2000; 218:115–24.

46. Relaix F, Zammit PS. Satellite cells are essential for skeletal muscle regeneration: the cell on the edge returns centre stage. Development 2012; 139:2845–56.

47. Webster C, Blau HM. Accelerated age-related decline in replicative life-span of Duchenne muscular dystrophy myoblasts: implications for cell and gene therapy. Somat Cell Mol Genet 1990; 16:557–65.

48. Novak JS, Mázala DAG, Nearing M, Hindupur R, Uapinyoying P, Habib NF, Dickson T, Ioffe OB, Harris BT, Fidelia-Lambert MN, et al. Human muscle stem cells are refractory to aging. Aging Cell 2021; :e13411.

49. Collins CA, Zammit PS, Ruiz AP, Morgan JE, Partridge TA. A population of myogenic stem cells that survives skeletal muscle aging. Stem Cells 2007; 25:885–94.

50. Douglas AGL, Wood MJA. Splicing therapy for neuromuscular disease. Mol Cell Neurosci 2013; 56:169–85.

51. Petrof BJ, Shrager JB, Stedman HH, Kelly AM, Sweeney HL. Dystrophin protects the sarcolemma from stresses developed during muscle contraction. Proc Natl Acad Sci U S A 1993; 90:3710–4.

52. Whitmore C, Morgan J. What do mouse models of muscular dystrophy tell us about the DAPC and its components? Int J Exp Pathol 2014; 95:365–77.

53. Messina S, Bitto A, Aguennouz M, Minutoli L, Monici MC, Altavilla D, Squadrito F, Vita G. Nuclear factor kappa-B blockade reduces skeletal muscle degeneration and enhances muscle function in Mdx mice. Exp Neurol 2006; 198:234–41.

54. Pastoret C, Sebille A. mdx mice show progressive weakness and muscle deterioration with age. J Neurol Sci 1995; 129:97–105.

55. Yokota T, Lu Q-L, Morgan JE, Davies KE, Fisher R, Takeda S, Partridge TA. Expansion of revertant fibers in dystrophic mdx muscles reflects activity of muscle precursor cells and serves as an index of muscle regeneration. J Cell Sci 2006; 119:2679–87.

56. Szabo SM, Salhany RM, Deighton A, Harwood M, Mah J, Gooch KL. The clinical course of Duchenne muscular dystrophy in the corticosteroid treatment era: a systematic literature review. Orphanet Journal of Rare Diseases 2021; 16:237.

57. Moriuchi T, Kagawa N, Mukoyama M, Hizawa K. Autopsy analyses of the muscular dystrophies. Tokushima J Exp Med 1993; 40:83–93.

58. Chiang DY, Allen HD, Kim JJ, Valdes SO, Wang Y, Pignatelli RH, Lotze TE, Miyake CY. Relation of Cardiac Dysfunction to Rhythm Abnormalities in Patients With Duchenne or Becker Muscular Dystrophies. Am J Cardiol 2016; 117:1349–54.

59. Ishikawa Y, Miura T, Ishikawa Y, Aoyagi T, Ogata H, Hamada S, Minami R. Duchenne muscular dystrophy: survival by cardio-respiratory interventions. Neuromuscul Disord 2011; 21:47–51.

60. Moizard MP, Toutain A, Fournier D, Berret F, Raynaud M, Billard C, Andres C, Moraine C. Severe cognitive impairment in DMD: obvious clinical indication for Dp71 isoform point mutation screening. Eur J Hum Genet 2000; 8:552–6.

61. Daoud F, Angeard N, Demerre B, Martie I, Benyaou R, Leturcq F, Cossée M, Deburgrave N, Saillour Y, Tuffery S, et al. Analysis of Dp71 contribution in the severity of mental retardation through comparison of Duchenne and Becker patients differing by mutation consequences on Dp71 expression. Hum Mol Genet 2009; 18:3779–94.

62. Connolly AM, Florence JM, Cradock MM, Malkus EC, Schierbecker JR, Siener CA, Wulf CO, Anand P, Golumbek PT, Zaidman CM, et al. Motor and cognitive assessment of infants and young boys with Duchenne Muscular Dystrophy: results from the Muscular Dystrophy Association DMD Clinical Research Network. Neuromuscul Disord 2013; 23:529–39.

63. Larson CM, Henderson RC. Bone mineral density and fractures in boys with Duchenne muscular dystrophy. J Pediatr Orthop 2000; 20:71–4.

64. Love DR, Hill DF, Dickson G, Spurr NK, Byth BC, Marsden RF, Walsh FS, Edwards YH, Davies KE. An autosomal transcript in skeletal muscle with homology to dystrophin. Nature 1989; 339:55–8.

65. Pearce M, Blake DJ, Tinsley JM, Byth BC, Campbell L, Monaco AP, Davies KE. The utrophin and dystrophin genes share similarities in genomic structure. Hum Mol Genet 1993; 2:1765–72.

66. Tomé FM, Matsumura K, Chevallay M, Campbell KP, Fardeau M. Expression of dystrophin-associated glycoproteins during human fetal muscle development: a preliminary immunocytochemical study. Neuromuscul Disord 1994; 4:343–8.

67. Weir AP, Burton EA, Harrod G, Davies KE. A- and B-utrophin have different expression patterns and are differentially up-regulated in mdx muscle. J Biol Chem 2002; 277:45285–90.

68. Tinsley JM, Potter AC, Phelps SR, Fisher R, Trickett JI, Davies KE. Amelioration of the dystrophic phenotype of mdx mice using a truncated utrophin transgene. Nature 1996; 384:349–53.

69. Tinsley J, Deconinck N, Fisher R, Kahn D, Phelps S, Gillis J-M, Davies K. Expression of full-length utrophin prevents muscular dystrophy in mdx mice. Nat Med 1998; 4:1441–4.

70. Kleopa KA, Drousiotou A, Mavrikiou E, Ormiston A, Kyriakides T. Naturally occurring utrophin correlates with disease severity in Duchenne muscular dystrophy. Hum Mol Genet 2006; 15:1623–8.

71. Vainzof M, Passos-Bueno MR, Man N, Zatz M. Absence of correlation between utrophin localization and quantity and the clinical severity in Duchenne/Becker dystrophies. Am J Med Genet 1995; 58:305–9.

72. Guiraud S, Squire SE, Edwards B, Chen H, Burns DT, Shah N, Babbs A, Davies SG, Wynne GM, Russell AJ, et al. Second-generation compound for the modulation of utrophin in the therapy of DMD. Hum Mol Genet 2015; 24:4212–24.

73. Gilbert R, Nalbantoglu J, Petrof BJ, Ebihara S, Guibinga GH, Tinsley JM, Kamen A, Massie B, Davies KE, Karpati G. Adenovirus-mediated utrophin gene transfer mitigates the dystrophic phenotype of mdx mouse muscles. Hum Gene Ther 1999; 10:1299–310.

74. Fallon JR, McNally EM. Non-Glycanated Biglycan and LTBP4: Leveraging the extracellular matrix for Duchenne Muscular Dystrophy therapeutics. Matrix Biol 2018; 68–69:616–27.

75. Wilkinson IVL, Perkins KJ, Dugdale H, Moir L, Vuorinen A, Chatzopoulou M, Squire SE, Monecke S, Lomow A, Geese M, et al. Chemical Proteomics and Phenotypic Profiling Identifies the Aryl Hydrocarbon Receptor as a Molecular Target of the Utrophin Modulator Ezutromid. Angewandte Chemie International Edition 2020; 59:2420–8.

76. McGreevy JW, Hakim CH, McIntosh MA, Duan D. Animal models of Duchenne muscular dystrophy: from basic mechanisms to gene therapy. Dis Model Mech 2015; 8:195–213.

77. Sicinski P, Geng Y, Ryder-Cook AS, Barnard EA, Darlison MG, Barnard PJ. The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. Science 1989; 244:1578–80.

78. Bulfield G, Siller WG, Wight PA, Moore KJ. X chromosome-linked muscular dystrophy (mdx) in the mouse. Proc Natl Acad Sci U S A 1984; 81:1189–92.

79. Chamberlain JS, Metzger J, Reyes M, Townsend D, Faulkner JA. Dystrophin-deficient mdx mice display a reduced life span and are susceptible to spontaneous rhabdomyosarcoma. FASEB J 2007; 21:2195–204.

80. Stedman HH, Sweeney HL, Shrager JB, Maguire HC, Panettieri RA, Petrof B, Narusawa M, Leferovich JM, Sladky JT, Kelly AM. The mdx mouse diaphragm reproduces the degenerative changes of Duchenne muscular dystrophy. Nature 1991; 352:536–9.

81. Bostick B, Yue Y, Long C, Duan D. Prevention of dystrophin-deficient cardiomyopathy in twenty-one-month-old carrier mice by mosaic dystrophin expression or complementary dystrophin/utrophin expression. Circ Res 2008; 102:121–30.

82. Lefaucheur JP, Pastoret C, Sebille A. Phenotype of dystrophinopathy in old mdx mice. Anat Rec 1995; 242:70–6.

83. Hakim CH, Grange RW, Duan D. The passive mechanical properties of the extensor digitorum longus muscle are compromised in 2- to 20-mo-old mdx mice. J Appl Physiol (1985) 2011; 110:1656–63.

84. van Westering TLE, Johansson HJ, Hanson B, Coenen-Stass AML, Lomonosova Y, Tanihata J, Motohashi N, Yokota T, Takeda S, Lehtiö J, et al. Mutation-independent Proteomic Signatures of Pathological Progression in Murine Models of Duchenne Muscular Dystrophy. Molecular & Cellular Proteomics 2020; 19:2047–68.

85. Cox GA, Phelps SF, Chapman VM, Chamberlain JS. New mdx mutation disrupts expression of muscle and nonmuscle isoforms of dystrophin. Nat Genet 1993; 4:87–93.

86. Chapman VM, Miller DR, Armstrong D, Caskey CT. Recovery of induced mutations for X chromosome-linked muscular dystrophy in mice. Proc Natl Acad Sci U S A 1989; 86:1292–6.

87. Li D, Yue Y, Duan D. Preservation of Muscle Force in Mdx3cv Mice Correlates with Low-Level Expression of a Near Full-Length Dystrophin Protein. The American Journal of Pathology 2008; 172:1332–41.

88. Danko I, Chapman V, Wolff JA. The Frequency of Revertants in mdx Mouse Genetic Models for Duchenne Muscular Dystrophy. Pediatr Res 1992; 32:128–31.

89. Deconinck AE, Rafael JA, Skinner JA, Brown SC, Potter AC, Metzinger L, Watt DJ, Dickson JG, Tinsley JM, Davies KE. Utrophin-dystrophin-deficient mice as a model for Duchenne muscular dystrophy. Cell 1997; 90:717–27.

90. Grady RM, Teng H, Nichol MC, Cunningham JC, Wilkinson RS, Sanes JR. Skeletal and cardiac myopathies in mice lacking utrophin and dystrophin: a model for Duchenne muscular dystrophy. Cell 1997; 90:729–38.

91. Rooney JE, Welser JV, Dechert MA, Flintoff-Dye NL, Kaufman SJ, Burkin DJ. Severe muscular dystrophy in mice that lack dystrophin and alpha7 integrin. J Cell Sci 2006; 119:2185–95.

92. Guo C, Willem M, Werner A, Raivich G, Emerson M, Neyses L, Mayer U. Absence of alpha 7 integrin in dystrophin-deficient mice causes a myopathy similar to Duchenne muscular dystrophy. Hum Mol Genet 2006; 15:989–98.

93. Banks GB, Combs AC, Odom GL, Bloch RJ, Chamberlain JS. Muscle structure influences utrophin expression in mdx mice. PLoS Genet 2014; 10:e1004431.

94. Megeney LA, Kablar B, Garrett K, Anderson JE, Rudnicki MA. MyoD is required for myogenic stem cell function in adult skeletal muscle. Genes Dev 1996; 10:1173–83.

95. Grady RM, Grange RW, Lau KS, Maimone MM, Nichol MC, Stull JT, Sanes JR. Role for alpha-dystrobrevin in the pathogenesis of dystrophin-dependent muscular dystrophies. Nat Cell Biol 1999; 1:215–20.

96. Han R, Rader EP, Levy JR, Bansal D, Campbell KP. Dystrophin deficiency exacerbates skeletal muscle pathology in dysferlin-null mice. Skelet Muscle 2011; 1:35.

97. Villalta SA, Rinaldi C, Deng B, Liu G, Fedor B, Tidball JG. Interleukin-10 reduces the pathology of mdx muscular dystrophy by deactivating M1 macrophages and modulating macrophage phenotype. Hum Mol Genet 2011; 20:790–805.

98. Isaac C, Wright A, Usas A, Li H, Tang Y, Mu X, Greco N, Dong Q, Vo N, Kang J, et al. Dystrophin and utrophin "double knockout" dystrophic mice exhibit a spectrum of degenerative musculoskeletal abnormalities. J Orthop Res 2013; 31:343–9.

99. Funkquist B, Haraldsson I, Stahre L. Primary progressive muscular dystrophy in the dog. Vet Rec 1980; 106:341–3.

100. Valentine BA, Cooper BJ, Cummings JF, deLahunta A. Progressive muscular dystrophy in a golden retriever dog: light microscope and ultrastructural features at 4 and 8 months. Acta Neuropathol 1986; 71:301–10.

101. Sharp NJ, Kornegay JN, Van Camp SD, Herbstreith MH, Secore SL, Kettle S, Hung WY, Constantinou CD, Dykstra MJ, Roses AD. An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy. Genomics 1992; 13:115–21.

102. Smith BF, Yue Y, Woods PR, Kornegay JN, Shin J-H, Williams RR, Duan D. An intronic LINE-1 element insertion in the dystrophin gene aborts dystrophin expression and results in Duchenne-like muscular dystrophy in the corgi breed. Lab Invest 2011; 91:216–31.

103. Valentine BA, Cooper BJ, de Lahunta A, O'Quinn R, Blue JT. Canine X-linked muscular dystrophy. An animal model of Duchenne muscular dystrophy: clinical studies. J Neurol Sci 1988; 88:69–81.

104. Shimatsu Y, Katagiri K, Furuta T, Nakura M, Tanioka Y, Yuasa K, Tomohiro M, Kornegay JN, Nonaka I, Takeda S. Canine X-linked muscular dystrophy in Japan (CXMDJ). Exp Anim 2003; 52:93–7.

105. Walmsley GL, Arechavala-Gomeza V, Fernandez-Fuente M, Burke MM, Nagel N, Holder A, Stanley R, Chandler K, Marks SL, Muntoni F, et al. A duchenne muscular dystrophy gene hot spot mutation in dystrophin-deficient cavalier king charles spaniels is amenable to exon 51 skipping. PLoS One 2010; 5:e8647.

106. Ricotti V, Ridout DA, Scott E, Quinlivan R, Robb SA, Manzur AY, Muntoni F, NorthStar Clinical Network. Long-term benefits and adverse effects of intermittent versus daily glucocorticoids in boys with Duchenne muscular dystrophy. J Neurol Neurosurg Psychiatry 2013; 84:698–705.

107. Fraser L-A, Adachi JD. Glucocorticoid-Induced Osteoporosis: Treatment Update and Review. Ther Adv Musculoskelet Dis 2009; 1:71–85.

108. Matthews E, Brassington R, Kuntzer T, Jichi F, Manzur AY. Corticosteroids for the treatment of Duchenne muscular dystrophy. Cochrane Database Syst Rev 2016; :CD003725.

109. Nagy S, Hafner P, Schmidt S, Rubino-Nacht D, Schädelin S, Bieri O, Fischer D. Tamoxifen in Duchenne muscular dystrophy (TAMDMD): study protocol for a

multicenter, randomized, placebo-controlled, double-blind phase 3 trial. Trials 2019; 20:637.

110. Koenig M, Beggs AH, Moyer M, Scherpf S, Heindrich K, Bettecken T, Meng G, Müller CR, Lindlöf M, Kaariainen H, et al. The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. Am J Hum Genet 1989; 45:498–506.

111. Malhotra SB, Hart KA, Klamut HJ, Thomas NS, Bodrug SE, Burghes AH, Bobrow M, Harper PS, Thompson MW, Ray PN. Frame-shift deletions in patients with Duchenne and Becker muscular dystrophy. Science 1988; 242:755–9.

112. Monaco AP, Bertelson CJ, Liechti-Gallati S, Moser H, Kunkel LM. An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. Genomics 1988; 2:90–5.

113. England SB, Nicholson LV, Johnson MA, Forrest SM, Love DR, Zubrzycka-Gaarn EE, Bulman DE, Harris JB, Davies KE. Very mild muscular dystrophy associated with the deletion of 46% of dystrophin. Nature 1990; 343:180–2.

114. Matsumura K, Burghes AH, Mora M, Tomé FM, Morandi L, Cornello F, Leturcq F, Jeanpierre M, Kaplan JC, Reinert P. Immunohistochemical analysis of dystrophin-associated proteins in Becker/Duchenne muscular dystrophy with huge in-frame deletions in the NH2-terminal and rod domains of dystrophin. J Clin Invest 1994; 93:99–105.

115. Davies KE, Guiraud S. Micro-dystrophin Genes Bring Hope of an Effective Therapy for Duchenne Muscular Dystrophy. Mol Ther 2019; 27:486–8.

116. Ramos JN, Hollinger K, Bengtsson NE, Allen JM, Hauschka SD, Chamberlain JS. Development of Novel Micro-dystrophins with Enhanced Functionality. Mol Ther 2019; 27:623–35.

117. Sun C, Shen L, Zhang Z, Xie X. Therapeutic Strategies for Duchenne Muscular Dystrophy: An Update. Genes (Basel) 2020; 11.

118. De Angelis FG, Sthandier O, Berarducci B, Toso S, Galluzzi G, Ricci E, Cossu G, Bozzoni I. Chimeric snRNA molecules carrying antisense sequences against the splice junctions of exon 51 of the dystrophin pre-mRNA induce exon skipping and restoration of a dystrophin synthesis in Delta 48-50 DMD cells. Proc Natl Acad Sci U S A 2002; 99:9456–61.

119. Goyenvalle A, Vulin A, Fougerousse F, Leturcq F, Kaplan J-C, Garcia L, Danos O. Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping. Science 2004; 306:1796–9.

120. Becker Y, Hadar J. Parvoviruses [Internet]. In: Becker Y, Hadar J, editors. Molecular Virology: Molecular and Medical Aspects of Disease-Causing Viruses of Man and Animals. Boston, MA: Springer US; 1983 [cited 2021 Jun 1]. page 131–5.Available from: https://doi.org/10.1007/978-1-4613-3906-9_11

121. Penaud-Budloo M, Le Guiner C, Nowrouzi A, Toromanoff A, Chérel Y, Chenuaud P, Schmidt M, von Kalle C, Rolling F, Moullier P, et al. Adeno-associated virus vector genomes persist as episomal chromatin in primate muscle. J Virol 2008; 82:7875–85.

122. Podsakoff G, Wong KK, Chatterjee S. Efficient gene transfer into nondividing cells by adeno-associated virus-based vectors. J Virol 1994; 68:5656–66.

123. Zaiss AK, Muruve DA. Immune responses to adeno-associated virus vectors. Curr Gene Ther 2005; 5:323–31.

124. Daya S, Berns KI. Gene therapy using adeno-associated virus vectors. Clin Microbiol Rev 2008; 21:583–93.

125. Chicoine LG, Rodino-Klapac LR, Shao G, Xu R, Bremer WG, Camboni M, Golden B, Montgomery CL, Shontz K, Heller KN, et al. Vascular delivery of rAAVrh74.MCK.GALGT2 to the gastrocnemius muscle of the rhesus macaque stimulates the expression of dystrophin and laminin α2 surrogates. Mol Ther 2014; 22:713–24.

126. Nguyen HH, Jayasinha V, Xia B, Hoyte K, Martin PT. Overexpression of the cytotoxic T cell GalNAc transferase in skeletal muscle inhibits muscular dystrophy in mdx mice. PNAS 2002; 99:5616–21.

127. Xu R, Singhal N, Serinagaoglu Y, Chandrasekharan K, Joshi M, Bauer JA, Janssen PML, Martin PT. Deletion of Galgt2 (B4Galnt2) reduces muscle growth in response to acute injury and increases muscle inflammation and pathology in dystrophin-deficient mice. Am J Pathol 2015; 185:2668–84.

128. Xia B, Hoyte K, Kammesheidt A, Deerinck T, Ellisman M, Martin PT. Overexpression of the CT GalNAc transferase in skeletal muscle alters myofiber growth, neuromuscular structure, and laminin expression. Dev Biol 2002; 242:58–73.

129. Ohtsuka Y, Udaka K, Yamashiro Y, Yagita H, Okumura K. Dystrophin acts as a transplantation rejection antigen in dystrophin-deficient mice: implication for gene therapy. J Immunol 1998; 160:4635–40.

130. Ebihara S, Guibinga GH, Gilbert R, Nalbantoglu J, Massie B, Karpati G, Petrof BJ. Differential effects of dystrophin and utrophin gene transfer in immunocompetent muscular dystrophy (mdx) mice. Physiol Genomics 2000; 3:133–44.

131. Delalande O, Molza A-E, Morais RDS, Chéron A, Pollet É, Raguenes-Nicol C, Tascon C, Giudice E, Guilbaud M, Nicolas A, et al. Dystrophin's central domain forms a complex filament that becomes disorganized by in-frame deletions. Journal of Biological Chemistry 2018; 293:6637–46.

132. Hakim CH, Wasala NB, Pan X, Kodippili K, Yue Y, Zhang K, Yao G, Haffner B, Duan SX, Ramos J, et al. A Five-Repeat Micro-Dystrophin Gene Ameliorated Dystrophic Phenotype in the Severe DBA/2J-mdx Model of Duchenne Muscular Dystrophy. Mol Ther Methods Clin Dev 2017; 6:216–30.

133. Le Guiner C, Servais L, Montus M, Larcher T, Fraysse B, Moullec S, Allais M, François V, Dutilleul M, Malerba A, et al. Long-term microdystrophin gene therapy is effective in a canine model of Duchenne muscular dystrophy. Nat Commun 2017; 8:16105.

134. Wells DJ, Wells KE, Walsh FS, Davies KE, Goldspink G, Love DR, Chan-Thomas P, Dunckley MG, Piper T, Dickson G. Human dystrophin expression corrects the myopathic phenotype in transgenic mdx mice. Hum Mol Genet 1992; 1:35–40.

135. Genethon announces First Patient dosed in Clinical Trial of Investigational Gene therapy GNT 0004 for Duchenne Muscular Dystrophy > Généthon [Internet]. [cited 2021 Jun 16]; Available from: https://www.genethon.fr/en/2021/04/20/genethon-announces-first-patient-dosed-in-clinical-trial-of-investigational-gene-therapy-gnt-0004-for-duchenne-muscular-dystrophy/

136. Bladen CL, Salgado D, Monges S, Foncuberta ME, Kekou K, Kosma K, Dawkins H, Lamont L, Roy AJ, Chamova T, et al. The TREAT-NMD DMD Global Database: Analysis of More than 7,000 Duchenne Muscular Dystrophy Mutations. Hum Mutat 2015; 36:395–402.

137. Béroud C, Tuffery-Giraud S, Matsuo M, Hamroun D, Humbertclaude V, Monnier N, Moizard M-P, Voelckel M-A, Calemard LM, Boisseau P, et al. Multiexon skipping leading to an artificial DMD protein lacking amino acids from exons 45 through 55 could rescue up to 63% of patients with Duchenne muscular dystrophy. Hum Mutat 2007; 28:196–202.

138. Nakamura A, Yoshida K, Fukushima K, Ueda H, Urasawa N, Koyama J, Yazaki Y, Yazaki M, Sakai T, Haruta S, et al. Follow-up of three patients with a large in-frame deletion of exons 45-55 in the Duchenne muscular dystrophy (DMD) gene. J Clin Neurosci 2008; 15:757–63.

139. Aartsma-Rus A, Straub V, Hemmings R, Haas M, Schlosser-Weber G, Stoyanova-Beninska V, Mercuri E, Muntoni F, Sepodes B, Vroom E, et al. Development of Exon Skipping Therapies for Duchenne Muscular Dystrophy: A Critical Review and a Perspective on the Outstanding Issues. Nucleic Acid Ther 2017; 27:251–9.

140. Aartsma-Rus A, Fokkema I, Verschuuren J, Ginjaar I, van Deutekom J, van Ommen G-J, den Dunnen JT. Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations. Hum Mutat 2009; 30:293–9.

141. Roberts TC, Langer R, Wood MJA. Advances in oligonucleotide drug delivery. Nature Reviews Drug Discovery 2020; 19:673–94.

142. van Vliet L, de Winter CL, van Deutekom JC, van Ommen G-JB, Aartsma-Rus A. Assessment of the feasibility of exon 45–55 multiexon skipping for duchenne muscular dystrophy. BMC Medical Genetics 2008; 9:105.

143. Tennyson CN, Klamut HJ, Worton RG. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nat Genet 1995; 9:184–90.

144. Aoki Y, Yokota T, Nagata T, Nakamura A, Tanihata J, Saito T, Duguez SMR, Nagaraju K, Hoffman EP, Partridge T, et al. Bodywide skipping of exons 45-55 in dystrophic

mdx52 mice by systemic antisense delivery. Proc Natl Acad Sci U S A 2012; 109:13763–8.

145. Mendell JR, Rodino-Klapac LR, Sahenk Z, Roush K, Bird L, Lowes LP, Alfano L, Gomez AM, Lewis S, Kota J, et al. Eteplirsen for the treatment of Duchenne muscular dystrophy. Ann Neurol 2013; 74:637–47.

146. Mendell JR, Goemans N, Lowes LP, Alfano LN, Berry K, Shao J, Kaye EM, Mercuri E, Eteplirsen Study Group and Telethon Foundation DMD Italian Network. Longitudinal effect of eteplirsen versus historical control on ambulation in Duchenne muscular dystrophy. Ann Neurol 2016; 79:257–71.

147. Wu B, Lu P, Benrashid E, Malik S, Ashar J, Doran TJ, Lu QL. Dose-dependent restoration of dystrophin expression in cardiac muscle of dystrophic mice by systemically delivered morpholino. Gene Ther 2010; 17:132–40.

148. Wood MJA. To skip or not to skip: that is the question for duchenne muscular dystrophy. Mol Ther 2013; 21:2131–2.

149. Lu Q-L, Cirak S, Partridge T. What Can We Learn From Clinical Trials of Exon Skipping for DMD? Mol Ther Nucleic Acids 2014; 3:e152.

150. Muntoni F, Fletcher S, Wilton S. Response to "Railroading at the FDA." Nat Biotechnol 2017; 35:207–9.

151. Aartsma-Rus A, Krieg AM. FDA Approves Eteplirsen for Duchenne Muscular Dystrophy: The Next Chapter in the Eteplirsen Saga. Nucleic Acid Ther 2017; 27:1–3.

152. Dowling JJ. Eteplirsen therapy for Duchenne muscular dystrophy: skipping to the front of the line. Nat Rev Neurol 2016; 12:675–6.

153. Charleston JS, Schnell FJ, Dworzak J, Donoghue C, Lewis S, Chen L, Young GD, Milici AJ, Voss J, DeAlwis U, et al. Eteplirsen treatment for Duchenne muscular dystrophy: Exon skipping and dystrophin production. Neurology 2018; 90:e2146–54.

154. Komaki H, Nagata T, Saito T, Masuda S, Takeshita E, Sasaki M, Tachimori H, Nakamura H, Aoki Y, Takeda S. Systemic administration of the antisense oligonucleotide NS-065/NCNP-01 for skipping of exon 53 in patients with Duchenne muscular dystrophy. Sci Transl Med 2018; 10.

155. Godfrey C, Muses S, McClorey G, Wells KE, Coursindel T, Terry RL, Betts C, Hammond S, O'Donovan L, Hildyard J, et al. How much dystrophin is enough: the physiological consequences of different levels of dystrophin in the mdx mouse. Hum Mol Genet 2015; 24:4225–37.

156. Our Pipeline [Internet]. [cited 2021 Jun 1]; Available from: https://www.sarepta.com/products-pipeline/pipeline

157. Betts C, Saleh AF, Arzumanov AA, Hammond SM, Godfrey C, Coursindel T, Gait MJ, Wood MJ. Pip6-PMO, A New Generation of Peptide-oligonucleotide Conjugates With Improved Cardiac Exon Skipping Activity for DMD Treatment. Mol Ther Nucleic Acids 2012; 1:e38.

158. Betts CA, Saleh AF, Carr CA, Hammond SM, Coenen-Stass AML, Godfrey C, McClorey G, Varela MA, Roberts TC, Clarke K, et al. Prevention of exercised induced cardiomyopathy following Pip-PMO treatment in dystrophic mdx mice. Sci Rep 2015; 5:8986.

159. Yin H, Saleh AF, Betts C, Camelliti P, Seow Y, Ashraf S, Arzumanov A, Hammond S, Merritt T, Gait MJ, et al. Pip5 transduction peptides direct high efficiency oligonucleotide-mediated dystrophin exon skipping in heart and phenotypic correction in mdx mice. Mol Ther 2011; 19:1295–303.

160. Roberts TC, Johansson HJ, McClorey G, Godfrey C, Blomberg KEM, Coursindel T, Gait MJ, Smith CIE, Lehtiö J, El Andaloussi S, et al. Multi-level omics analysis in a murine model of dystrophin loss and therapeutic restoration. Hum Mol Genet 2015; 24:6756–68.

161. Yin H, Moulton HM, Seow Y, Boyd C, Boutilier J, Iverson P, Wood MJA. Cell-penetrating peptide-conjugated antisense oligonucleotides restore systemic muscle and cardiac dystrophin expression and function. Hum Mol Genet 2008; 17:3909–18.

162. Wu B, Moulton HM, Iversen PL, Jiang J, Li J, Li J, Spurney CF, Sali A, Guerron AD, Nagaraju K, et al. Effective rescue of dystrophin improves cardiac function in dystrophin-deficient mice by a modified morpholino oligomer. Proceedings of the National Academy of Sciences 2008; 105:14814–9.

163. Peccate C, Mollard A, Le Hir M, Julien L, McClorey G, Jarmin S, Le Heron A, Dickson G, Benkhelifa-Ziyyat S, Piétri-Rouxel F, et al. Antisense pre-treatment increases gene therapy efficacy in dystrophic muscles. Hum Mol Genet 2016; 25:3555–63.

164. Ferrari FK, Samulski T, Shenk T, Samulski RJ. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J Virol 1996; 70:3227–34.

165. Fisher KJ, Gao GP, Weitzman MD, DeMatteo R, Burda JF, Wilson JM. Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. J Virol 1996; 70:520–32.

166. Le Hir M, Goyenvalle A, Peccate C, Précigout G, Davies KE, Voit T, Garcia L, Lorain S. AAV genome loss from dystrophic mouse muscles during AAV-U7 snRNA-mediated exon-skipping therapy. Mol Ther 2013; 21:1551–8.

167. Sussman D, Chadsey M, Fauce S, Engel A, Bruett A, Monnat R, Stoddard BL, Seligman LM. Isolation and characterization of new homing endonuclease specificities at individual target site positions. J Mol Biol 2004; 342:31–41.

168. Rosen LE, Morrison HA, Masri S, Brown MJ, Springstubb B, Sussman D, Stoddard BL, Seligman LM. Homing endonuclease I-CreI derivatives with novel DNA target specificities. Nucleic Acids Res 2006; 34:4791–800.

169. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 2011; 29:143–8.

170. Urnov FD, Miller JC, Lee Y-L, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 2005; 435:646–51.

171. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. Science 2013; 339:823–6.

172. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. Multiplex genome engineering using CRISPR/Cas systems. Science 2013; 339:819–23.

173. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012; 337:816–21.

174. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007; 315:1709–12.

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

176. Mojica FJ, Díez-Villaseñor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol 2000; 36:244–6.

177. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology (Reading) 2009; 155:733–40.

178. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature 2015; 520:186–91.

179. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. Cell 2014; 156:935–49.

180. Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, Li Y, Kurabayashi A, Ishitani R, Zhang F, Nureki O. Crystal Structure of Staphylococcus aureus Cas9. Cell 2015; 162:1113–26.

181. Tabebordbar M, Zhu K, Cheng JKW, Chew WL, Widrick JJ, Yan WX, Maesner C, Wu EY, Xiao R, Ran FA, et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science 2016; 351:407–11.

182. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 2014; 507:62–7.

183. Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 2013; 154:1380–9.

184. Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG, Chandrasegaran S. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol 2001; 21:289–97.

185. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 2011; 471:602–7.

186. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell 2014; 159:647–61.

187. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 2013; 31:827–32.

188. Cho SW, Kim S, Kim JM, Kim J-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nature Biotechnology 2013; 31:230–2.

189. Wu Z, Yang H, Colosi P. Effect of genome size on AAV vector packaging. Mol Ther 2010; 18:80–6.

190. Nelson CE, Hakim CH, Ousterout DG, Thakore PI, Moreb EA, Castellanos Rivera RM, Madhavan S, Pan X, Ran FA, Yan WX, et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science 2016; 351:403–7.

191. Nelson CE, Wu Y, Gemberling MP, Oliver ML, Waller MA, Bohning JD, Robinson-Hamm JN, Bulaklak K, Castellanos Rivera RM, Collier JH, et al. Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. Nat Med 2019; 25:427–32.

192. Hakim CH, Wasala NB, Nelson CE, Wasala LP, Yue Y, Louderman JA, Lessa TB, Dai A, Zhang K, Jenkins GJ, et al. AAV CRISPR editing rescues cardiac and muscle function for 18 months in dystrophic mice. JCI Insight 2018; 3:124297.

193. Ousterout DG, Kabadi AM, Thakore PI, Majoros WH, Reddy TE, Gersbach CA. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. Nature Communications 2015; 6:6244.

194. Long C, Amoasii L, Mireault AA, McAnally JR, Li H, Sanchez-Ortiz E, Bhattacharyya S, Shelton JM, Bassel-Duby R, Olson EN. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. Science 2016; 351:400–3.

195. Bengtsson NE, Hall JK, Odom GL, Phelps MP, Andrus CR, Hawkins RD, Hauschka SD, Chamberlain JR, Chamberlain JS. Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy. Nat Commun 2017; 8:14454.

196. Amoasii L, Long C, Li H, Mireault AA, Shelton JM, Sanchez-Ortiz E, McAnally JR, Bhattacharyya S, Schmidt F, Grimm D, et al. Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. Sci Transl Med 2017; 9.

197. Amoasii L, Hildyard JCW, Li H, Sanchez-Ortiz E, Mireault A, Caballero D, Harron R, Stathopoulou T-R, Massey C, Shelton JM, et al. Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. Science 2018; 362:86–91.

198. Min Y-L, Li H, Rodriguez-Caycedo C, Mireault AA, Huang J, Shelton JM, McAnally JR, Amoasii L, Mammen PPA, Bassel-Duby R, et al. CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells. Sci Adv 2019; 5:eaav4324.

199. Zhang Y, Li H, Min Y-L, Sanchez-Ortiz E, Huang J, Mireault AA, Shelton JM, Kim J, Mammen PPA, Bassel-Duby R, et al. Enhanced CRISPR-Cas9 correction of Duchenne muscular dystrophy in mice by a self-complementary AAV delivery system. Sci Adv 2020; 6:eaay6812.

200. Zhang Y, Long C, Li H, McAnally JR, Baskin KK, Shelton JM, Bassel-Duby R, Olson EN. CRISPR-Cpf1 correction of muscular dystrophy mutations in human cardiomyocytes and mice. Sci Adv 2017; 3:e1602814.

201. Long C, McAnally JR, Shelton JM, Mireault AA, Bassel-Duby R, Olson EN. Prevention of muscular dystrophy in mice by CRISPR/Cas9–mediated editing of germline DNA. Science 2014; 345:1184–8.

202. Heyer W-D, Ehmsen KT, Liu J. Regulation of homologous recombination in eukaryotes. Annu Rev Genet 2010; 44:113–39.

203. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 2014; 3:e04766.

204. Yang D, Scavuzzo MA, Chmielowiec J, Sharp R, Bajic A, Borowiak M. Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. Sci Rep 2016; 6:21264.

205. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 2015; 163:759–71.

206. Kim D, Kim J, Hur JK, Been KW, Yoon S-H, Kim J-S. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat Biotechnol 2016; 34:863–8.

207. Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM, Lopez JM, McCaw ZR, Aryee MJ, Joung JK. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat Biotechnol 2016; 34:869–74.

208. Tóth E, Weinhardt N, Bencsura P, Huszár K, Kulcsár PI, Tálas A, Fodor E, Welker E. Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells. Biol Direct 2016; 11:46.

209. Mort M, Ivanov D, Cooper DN, Chuzhanova NA. A meta-analysis of nonsense mutations causing human genetic disease. Hum Mutat 2008; 29:1037–47.

210. Aartsma-Rus A, Van Deutekom JCT, Fokkema IF, Van Ommen G-JB, Den Dunnen JT. 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 2006; 34:135–44.

211. Cunniff C, Andrews J, Meaney FJ, Mathews KD, Matthews D, Ciafaloni E, Miller TM, Bodensteiner JB, Miller LA, James KA, et al. Mutation analysis in a population-based cohort of boys with Duchenne or Becker muscular dystrophy. J Child Neurol 2009; 24:425–30.

212. Mendell JR, Buzin CH, Feng J, Yan J, Serrano C, Sangani DS, Wall C, Prior TW, Sommer SS. Diagnosis of Duchenne dystrophy by enhanced detection of small mutations. Neurology 2001; 57:645–50.

213. Flanigan KM, Dunn D, von Niederhausern A, Soltanzadeh P, Gappmaier E, Howard MT, Sampson J, Mendell J, Wall C, King W, et al. Mutational Spectrum of DMD Mutations in Dystrophinopathy Patients: Application of Modern Diagnostic Techniques to a Large Cohort. Hum Mutat 2009; 30:1657–66.

214. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 2016; 533:420–4.

215. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 2016; 353.

216. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 2017; 551:464–71.

217. Dianov GL, Hübscher U. Mammalian base excision repair: the forgotten archangel. Nucleic Acids Res 2013; 41:3483–90.

218. Cullot G, Boutin J, Toutain J, Prat F, Pennamen P, Rooryck C, Teichmann M, Rousseau E, Lamrissi-Garcia I, Guyonnet-Duperat V, et al. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. Nat Commun 2019; 10:1136.

219. Ryu S-M, Koo T, Kim K, Lim K, Baek G, Kim S-T, Kim HS, Kim D-E, Lee H, Chung E, et al. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. Nat Biotechnol 2018; 36:536–9.

220. Sun L, Li J, Xiao X. Overcoming adeno-associated virus vector size limitation through viral DNA heterodimerization. Nat Med 2000; 6:599–602.

221. Lai Y, Yue Y, Liu M, Ghosh A, Engelhardt JF, Chamberlain JS, Duan D. Efficient in vivo gene expression by trans-splicing adeno-associated viral vectors. Nat Biotechnol 2005; 23:1435–9.

222. Cox DBT, Gootenberg JS, Abudayyeh OO, Franklin B, Kellner MJ, Joung J, Zhang F. RNA Editing with CRISPR-Cas13. Science 2017; 358:1019–27.

223. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem 2010; 79:321–49.

224. Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, Liu KI, Zhang R, Ramaswami G, Ariyoshi K, et al. Dynamic landscape and regulation of RNA editing in mammals. Nature 2017; 550:249–54.

225. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, Shmakov S, Makarova KS, Semenova E, Minakhin L, et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 2016; 353:aaf5573.

226. Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K, et al. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol Cell 2015; 60:385–97.

227. Bass BL, Weintraub H. An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell 1988; 55:1089–98.

228. Matthews MM, Thomas JM, Zheng Y, Tran K, Phelps KJ, Scott AI, Havel J, Fisher AJ, Beal PA. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. Nat Struct Mol Biol 2016; 23:426–33.

229. Wong SK, Sato S, Lazinski DW. Substrate recognition by ADAR1 and ADAR2. RNA 2001; 7:846–58.

230. Katrekar D, Chen G, Meluzzi D, Ganesh A, Worlikar A, Shih Y-R, Varghese S, Mali P. In vivo RNA editing of point mutations via RNA-guided adenosine deaminases. Nat Methods 2019; 16:239–42.

231. Chen J-F, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon FL, Wang D-Z. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet 2006; 38:228–33.

232. Chen J-F, Tao Y, Li J, Deng Z, Yan Z, Xiao X, Wang D-Z. microRNA-1 and microRNA-206 regulate skeletal muscle satellite cell proliferation and differentiation by repressing Pax7. J Cell Biol 2010; 190:867–79.

233. Roberts TC, Godfrey C, McClorey G, Vader P, Briggs D, Gardiner C, Aoki Y, Sargent I, Morgan JE, Wood MJA. Extracellular microRNAs are dynamic non-vesicular biomarkers of muscle turnover. Nucleic Acids Res 2013; 41:9500–13.

234. Coenen-Stass AML, Betts CA, Lee YF, Mäger I, Turunen MP, El Andaloussi S, Morgan JE, Wood MJA, Roberts TC. Selective release of muscle-specific, extracellular microRNAs during myogenic differentiation. Hum Mol Genet 2016; 25:3960–74.

235. Filipowicz W, Jaskiewicz L, Kolb FA, Pillai RS. Post-transcriptional gene silencing by siRNAs and miRNAs. Curr Opin Struct Biol 2005; 15:331–41.

236. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of Tissue-Specific MicroRNAs from Mouse. Current Biology 2002; 12:735–9.

237. McCarthy JJ. The MyomiR Network in Skeletal Muscle Plasticity. Exercise and Sport Sciences Reviews 2011; 39:150–4.

238. Mizuno H, Nakamura A, Aoki Y, Ito N, Kishi S, Yamamoto K, Sekiguchi M, Takeda S, Hashido K. Identification of Muscle-Specific MicroRNAs in Serum of Muscular Dystrophy Animal Models: Promising Novel Blood-Based Markers for Muscular Dystrophy. PLOS ONE 2011; 6:e18388.

239. Roberts TC, Blomberg KEM, McClorey G, El Andaloussi S, Godfrey C, Betts C, Coursindel T, Gait MJ, Smith CIE, Wood MJA. Expression analysis in multiple muscle groups and serum reveals complexity in the microRNA transcriptome of the mdx mouse with implications for therapy. Mol Ther Nucleic Acids 2012; 1:e39.

240. Vignier N, Amor F, Fogel P, Duvallet A, Poupiot J, Charrier S, Arock M, Montus M, Nelson I, Richard I, et al. Distinctive Serum miRNA Profile in Mouse Models of Striated Muscular Pathologies. PLOS ONE 2013; 8:e55281.

241. Goyenvalle A, Babbs A, Wright J, Wilkins V, Powell D, Garcia L, Davies KE. Rescue of severely affected dystrophin/utrophin-deficient mice through scAAV-U7snRNA-mediated exon skipping. Hum Mol Genet 2012; 21:2559–71.

242. Jeanson-Leh L, Lameth J, Krimi S, Buisset J, Amor F, Le Guiner C, Barthélémy I, Servais L, Blot S, Voit T, et al. Serum Profiling Identifies Novel Muscle miRNA and Cardiomyopathy-Related miRNA Biomarkers in Golden Retriever Muscular Dystrophy Dogs and Duchenne Muscular Dystrophy Patients. The American Journal of Pathology 2014; 184:2885–98.

243. Coenen-Stass AML, Sork H, Gatto S, Godfrey C, Bhomra A, Krjutškov K, Hart JR, Westholm JO, O'Donovan L, Roos A, et al. Comprehensive RNA-Sequencing Analysis in Serum and Muscle Reveals Novel Small RNA Signatures with Biomarker Potential for DMD. Mol Ther Nucleic Acids 2018; 13:1–15.

244. Coenen-Stass AML, Wood MJA, Roberts TC. Biomarker Potential of Extracellular miRNAs in Duchenne Muscular Dystrophy. Trends in Molecular Medicine 2017; 23:989–1001.

245. Duguez S, Duddy W, Johnston H, Lainé J, Le Bihan MC, Brown KJ, Bigot A, Hathout Y, Butler-Browne G, Partridge T. Dystrophin deficiency leads to disturbance of LAMP1-vesicle-associated protein secretion. Cell Mol Life Sci 2013; 70:2159–74.

246. Cacchiarelli D, Legnini I, Martone J, Cazzella V, D'Amico A, Bertini E, Bozzoni I. miRNAs as serum biomarkers for Duchenne muscular dystrophy. EMBO Molecular Medicine 2011; 3:258–65.

247. Zaharieva IT, Calissano M, Scoto M, Preston M, Cirak S, Feng L, Collins J, Kole R, Guglieri M, Straub V, et al. Dystromirs as serum biomarkers for monitoring the disease severity in Duchenne muscular Dystrophy. PLoS One 2013; 8:e80263.

248. Li X, Li Y, Zhao L, Zhang D, Yao X, Zhang H, Wang Y-C, Wang X-Y, Xia H, Yan J, et al. Circulating Muscle-specific miRNAs in Duchenne Muscular Dystrophy Patients. Mol Ther Nucleic Acids 2014; 3:e177.

249. Liu N, Williams AH, Maxeiner JM, Bezprozvannaya S, Shelton JM, Richardson JA, Bassel-Duby R, Olson EN. microRNA-206 promotes skeletal muscle regeneration and delays progression of Duchenne muscular dystrophy in mice. J Clin Invest 2012; 122:2054–65.

250. McCarthy JJ. MicroRNA-206: the skeletal muscle-specific myomiR. Biochim Biophys Acta 2008; 1779:682–91.

251. Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ, Lötvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nature Cell Biology 2007; 9:654–9.

252. Hunter MP, Ismail N, Zhang X, Aguda BD, Lee EJ, Yu L, Xiao T, Schafer J, Lee M-LT, Schmittgen TD, et al. Detection of microRNA expression in human peripheral blood microvesicles. PLoS One 2008; 3:e3694.

253. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. Nature Cell Biology 2011; 13:423–33.

254. Wagner J, Riwanto M, Besler C, Knau A, Fichtlscherer S, Röxe T, Zeiher AM, Landmesser U, Dimmeler S. Characterization of levels and cellular transfer of circulating lipoprotein-bound microRNAs. Arterioscler Thromb Vasc Biol 2013; 33:1392–400.

255. Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, Mitchell PS, Bennett CF, Pogosova-Agadjanyan EL, Stirewalt DL, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc Natl Acad Sci U S A 2011; 108:5003–8.

256. Wang K, Zhang S, Weber J, Baxter D, Galas DJ. Export of microRNAs and microRNA-protective protein by mammalian cells. Nucleic Acids Res 2010; 38:7248–59.

257. Sun C, Serra C, Lee G, Wagner KR. Stem cell-based therapies for Duchenne muscular dystrophy. Exp Neurol 2020; 323:113086.

258. Alvarez-Erviti L, Seow Y, Yin H, Betts C, Lakhal S, Wood MJA. Delivery of siRNA to the mouse brain by systemic injection of targeted exosomes. Nat Biotechnol 2011; 29:341–5.

259. EL Andaloussi S, Mäger I, Breakefield XO, Wood MJA. Extracellular vesicles: biology and emerging therapeutic opportunities. Nat Rev Drug Discov 2013; 12:347–57.

260. Wolf P. The nature and significance of platelet products in human plasma. Br J Haematol 1967; 13:269–88.

261. van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of extracellular vesicles. Nat Rev Mol Cell Biol 2018; 19:213–28.

262. Al-Nedawi K, Meehan B, Micallef J, Lhotak V, May L, Guha A, Rak J. Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. Nat Cell Biol 2008; 10:619–24.

263. Barros FM, Carneiro F, Machado JC, Melo SA. Exosomes and Immune Response in Cancer: Friends or Foes? Front Immunol 2018; 9:730.

264. Mashouri L, Yousefi H, Aref AR, Ahadi AM, Molaei F, Alahari SK. Exosomes: composition, biogenesis, and mechanisms in cancer metastasis and drug resistance. Mol Cancer 2019; 18:75.

265. Santavanond JP, Rutter SF, Atkin-Smith GK, Poon IKH. Apoptotic Bodies: Mechanism of Formation, Isolation and Functional Relevance. Subcell Biochem 2021; 97:61–88.

266. Johnstone RM, Adam M, Hammond JR, Orr L, Turbide C. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). J Biol Chem 1987; 262:9412–20.

267. Théry C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R, Antoniou A, Arab T, Archer F, Atkin-Smith GK, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. Journal of Extracellular Vesicles 2018; 7:1535750.

268. EV-TRACK Consortium, Van Deun J, Mestdagh P, Agostinis P, Akay Ö, Anand S, Anckaert J, Martinez ZA, Baetens T, Beghein E, et al. EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research. Nat Methods 2017; 14:228–32.

269. Wenzel EM, Schultz SW, Schink KO, Pedersen NM, Nähse V, Carlson A, Brech A, Stenmark H, Raiborg C. Concerted ESCRT and clathrin recruitment waves define the timing and morphology of intraluminal vesicle formation. Nat Commun 2018; 9:2932.

270. Yáñez-Mó M, Siljander PR-M, Andreu Z, Zavec AB, Borràs FE, Buzas EI, Buzas K, Casal E, Cappello F, Carvalho J, et al. Biological properties of extracellular vesicles and their physiological functions. J Extracell Vesicles 2015; 4:27066.

271. Jahn R, Scheller RH. SNAREs--engines for membrane fusion. Nat Rev Mol Cell Biol 2006; 7:631–43.

272. Trajkovic K, Hsu C, Chiantia S, Rajendran L, Wenzel D, Wieland F, Schwille P, Brügger B, Simons M. Ceramide triggers budding of exosome vesicles into multivesicular endosomes. Science 2008; 319:1244–7.

273. Hoshino D, Kirkbride KC, Costello K, Clark ES, Sinha S, Grega-Larson N, Tyska MJ, Weaver AM. Exosome Secretion Is Enhanced by Invadopodia and Drives Invasive Behavior. Cell Reports 2013; 5:1159–68.

274. Chairoungdua A, Smith DL, Pochard P, Hull M, Caplan MJ. Exosome release of β-catenin: a novel mechanism that antagonizes Wnt signaling. J Cell Biol 2010; 190:1079–91.

275. Cheng Q, Li X, Wang Y, Dong M, Zhan F-H, Liu J. The ceramide pathway is involved in the survival, apoptosis and exosome functions of human multiple myeloma cells in vitro. Acta Pharmacol Sin 2018; 39:561–8.

276. Jankovičová J, Sečová P, Michalková K, Antalíková J. Tetraspanins, More than Markers of Extracellular Vesicles in Reproduction. Int J Mol Sci 2020; 21.

277. Hessvik NP, Llorente A. Current knowledge on exosome biogenesis and release. Cell Mol Life Sci 2018; 75:193–208.

278. Mittelbrunn M, Vicente Manzanares M, Sánchez-Madrid F. Organizing polarized delivery of exosomes at synapses. Traffic 2015; 16:327–37.

279. Savina A, Vidal M, Colombo MI. The exosome pathway in K562 cells is regulated by Rab11. J Cell Sci 2002; 115:2505–15.

280. Ostrowski M, Carmo NB, Krumeich S, Fanget I, Raposo G, Savina A, Moita CF, Schauer K, Hume AN, Freitas RP, et al. Rab27a and Rab27b control different steps of the exosome secretion pathway. Nature Cell Biology 2010; 12:19–30.

281. Baietti MF, Zhang Z, Mortier E, Melchior A, Degeest G, Geeraerts A, Ivarsson Y, Depoortere F, Coomans C, Vermeiren E, et al. Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. Nat Cell Biol 2012; 14:677–85.

282. Bonifacino JS, Glick BS. The mechanisms of vesicle budding and fusion. Cell 2004; 116:153–66.

283. Hessvik NP, Øverbye A, Brech A, Torgersen ML, Jakobsen IS, Sandvig K, Llorente A. PIKfyve inhibition increases exosome release and induces secretory autophagy. Cell Mol Life Sci 2016; 73:4717–37.

284. Dias MVS, Teixeira BL, Rodrigues BR, Sinigaglia-Coimbra R, Porto-Carreiro I, Roffé M, Hajj GNM, Martins VR. PRNP/prion protein regulates the secretion of exosomes modulating CAV1/caveolin-1-suppressed autophagy. Autophagy 2016; 12:2113–28.

285. Hoshino A, Costa-Silva B, Shen T-L, Rodrigues G, Hashimoto A, Tesic Mark M, Molina H, Kohsaka S, Di Giannatale A, Ceder S, et al. Tumour exosome integrins determine organotropic metastasis. Nature 2015; 527:329–35.

286. Peinado H, Alečković M, Lavotshkin S, Matei I, Costa-Silva B, Moreno-Bueno G, Hergueta-Redondo M, Williams C, García-Santos G, Ghajar CM, et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. Nat Med 2012; 18:883–91.

287. Budnik V, Ruiz-Cañada C, Wendler F. Extracellular vesicles round off communication in the nervous system. Nat Rev Neurosci 2016; 17:160–72.

288. Rajendran L, Honsho M, Zahn TR, Keller P, Geiger KD, Verkade P, Simons K. Alzheimer's disease beta-amyloid peptides are released in association with exosomes. Proc Natl Acad Sci U S A 2006; 103:11172–7.

289. Joshi P, Benussi L, Furlan R, Ghidoni R, Verderio C. Extracellular vesicles in Alzheimer's disease: friends or foes? Focus on aβ-vesicle interaction. Int J Mol Sci 2015; 16:4800–13.

290. Yuyama K, Sun H, Sakai S, Mitsutake S, Okada M, Tahara H, Furukawa J-I, Fujitani N, Shinohara Y, Igarashi Y. Decreased amyloid-β pathologies by intracerebral loading of glycosphingolipid-enriched exosomes in Alzheimer model mice. J Biol Chem 2014; 289:24488–98.

291. Agosta F, Dalla Libera D, Spinelli EG, Finardi A, Canu E, Bergami A, Bocchio Chiavetto L, Baronio M, Comi G, Martino G, et al. Myeloid microvesicles in cerebrospinal fluid are associated with myelin damage and neuronal loss in mild cognitive impairment and Alzheimer disease. Ann Neurol 2014; 76:813–25.

292. Haney MJ, Klyachko NL, Zhao Y, Gupta R, Plotnikova EG, He Z, Patel T, Piroyan A, Sokolsky M, Kabanov AV, et al. Exosomes as drug delivery vehicles for Parkinson's disease therapy. J Control Release 2015; 207:18–30.

293. Gagliardi D, Bresolin N, Comi GP, Corti S. Extracellular vesicles and amyotrophic lateral sclerosis: from misfolded protein vehicles to promising clinical biomarkers. Cell Mol Life Sci 2021; 78:561–72.

294. Catapano F, Scaglioni D, Maresh K, Ala P, Domingos J, Selby V, Ricotti V, Phillips L, Servais L, Seferian A, et al. Novel free-circulating and extracellular vesicle-derived miRNAs dysregulated in Duchenne muscular dystrophy. Epigenomics 2020; 12:1899–915.

295. Roberts TC, Ezzat K, El Andaloussi S, Weinberg MS. Synthetic SiRNA Delivery: Progress and Prospects. Methods Mol Biol 2016; 1364:291–310.

296. Witwer KW, Wolfram J. Extracellular vesicles versus synthetic nanoparticles for drug delivery. Nat Rev Mater 2021; 6:103–6.

297. Conceição M, Forcina L, Wiklander OPB, Gupta D, Nordin JZ, Vrellaku B, McClorey G, Mäger I, Görgens A, Lundin P, et al. Engineered extracellular vesicle decoy receptor-mediated modulation of the IL6 trans-signalling pathway in muscle. Biomaterials 2021; 266:120435.

298. Saari H, Turunen T, Lõhmus A, Turunen M, Jalasvuori M, Butcher SJ, Ylä-Herttuala S, Viitala T, Cerullo V, Siljander PRM, et al. Extracellular vesicles provide a capsid-free vector for oncolytic adenoviral DNA delivery. J Extracell Vesicles 2020; 9:1747206.

299. Gee P, Lung MSY, Okuzaki Y, Sasakawa N, Iguchi T, Makita Y, Hozumi H, Miura Y, Yang LF, Iwasaki M, et al. Extracellular nanovesicles for packaging of CRISPR-Cas9 protein and sgRNA to induce therapeutic exon skipping. Nat Commun 2020; 11:1334.

300. Heldring N, Mäger I, Wood MJA, Le Blanc K, Andaloussi SEL. Therapeutic Potential of Multipotent Mesenchymal Stromal Cells and Their Extracellular Vesicles. Human Gene Therapy 2015; 26:506–17.

301. Giebel B, Kordelas L, Börger V. Clinical potential of mesenchymal stem/stromal cell-derived extracellular vesicles. Stem Cell Investig 2017; 4:84.

302. Elahi FM, Farwell DG, Nolta JA, Anderson JD. Preclinical translation of exosomes derived from mesenchymal stem/stromal cells. Stem Cells 2020; 38:15–21.

303. Timmers L, Lim SK, Arslan F, Armstrong JS, Hoefer IE, Doevendans PA, Piek JJ, El Oakley RM, Choo A, Lee CN, et al. Reduction of myocardial infarct size by human mesenchymal stem cell conditioned medium. Stem Cell Research 2008; 1:129–37.

304. Lai RC, Arslan F, Lee MM, Sze NSK, Choo A, Chen TS, Salto-Tellez M, Timmers L, Lee CN, El Oakley RM, et al. Exosome secreted by MSC reduces myocardial ischemia/reperfusion injury. Stem Cell Research 2010; 4:214–22.

305. van Koppen A, Joles JA, van Balkom BWM, Lim SK, de Kleijn D, Giles RH, Verhaar MC. Human embryonic mesenchymal stem cell-derived conditioned medium rescues kidney function in rats with established chronic kidney disease. PLoS One 2012; 7:e38746.

306. Forterre A, Jalabert A, Berger E, Baudet M, Chikh K, Errazuriz E, De Larichaudy J, Chanon S, Weiss-Gayet M, Hesse A-M, et al. Proteomic analysis of C2C12 myoblast and myotube exosome-like vesicles: a new paradigm for myoblast-myotube cross talk? PLoS One 2014; 9:e84153.

307. Fry CS, Kirby TJ, Kosmac K, McCarthy JJ, Peterson CA. Myogenic Progenitor Cells Control Extracellular Matrix Production by Fibroblasts during Skeletal Muscle Hypertrophy. Cell Stem Cell 2017; 20:56–69.

308. Choi JS, Yoon HI, Lee KS, Choi YC, Yang SH, Kim I-S, Cho YW. Exosomes from differentiating human skeletal muscle cells trigger myogenesis of stem cells and provide biochemical cues for skeletal muscle regeneration. J Control Release 2016; 222:107–15.

309. Nakamura Y, Miyaki S, Ishitobi H, Matsuyama S, Nakasa T, Kamei N, Akimoto T, Higashi Y, Ochi M. Mesenchymal-stem-cell-derived exosomes accelerate skeletal muscle regeneration. FEBS Letters 2015; 589:1257–65.

310. Fais S, O'Driscoll L, Borras FE, Buzas E, Camussi G, Cappello F, Carvalho J, Cordeiro da Silva A, Del Portillo H, El Andaloussi S, et al. Evidence-Based Clinical Use of Nanoscale Extracellular Vesicles in Nanomedicine. ACS Nano 2016; 10:3886–99.

311. Forterre A, Jalabert A, Chikh K, Pesenti S, Euthine V, Granjon A, Errazuriz E, Lefai E, Vidal H, Rome S. Myotube-derived exosomal miRNAs downregulate Sirtuin1 in myoblasts during muscle cell differentiation. Cell Cycle 2014; 13:78–89.

312. Mytidou C, Koutsoulidou A, Katsioloudi A, Prokopi M, Kapnisis K, Michailidou K, Anayiotos A, Phylactou LA. Muscle-derived exosomes encapsulate myomiRs and are involved in local skeletal muscle tissue communication. The FASEB Journal 2021; 35:e21279.

313. Kozak M. The scanning model for translation: an update. Journal of Cell Biology 1989; 108:229–41.

314. Kozak M. Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. Nucleic Acids Res 1984; 12:3873–93.

315. Kozak M. How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell 1978; 15:1109–23.

316. Donnelly N, Gorman AM, Gupta S, Samali A. The eIF2α kinases: their structures and functions. Cell Mol Life Sci 2013; 70:3493–511.

317. Kozak M. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. The EMBO Journal 1997; 16:2482–92.

318. Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 1986; 44:283–92.

319. Pisareva VP, Skabkin MA, Hellen CUT, Pestova TV, Pisarev AV. Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. EMBO J 2011; 30:1804–17.

320. Alkalaeva EZ, Pisarev AV, Frolova LY, Kisselev LL, Pestova TV. In Vitro Reconstitution of Eukaryotic Translation Reveals Cooperativity between Release Factors eRF1 and eRF3. Cell 2006; 125:1125–36.

321. Frolova LY, Tsivkovskii RY, Sivolobova GF, Oparina NY, Serpinsky OI, Blinov VM, Tatkov SI, Kisselev LL. Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. RNA 1999; 5:1014–20.

322. Pisarev AV, Skabkin MA, Pisareva VP, Skabkina OV, Rakotondrafara AM, Hentze MW, Hellen CUT, Pestova TV. The Role of ABCE1 in Eukaryotic Posttermination Ribosomal Recycling. Molecular Cell 2010; 37:196–210.

323. Santiago TC, Purvis IJ, Bettany AJ, Brown AJ. The relationship between mRNA stability and length in Saccharomyces cerevisiae. Nucleic Acids Res 1986; 14:8347–60.

324. Matsui M, Yachie N, Okada Y, Saito R, Tomita M. Bioinformatic analysis of post-transcriptional regulation by uORF in human and mouse. FEBS Lett 2007; 581:4184–8.

325. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 2012; 13:227–32.

326. Ly T, Ahmad Y, Shlien A, Soroka D, Mills A, Emanuele MJ, Stratton MR, Lamond AI. A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. Elife 2014; 3:e01630.

327. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 2010; 466:835–40.

328. Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol 2004; 2:e391.

329. Bazak L, Haviv A, Barak M, Jacob-Hirsch J, Deng P, Zhang R, Isaacs FJ, Rechavi G, Li JB, Eisenberg E, et al. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res 2014; 24:365–76.

330. Blow M, Futreal PA, Wooster R, Stratton MR. A survey of RNA editing in human brain. Genome Res 2004; 14:2379–87.

331. Kim DDY, Kim TTY, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A. Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res 2004; 14:1719–25.

332. Nishikura K, Sakurai M, Ariyoshi K, Ota H. Antagonistic and stimulative roles of ADAR1 in RNA silencing. RNA Biol 2013; 10:1240–7.

333. Ota H, Sakurai M, Gupta R, Valente L, Wulff B-E, Ariyoshi K, Iizasa H, Davuluri RV, Nishikura K. ADAR1 forms a complex with Dicer to promote microRNA processing and RNA-induced gene silencing. Cell 2013; 153:575–89.

334. Rodnina MV, Korniy N, Klimova M, Karki P, Peng B-Z, Senyushkina T, Belardinelli R, Maracci C, Wohlgemuth I, Samatova E, et al. Translational recoding: canonical translation mechanisms reinterpreted. Nucleic Acids Res 2020; 48:1056–67.

335. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, et al. The complete sequence of a human genome. bioRxiv 2021; :2021.05.26.445798.

336. Couso J-P, Patraquim P. Classification and function of small open reading frames. Nat Rev Mol Cell Biol 2017; 18:575–89.

337. Basrai MA, Hieter P, Boeke JD. Small open reading frames: beautiful needles in the haystack. Genome Res 1997; 7:768–71.

338. Anderson DM, Anderson KM, Chang C-L, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 2015; 160:595–606.

339. Jackson R, Kroehling L, Khitun A, Bailis W, Jarret A, York AG, Khan OM, Brewer JR, Skadow MH, Duizer C, et al. The translation of non-canonical open reading frames controls mucosal immunity. Nature 2018; 564:434–8.

340. Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, Reese AL, McAnally JR, Chen X, Kavalali ET, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science 2016; 351:271–5.

341. Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, Hashimoto Y, Kobayashi S, Payre F, Kageyama Y. Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. Science 2010; 329:336–9.

342. Magny EG, Pueyo JI, Pearl FMG, Cespedes MA, Niven JE, Bishop SA, Couso JP. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. Science 2013; 341:1116–20.

343. D'Lima NG, Ma J, Winkler L, Chu Q, Loh KH, Corpuz EO, Budnik BA, Lykke-Andersen J, Saghatelian A, Slavoff SA. A human microprotein that interacts with the mRNA decapping complex. Nat Chem Biol 2017; 13:174–80.

344. Stein CS, Jadiya P, Zhang X, McLendon JM, Abouassaly GM, Witmer NH, Anderson EJ, Elrod JW, Boudreau RL. Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. Cell Rep 2018; 23:3710-3720.e8.

345. Makarewich CA, Baskin KK, Munir AZ, Bezprozvannaya S, Sharma G, Khemtong C, Shah AM, McAnally JR, Malloy CR, Szweda LI, et al. MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid β-Oxidation. Cell Rep 2018; 23:3701–9.

346. Matsumoto A, Pasut A, Matsumoto M, Yamashita R, Fung J, Monteleone E, Saghatelian A, Nakayama KI, Clohessy JG, Pandolfi PP. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. Nature 2017; 541:228–32.

347. Slavoff SA, Heo J, Budnik BA, Hanakahi LA, Saghatelian A. A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining. J Biol Chem 2014; 289:10950–7.

348. Bi P, Ramirez-Martinez A, Li H, Cannavino J, McAnally JR, Shelton JM, Sánchez-Ortiz E, Bassel-Duby R, Olson EN. Control of muscle formation by the fusogenic micropeptide myomixer. Science 2017; 356:323–7.

349. Zhang Q, Vashisht AA, O'Rourke J, Corbel SY, Moran R, Romero A, Miraglia L, Zhang J, Durrant E, Schmedt C, et al. The microprotein Minion controls cell fusion and muscle formation. Nature Communications 2017; 8:15664.

350. Pauli A, Norris ML, Valen E, Chew G-L, Gagnon JA, Zimmerman S, Mitchell A, Ma J, Dubrulle J, Reyon D, et al. Toddler: an embryonic signal that promotes cell movement via Apelin receptors. Science 2014; 343:1248636.

351. Odermatt A, Taschner PE, Scherer SW, Beatty B, Khanna VK, Cornblath DR, Chaudhry V, Yee WC, Schrank B, Karpati G, et al. Characterization of the gene encoding human sarcolipin (SLN), a proteolipid associated with SERCA1: absence of structural mutations in five patients with Brody disease. Genomics 1997; 45:541–53.

352. Martinez TF, Chu Q, Donaldson C, Tan D, Shokhirev MN, Saghatelian A. Accurate annotation of human protein-coding small open reading frames. Nature Chemical Biology 2020; 16:458–68.

353. Mackowiak SD, Zauber H, Bielow C, Thiel D, Kutz K, Calviello L, Mastrobuoni G, Rajewsky N, Kempa S, Selbach M, et al. Extensive identification and analysis of conserved small ORFs in animals. Genome Biol 2015; 16:179.

354. Joliot A, Pernelle C, Deagostini-Bazin H, Prochiantz A. Antennapedia homeobox peptide regulates neural morphogenesis. Proc Natl Acad Sci U S A 1991; 88:1864–8.

355. Joliot A, Prochiantz A. Transduction peptides: from technology to physiology. Nat Cell Biol 2004; 6:189–96.

356. Sarepta Therapeutics Announces Positive Clinical Results from MOMENTUM, a Phase 2 Clinical Trial of SRP-5051 in Patients with Duchenne Muscular Dystrophy Amenable to Skipping Exon 51 | Sarepta Therapeutics, Inc. [Internet]. [cited 2021 Jun 12];

Available from: https://investorrelations.sarepta.com/news-releases/news-release-details/sarepta-therapeutics-announces-positive-clinical-results

357. Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM. Long non-coding RNAs as a source of new peptides. Elife 2014; 3:e03523.

358. Huang J-Z, Chen M, Chen D, Gao X-C, Zhu S, Huang H, Hu M, Zhu H, Yan G-R. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. Mol Cell 2017; 68:171-184.e6.

359. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proceedings of the National Academy of Sciences 2009; 106:7507–12.

360. Johnstone TG, Bazzini AA, Giraldez AJ. Upstream ORFs are prevalent translational repressors in vertebrates. The EMBO Journal 2016; 35:706–23.

361. Iacono M, Mignone F, Pesole G. uAUG and uORFs in human and rodent 5'untranslated mRNAs. Gene 2005; 349:97–105.

362. Wethmar K. The regulatory potential of upstream open reading frames in eukaryotic gene expression. Wiley Interdiscip Rev RNA 2014; 5:765–78.

363. Chen J, Brunner A-D, Cogan JZ, Nuñez JK, Fields AP, Adamson B, Itzhak DN, Li JY, Mann M, Leonetti MD, et al. Pervasive functional translation of noncanonical human open reading frames. Science 2020; 367:1140–6.

364. Suzuki Y, Ishihara D, Sasaki M, Nakagawa H, Hata H, Tsunoda T, Watanabe M, Komatsu T, Ota T, Isogai T, et al. Statistical analysis of the 5' untranslated region of human mRNA using "Oligo-Capped" cDNA libraries. Genomics 2000; 64:286–97.

365. Rogozin IB, Kochetov AV, Kondrashov FA, Koonin EV, Milanesi L. Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a "weak" context of the start codon. Bioinformatics 2001; 17:890–900.

366. Mignone F, Gissi C, Liuni S, Pesole G. Untranslated regions of mRNAs. Genome Biol 2002; 3:reviews0004.1.

367. Morris DR, Geballe AP. Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol 2000; 20:8635–42.

368. Kochetov AV, Ischenko IV, Vorobiev DG, Kel AE, Babenko VN, Kisselev LL, Kolchanov NA. Eukaryotic mRNAs encoding abundant and scarce proteins are statistically dissimilar in many structural features. FEBS Lett 1998; 440:351–5.

369. Meijer HA, Thomas AAM. Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. Biochem J 2002; 367:1–11.

370. Meijer HA, Thomas AAM. Ribosomes stalling on uORF1 in the Xenopus Cx41 5' UTR inhibit downstream translation initiation. Nucleic Acids Res 2003; 31:3174–84.

371. Calkhoven CF, Müller C, Leutz A. Translational control of C/EBPα and C/EBPβ isoform expression. Genes Dev 2000; 14:1920–32.

372. Suzuki Y, Holmes JB, Cerritelli SM, Sakhuja K, Minczuk M, Holt IJ, Crouch RJ. An upstream open reading frame and the context of the two AUG codons affect the abundance of mitochondrial and nuclear RNase H1. Mol Cell Biol 2010; 30:5123–34.

373. Timchenko NA, Welm AL, Lu X, Timchenko LT. CUG repeat binding protein (CUGBP1) interacts with the 5' region of C/EBPbeta mRNA and regulates translation of C/EBPbeta isoforms. Nucleic Acids Res 1999; 27:4517–25.

374. Timchenko LT, Miller JW, Timchenko NA, DeVore DR, Datar KV, Lin L, Roberts R, Caskey CT, Swanson MS. Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. Nucleic Acids Res 1996; 24:4407–14.

375. Wethmar K, Smink JJ, Leutz A. Upstream open reading frames: Molecular switches in (patho)physiology. Bioessays 2010; 32:885–93.

376. Kozak M. Constraints on reinitiation of translation in mammals. Nucleic Acids Res 2001; 29:5226–32.

377. Rajkowitsch L, Vilela C, Berthelot K, Ramirez CV, McCarthy JEG. Reinitiation and recycling are distinct processes occurring downstream of translation termination in yeast. J Mol Biol 2004; 335:71–85.

378. Child SJ, Miller MK, Geballe AP. Translational control by an upstream open reading frame in the HER-2/neu transcript. J Biol Chem 1999; 274:24335–41.

379. Zhou W, Song W. Leaky scanning and reinitiation regulate BACE1 gene expression. Mol Cell Biol 2006; 26:3353–64.

380. Wang X-Q, Rothnagel JA. 5'-untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. Nucleic Acids Res 2004; 32:1382–91.

381. Wethmar K, Schulz J, Muro EM, Talyan S, Andrade-Navarro MA, Leutz A. Comprehensive translational control of tyrosine kinase expression by upstream open reading frames. Oncogene 2016; 35:1736–42.

382. Barbosa C, Peixeiro I, Romão L. Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. PLOS Genetics 2013; 9:e1003529.

383. Kozak M. An analysis of vertebrate mRNA sequences: intimations of translational control. Journal of Cell Biology 1991; 115:887–903.

384. Schulz J, Mah N, Neuenschwander M, Kischka T, Ratei R, Schlag PM, Castaños-Vélez E, Fichtner I, Tunn P-U, Denkert C, et al. Loss-of-function uORF mutations in human malignancies. Sci Rep 2018; 8:2395.

385. Spevak CC, Park E-H, Geballe AP, Pelletier J, Sachs MS. her-2 upstream open reading frame effects on the use of downstream initiation codons. Biochem Biophys Res Commun 2006; 350:834–41.

386. McGillivray P, Ault R, Pawashe M, Kitchen R, Balasubramanian S, Gerstein M. A comprehensive catalog of predicted functional upstream open reading frames in humans. Nucleic Acids Res 2018; 46:3326–38.

387. Bach J, Endler G, Winkelmann BR, Boehm BO, Maerz W, Mannhalter C, Hellstern P. Coagulation factor XII (FXII) activity, activated FXII, distribution of FXII C46T gene polymorphism and coronary risk. J Thromb Haemost 2008; 6:291–6.

388. Bertina RM, Poort SR, Vos HL, Rosendaal FR. The 46C-->T polymorphism in the factor XII gene (F12) and the risk of venous thrombosis. J Thromb Haemost 2005; 3:597–9.

389. Endler G, Exner M, Mannhalter C, Meier S, Ruzicka K, Handler S, Panzer S, Wagner O, Quehenberger P. A common C-->T polymorphism at nt 46 in the promoter region of coagulation factor XII is associated with decreased factor XII activity. Thromb Res 2001; 101:255–60.

390. Kanaji T, Okamura T, Osaki K, Kuroiwa M, Shimoda K, Hamasaki N, Niho Y. A common genetic polymorphism (46 C to T substitution) in the 5'-untranslated region of the coagulation factor XII gene is associated with low translation efficiency and decrease in plasma factor XII level. Blood 1998; 91:2010–4.

391. Tirado I, Soria JM, Mateo J, Oliver A, Souto JC, Santamaria A, Felices R, Borrell M, Fontcuberta J. Association after linkage analysis indicates that homozygosity for the 46C-->T polymorphism in the F12 gene is a genetic risk factor for venous thrombosis. Thromb Haemost 2004; 91:899–904.

392. Liang X, Shen W, Sun H, Migawa MT, Vickers TA, Crooke ST. Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. Nature Biotechnology 2016; 34:875–80.

393. Liang X, Sun H, Shen W, Wang S, Yao J, Migawa MT, Bui H-H, Damle SS, Riney S, Graham MJ, et al. Antisense oligonucleotides targeting translation inhibitory elements in 5′ UTRs can selectively increase protein levels. Nucleic Acids Res 2017; 45:9528–46.

394. Hardy S, Kostantin E, Wang SJ, Hristova T, Galicia-Vázquez G, Baranov PV, Pelletier J, Tremblay ML. Magnesium-sensitive upstream ORF controls PRL phosphatase expression to mediate energy metabolism. Proc Natl Acad Sci U S A 2019; 116:2925–34.

395. Zhu H, Li C, Gao C. Applications of CRISPR–Cas in agriculture and plant biotechnology. Nature Reviews Molecular Cell Biology 2020; 21:661–77.

396. Si X, Zhang H, Wang Y, Chen K, Gao C. Manipulating gene translation in plants by CRISPR–Cas9-mediated genome editing of upstream open reading frames. Nature Protocols 2020; 15:338–63.

397. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature 2015; 520:186–91.

398. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 2019; 576:149–57.

399. Wiestner A, Schlemper RJ, van der Maas AP, Skoda RC. An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. Nat Genet 1998; 18:49–52.

400. Wen Y, Liu Y, Xu Y, Zhao Y, Hua R, Wang K, Sun M, Li Y, Yang S, Zhang X-J, et al. Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. Nat Genet 2009; 41:228–33.

401. Glubb DM, Cerri E, Giese A, Zhang W, Mirza O, Thompson EE, Chen P, Das S, Jassem J, Rzyman W, et al. Novel functional germline variants in the VEGF receptor 2 gene and their effect on gene expression and microvessel density in lung cancer. Clin Cancer Res 2011; 17:5257–67.

402. Uzunoglu FG, Kolbe J, Wikman H, Güngör C, Bohn BA, Nentwich MF, Reeh M, König AM, Bockhorn M, Kutup A, et al. VEGFR-2, CXCR-2 and PAR-1 germline polymorphisms as predictors of survival in pancreatic carcinoma. Ann Oncol 2013; 24:1282–90.

403. Liu L, Dilworth D, Gao L, Monzon J, Summers A, Lassam N, Hogg D. Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. Nat Genet 1999; 21:128–32.

404. Poulat F, Desclozeaux M, Tuffery S, Jay P, Boizet B, Berta P. Mutation in the 5' noncoding region of the SRY gene in an XY sex-reversed patient. Hum Mutat 1998; Suppl 1:S192-194.

405. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RLLF, Daack-Hirsch S, Sander A, et al. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet 2002; 32:285–9.

406. Tassin J, Dürr A, Bonnet AM, Gil R, Vidailhet M, Lücking CB, Goas JY, Durif F, Abada M, Echenne B, et al. Levodopa-responsive dystonia. GTP cyclohydrolase I or parkin mutations? Brain 2000; 123 (Pt 6):1112–21.

407. Matthes T, Aguilar-Martinez P, Pizzi-Bosman L, Darbellay R, Rubbia-Brandt L, Giostra E, Michel M, Ganz T, Beris P. Severe hemochromatosis in a Portuguese family associated with a new mutation in the 5′-UTR of the HAMP gene. Blood 2004; 104:2181–3.

408. Huopio H, Jääskeläinen J, Komulainen J, Miettinen R, Kärkkäinen P, Laakso M, Tapanainen P, Voutilainen R, Otonkoski T. Acute insulin response tests for the differential diagnosis of congenital hyperinsulinism. J Clin Endocrinol Metab 2002; 87:4502–7.

409. Sözen MM, Whittall R, Oner C, Tokatli A, Kalkanoğlu HS, Dursun A, Coşkun T, Oner R, Humphries SE. The molecular basis of familial hypercholesterolaemia in Turkish patients. Atherosclerosis 2005; 180:63–71.

410. Braverman N, Chen L, Lin P, Obie C, Steel G, Douglas P, Chakraborty PK, Clarke JTR, Boneh A, Moser A, et al. Mutation analysis of PEX7 in 60 probands with rhizomelic chondrodysplasia punctata and functional correlations of genotype with phenotype. Hum Mutat 2002; 20:284–97.

411. Krude H, Biebermann H, Luck W, Horn R, Brabant G, Grüters A. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. Nat Genet 1998; 19:155–7.

412. Groussin L, Kirschner LS, Vincent-Dejean C, Perlemoine K, Jullian E, Delemer B, Zacharieva S, Pignatelli D, Carney JA, Luton JP, et al. Molecular analysis of the cyclic AMP-dependent protein kinase A (PKA) regulatory subunit 1A (PRKAR1A) gene in patients with Carney complex and primary pigmented nodular adrenocortical disease (PPNAD) reveals novel mutations and clues for pathophysiology: augmented PKA signaling is associated with adrenal tumorigenesis in PPNAD. Am J Hum Genet 2002; 71:1433–42.

413. Witt H, Luck W, Hennies HC, Classen M, Kage A, Lass U, Landt O, Becker M. Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. Nat Genet 2000; 25:213–6.

414. Oner R, Agarwal S, Dimovski AJ, Efremov GD, Petkov GH, Altay C, Gurgey A, Huisman TH. The G----A mutation at position +22 3' to the Cap site of the beta-globin gene as a possible cause for a beta-thalassemia. Hemoglobin 1991; 15:67–76.

415. Lee DSM, Park J, Kromer A, Baras A, Rader DJ, Ritchie MD, Ghanem LR, Barash Y. Disrupting upstream translation in mRNAs is associated with human disease. Nat Commun 2021; 12:1515.

416. Mehta A, Trotta CR, Peltz SW. Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. Genes Dev 2006; 20:939–53.

417. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 2001; 29:e45.

418. Aartsma-Rus A, Deutekom JCTV, Fokkema IF, Ommen G-JBV, Dunnen JTD. 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 2006; 34:135–44.

419. Bladen CL, Salgado D, Monges S, Foncuberta ME, Kekou K, Kosma K, Dawkins H, Lamont L, Roy AJ, Chamova T, et al. The TREAT-NMD DMD Global Database: Analysis of More than 7,000 Duchenne Muscular Dystrophy Mutations. Human Mutation 2015; 36:395–402.

420. Sussman D, Chadsey M, Fauce S, Engel A, Bruett A, Monnat R, Stoddard BL, Seligman LM. Isolation and characterization of new homing endonuclease specificities at individual target site positions. Journal of Molecular Biology 2004; 342:31–41.

421. Rosen LE, Morrison HA, Masri S, Brown MJ, Springstubb B, Sussman D, Stoddard BL, Seligman LM. Homing endonuclease I-CreI derivatives with novel DNA target specificities. Nucleic Acids Res 2006; 34:4791–800.

422. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012; 337:816–21.

423. Podsakoff G, Wong KK, Chatterjee S. Efficient gene transfer into nondividing cells by adeno-associated virus-based vectors. Journal of Virology 1994; 68:5656–66.

424. Muruve AKZ and DA. Immune Responses to Adeno-Associated Virus Vectors [Internet]. Current Gene Therapy2005 [cited 2020 Nov 18]; 5:323–31. Available from: https://www.eurekaselect.com/60737/article

425. Inagaki K, Fuess S, Storm TA, Gibson GA, Mctiernan CF, Kay MA, Nakai H. Robust Systemic Transduction with AAV9 Vectors in Mice: Efficient Global Cardiac Gene Transfer Superior to That of AAV8. Mol Ther 2006; 14:45–53.

426. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 2013; 339:819–23.

427. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-Guided Human Genome Engineering via Cas9. Science 2013; 339:823–6.

428. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature 2015; 520:186–91.

429. Wu Z, Yang H, Colosi P. Effect of genome size on AAV vector packaging. Mol Ther 2010; 18:80–6.

430. Bulfield G, Siller WG, Wight PA, Moore KJ. X chromosome-linked muscular dystrophy (mdx) in the mouse. Proc Natl Acad Sci U S A 1984; 81:1189–92.

431. Hakim CH, Wasala NB, Nelson CE, Wasala LP, Yue Y, Louderman JA, Lessa TB, Dai A, Zhang K, Jenkins GJ, et al. AAV CRISPR editing rescues cardiac and muscle function for 18 months in dystrophic mice. JCI Insight 2018; 3:e124297.

432. Long C, Amoasii L, Mireault AA, McAnally JR, Li H, Sanchez-Ortiz E, Bhattacharyya S, Shelton JM, Bassel-Duby R, Olson EN. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. Science 2016; 351:400–3.

433. Nelson CE, Wu Y, Gemberling MP, Oliver ML, Waller MA, Bohning JD, Robinson-Hamm JN, Bulaklak K, Castellanos Rivera RM, Collier JH, et al. Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. Nat Med 2019; 25:427–32.

434. Nelson CE, Hakim CH, Ousterout DG, Thakore PI, Moreb EA, Rivera RMC, Madhavan S, Pan X, Ran FA, Yan WX, et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science 2016; 351:403–7.

435. Tabebordbar M, Zhu K, Cheng JKW, Chew WL, Widrick JJ, Yan WX, Maesner C, Wu EY, Xiao R, Ran FA, et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science 2016; 351:407–11.

436. Betts CA, Saleh AF, Carr CA, Muses S, Wells KE, Hammond SM, Godfrey C, McClorey G, Woffindale C, Clarke K, et al. Implications for Cardiac Function Following Rescue of the Dystrophic Diaphragm in a Mouse Model of Duchenne Muscular Dystrophy. Scientific Reports 2015; 5:11632.

437. Costas JM, Nye DJ, Henley JB, Plochocki JH. Voluntary exercise induces structural remodeling in the hearts of dystrophin-deficient mice. Muscle Nerve 2010; 42:881–5.

438. Quinlan JG, Hahn HS, Wong BL, Lorenz JN, Wenisch AS, Levin LS. Evolution of the mdx mouse cardiomyopathy: physiological and morphological findings. Neuromuscular Disorders 2004; 14:491–6.

439. Townsend D, Blankinship MJ, Allen JM, Gregorevic P, Chamberlain JS, Metzger JM. Systemic Administration of Micro-dystrophin Restores Cardiac Geometry and Prevents Dobutamine-induced Cardiac Pump Failure. Molecular Therapy 2007; 15:1086–92.

440. Yasuda S, Townsend D, Michele DE, Favre EG, Day SM, Metzger JM. Dystrophic heart failure blocked by membrane sealant poloxamer. Nature 2005; 436:1025–9.

441. Chamberlain JS, Metzger J, Reyes M, Townsend D, Faulkner JA. Dystrophin-deficient mdx mice display a reduced life span and are susceptible to spontaneous rhabdomyosarcoma. FASEB J 2007; 21:2195–204.

442. Grady RM, Teng H, Nichol MC, Cunningham JC, Wilkinson RS, Sanes JR. Skeletal and Cardiac Myopathies in Mice Lacking Utrophin and Dystrophin: A Model for Duchenne Muscular Dystrophy. Cell 1997; 90:729–38.

443. Deconinck AE, Rafael JA, Skinner JA, Brown SC, Potter AC, Metzinger L, Watt DJ, Dickson JG, Tinsley JM, Davies KE. Utrophin-Dystrophin-Deficient Mice as a Model for Duchenne Muscular Dystrophy. Cell 1997; 90:717–27.

444. Goyenvalle A, Babbs A, Wright J, Wilkins V, Powell D, Garcia L, Davies KE. Rescue of severely affected dystrophin/utrophin-deficient mice through scAAV-U7snRNA-mediated exon skipping. Hum Mol Genet 2012; 21:2559–71.

445. Goyenvalle A, Babbs A, Powell D, Kole R, Fletcher S, Wilton SD, Davies KE. Prevention of dystrophic pathology in severely affected dystrophin/utrophin-deficient mice by morpholino-oligomer-mediated exon-skipping. Mol Ther 2010; 18:198–205.

446. Finkel RS, Mercuri E, Darras BT, Connolly AM, Kuntz NL, Kirschner J, Chiriboga CA, Saito K, Servais L, Tizzano E, et al. Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. N Engl J Med 2017; 377:1723–32.

447. De Vivo DC, Bertini E, Swoboda KJ, Hwu W-L, Crawford TO, Finkel RS, Kirschner J, Kuntz NL, Parsons JA, Ryan MM, et al. Nusinersen initiated in infants during the presymptomatic stage of spinal muscular atrophy: Interim efficacy and safety results from the Phase 2 NURTURE study. Neuromuscular Disorders 2019; 29:842–56.

448. Rosant C, Nagel M-D, Pérot C. Aging affects passive stiffness and spindle function of the rat soleus muscle. Experimental Gerontology 2007; 42:301–8.

449. Deconinck AE, Potter AC, Tinsley JM, Wood SJ, Vater R, Young C, Metzinger L, Vincent A, Slater CR, Davies KE. Postsynaptic abnormalities at the neuromuscular junctions of utrophin-deficient mice. J Cell Biol 1997; 136:883–94.

450. Engler C, Kandzia R, Marillonnet S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLOS ONE 2008; 3:e3647.

451. Bjourson AJ, Cooper JE. Band-stab PCR: a simple technique for the purification of individual PCR products. Nucleic Acids Res 1992; 20:4675.

452. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature 2015; 520:186–91.

453. Lai Y, Yue Y, Duan D. Evidence for the Failure of Adeno-associated Virus Serotype 5 to Package a Viral Genome |[ge]|8.2 kb. Mol Ther 2009; 18:75–9.

454. Immunohistochemistry (IHC): the complete guide | Abcam [Internet]. [cited 2021 Jun 21]; Available from: https://www.abcam.com/content/immunohistochemistry-the-complete-guide

455. Carroll KJ, Makarewich CA, McAnally J, Anderson DM, Zentilin L, Liu N, Giacca M, Bassel-Duby R, Olson EN. A mouse model for adult cardiac-specific gene deletion with CRISPR/Cas9. Proc Natl Acad Sci U S A 2016; 113:338–43.

456. Westering TLE van, Lomonosova Y, Coenen-Stass AML, Betts CA, Bhomra A, Hulsker M, Clark LE, McClorey G, Aartsma-Rus A, Putten M van, et al. Uniform sarcolemmal dystrophin expression is required to prevent extracellular microRNA release and improve dystrophic pathology. Journal of Cachexia, Sarcopenia and Muscle 2020; 11:578–93.

457. Bengtsson NE, Hall JK, Odom GL, Phelps MP, Andrus CR, Hawkins RD, Hauschka SD, Chamberlain JR, Chamberlain JS. Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy. Nature Communications 2017; 8:14454.

458. El Refaey M, Xu L, Gao Y, Canan BD, Adesanya TMA, Warner SC, Akagi K, Symer DE, Mohler PJ, Ma J, et al. In Vivo Genome Editing Restores Dystrophin Expression and Cardiac Function in Dystrophic Mice. Circ Res 2017; 121:923–9.

459. Wang Z, Zhu T, Qiao C, Zhou L, Wang B, Zhang J, Chen C, Li J, Xiao X. Adeno-associated virus serotype 8 efficiently delivers genes to muscle and heart. Nat Biotechnol 2005; 23:321–8.

460. Gregorevic P, Blankinship MJ, Allen JM, Crawford RW, Meuse L, Miller DG, Russell DW, Chamberlain JS. Systemic delivery of genes to striated muscles using adeno-associated viral vectors. Nature Medicine 2004; 10:828–34.

461. Verhaart IEC, van Vliet-van den Dool L, Sipkens JA, de Kimpe SJ, Kolfschoten IGM, van Deutekom JCT, Liefaard L, Ridings JE, Hood SR, Aartsma-Rus A. The Dynamics of Compound, Transcript, and Protein Effects After Treatment With 2OMePS Antisense Oligonucleotides in mdx Mice. Molecular Therapy - Nucleic Acids 2014; 3:e148.

462. Wu B, Lu P, Cloer C, Shaban M, Grewal S, Milazi S, Shah SN, Moulton HM, Lu QL. Long-term rescue of dystrophin expression and improvement in muscle pathology and function in dystrophic mdx mice by peptide-conjugated morpholino. Am J Pathol 2012; 181:392–400.

463. Bostick Brian, Yue Yongping, Long Chun, Duan Dongsheng. Prevention of Dystrophin-Deficient Cardiomyopathy in Twenty-One-Month-Old Carrier Mice by Mosaic Dystrophin Expression or Complementary Dystrophin/Utrophin Expression. Circulation Research 2008; 102:121–30.

464. Howard ZM, Dorn LE, Lowe J, Gertzen MD, Ciccone P, Rastogi N, Odom GL, Accornero F, Chamberlain JS, Rafael-Fortney JA. Micro-dystrophin gene therapy prevents heart failure in an improved Duchenne muscular dystrophy cardiomyopathy mouse model. JCI Insight 2021; 6:146511.

465. Tinsley J, Deconinck N, Fisher R, Kahn D, Phelps S, Gillis JM, Davies K. Expression of full-length utrophin prevents muscular dystrophy in mdx mice. Nat Med 1998; 4:1441–4.

466. van Putten M, Kumar D, Hulsker M, Hoogaars WMH, Plomp JJ, van Opstal A, van Iterson M, Admiraal P, van Ommen G-JB, 't Hoen PAC, et al. Comparison of skeletal muscle pathology and motor function of dystrophin and utrophin deficient mouse strains. Neuromuscul Disord 2012; 22:406–17.

467. Zhou L, Rafael-Fortney JA, Huang P, Zhao XS, Cheng G, Zhou X, Kaminski HJ, Liu L, Ransohoff RM. Haploinsufficiency of utrophin gene worsens skeletal muscle inflammation and fibrosis in mdx mice. Journal of the Neurological Sciences 2008; 264:106–11.

468. Godfrey C, Muses S, McClorey G, Wells KE, Coursindel T, Terry RL, Betts C, Hammond S, O'Donovan L, Hildyard J, et al. How much dystrophin is enough: the physiological consequences of different levels of dystrophin in the mdx mouse. Hum Mol Genet 2015; 24:4225–37.

469. Wells DJ, Wells KE, Asante EA, Turner G, Sunada Y, Campbell KP, Walsh FS, Dickson G. Expression of human full-length and minidystrophin in transgenic mdx mice: implications for gene therapy of Duchenne muscular dystrophy. Human Molecular Genetics 1995; 4:1245–50.

470. Rafael JA, Tinsley JM, Potter AC, Deconinck AE, Davies KE. Skeletal muscle-specific expression of a utrophin transgene rescues utrophin-dystrophin deficient mice. Nat Genet 1998; 19:79–82.

471. Li D, Yue Y, Duan D. Marginal level dystrophin expression improves clinical outcome in a strain of dystrophin/utrophin double knockout mice. PLoS One 2010; 5:e15286.

472. van Putten M, Hulsker M, Young C, Nadarajah VD, Heemskerk H, van der Weerd L, 't Hoen PAC, van Ommen G-JB, Aartsma-Rus AM. Low dystrophin levels increase survival and improve muscle pathology and function in dystrophin/utrophin double-knockout mice. FASEB J 2013; 27:2484–95.

473. Hoffman EP, Fischbeck KH, Brown RH, Johnson M, Medori R, Loike JD, Harris JB, Waterston R, Brooke M, Specht L. Characterization of dystrophin in muscle-biopsy specimens from patients with Duchenne's or Becker's muscular dystrophy. N Engl J Med 1988; 318:1363–8.

474. Wu B, Cloer C, Lu P, Milazi S, Shaban M, Shah SN, Marston-Poe L, Moulton HM, Lu QL. Exon skipping restores dystrophin expression, but fails to prevent disease progression in later stage dystrophic dko mice. Gene Ther 2014; 21:785–93.

475. Crisp A, Yin H, Goyenvalle A, Betts C, Moulton HM, Seow Y, Babbs A, Merritt T, Saleh AF, Gait MJ, et al. Diaphragm rescue alone prevents heart dysfunction in dystrophic mice. Human Molecular Genetics 2011; 20:413–21.

476. Goyenvalle A, Griffith G, Babbs A, Andaloussi SE, Ezzat K, Avril A, Dugovic B, Chaussenot R, Ferry A, Voit T, et al. Functional correction in mouse models of muscular dystrophy using exon-skipping tricyclo-DNA oligomers. Nature Medicine 2015; 21:270–5.

477. McCarty DM, Fu H, Monahan PE, Toulson CE, Naik P, Samulski RJ. Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction in vivo. Gene Therapy 2003; 10:2112–8.

478. Wang Z, Ma H-I, Li J, Sun L, Zhang J, Xiao X. Rapid and highly efficient transduction by double-stranded adeno-associated virus vectors in vitro and in vivo. Gene Therapy 2003; 10:2105–11.

479. Ferrari FK, Samulski T, Shenk T, Samulski RJ. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. Journal of Virology 1996; 70:3227–34.

480. Fisher KJ, Gao GP, Weitzman MD, DeMatteo R, Burda JF, Wilson JM. Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. J Virol 1996; 70:520–32.

481. Daya S, Berns KI. Gene Therapy Using Adeno-Associated Virus Vectors. Clin Microbiol Rev 2008; 21:583–93.

482. Le Hir M, Goyenvalle A, Peccate C, Précigout G, Davies KE, Voit T, Garcia L, Lorain S. AAV Genome Loss From Dystrophic Mouse Muscles During AAV-U7 snRNA-mediated Exon-skipping Therapy. Mol Ther 2013; 21:1551–8.

483. Hanson B, Wood MJA, Roberts TC. Molecular correction of Duchenne muscular dystrophy by splice modulation and gene editing. RNA Biol 2021; 20:1–15.

484. Hall ZW, Ralston E. Nuclear domains in muscle cells. Cell 1989; 59:771–2.

485. Min Y-L, Li H, Rodriguez-Caycedo C, Mireault AA, Huang J, Shelton JM, McAnally JR, Amoasii L, Mammen PPA, Bassel-Duby R, et al. CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells. Science Advances 2019; 5:eaav4324.

486. Hanlon KS, Kleinstiver BP, Garcia SP, Zaborowski MP, Volak A, Spirig SE, Muller A, Sousa AA, Tsai SQ, Bengtsson NE, et al. High levels of AAV vector integration into CRISPR-induced DNA breaks. Nature Communications 2019; 10:4439.

487. Amoasii L, Long C, Li H, Mireault AA, Shelton JM, Sanchez-Ortiz E, McAnally JR, Bhattacharyya S, Schmidt F, Grimm D, et al. Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. Sci Transl Med 2017; 9:eaan8081.

488. Zhang Y, Long C, Li H, McAnally JR, Baskin KK, Shelton JM, Bassel-Duby R, Olson EN. CRISPR-Cpf1 correction of muscular dystrophy mutations in human cardiomyocytes and mice. Science Advances 2017; 3:e1602814.

489. Amoasii L, Hildyard JCW, Li H, Sanchez-Ortiz E, Mireault A, Caballero D, Harron R, Stathopoulou T-R, Massey C, Shelton JM, et al. Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. Science 2018; 362:86–91.

490. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 2016; 533:420–4.

491. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 2017; 551:464.

492. Ryu S-M, Koo T, Kim K, Lim K, Baek G, Kim S-T, Kim HS, Kim D, Lee H, Chung E, et al. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. Nature Biotechnology 2018; 36:536–9.

493. Chemello F, Chai AC, Li H, Rodriguez-Caycedo C, Sanchez-Ortiz E, Atmanli A, Mireault AA, Liu N, Bassel-Duby R, Olson EN. Precise correction of Duchenne muscular dystrophy exon deletion mutations by base and prime editing. Science Advances 2021; 7:eabg4910.

494. Cox DBT, Gootenberg JS, Abudayyeh OO, Franklin B, Kellner MJ, Joung J, Zhang F. RNA editing with CRISPR-Cas13. Science 2017; 358:1019–27.

495. Xu X, Wan T, Xin H, Li D, Pan H, Wu J, Ping Y. Delivery of CRISPR/Cas9 for therapeutic genome editing. The Journal of Gene Medicine 2019; 21:e3107.

496. Lee K, Conboy M, Park HM, Jiang F, Kim HJ, Dewitt MA, Mackley VA, Chang K, Rao A, Skinner C, et al. Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. Nat Biomed Eng 2017; 1:889–901.

497. Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, Hu JH, Maeder ML, Joung JK, Chen Z-Y, Liu DR. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat Biotechnol 2015; 33:73–80.

498. Gao X, Tao Y, Lamas V, Huang M, Yeh W-H, Pan B, Hu Y-J, Hu JH, Thompson DB, Shu Y, et al. Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. Nature 2018; 553:217–21.

499. Ramakrishna S, Kwaku Dad A-B, Beloor J, Gopalappa R, Lee S-K, Kim H. Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. Genome Res 2014; 24:1020–7.

500. Gee P, Lung MSY, Okuzaki Y, Sasakawa N, Iguchi T, Makita Y, Hozumi H, Miura Y, Yang LF, Iwasaki M, et al. Extracellular nanovesicles for packaging of CRISPR-Cas9 protein and sgRNA to induce therapeutic exon skipping. Nature Communications 2020; 11:1334.

501. Zhang Y, Li H, Min Y-L, Sanchez-Ortiz E, Huang J, Mireault AA, Shelton JM, Kim J, Mammen PPA, Bassel-Duby R, et al. Enhanced CRISPR-Cas9 correction of Duchenne muscular dystrophy in mice by a self-complementary AAV delivery system. Sci Adv 2020; 6:eaay6812.

502. High-dose AAV gene therapy deaths. Nature Biotechnology 2020; 38:910–910.

503. Lysogene Confirms Child's Death in Phase II/III Gene Therapy Trial. [Internet]. Available from: https://www.genengnews.com/news/lysogene-confirms-childs-death-in-phase-ii-iii-gene-therapy-trial/.

504. Nance ME, Shi R, Hakim CH, Wasala NB, Yue Y, Pan X, Zhang T, Robinson CA, Duan SX, Yao G, et al. AAV9 Edits Muscle Stem Cells in Normal and Dystrophic Adult Mice. Mol Ther 2019; 27:1568–85.

505. Arnett AL, Konieczny P, Ramos JN, Hall J, Odom G, Yablonka-Reuveni Z, Chamberlain JR, Chamberlain JS. Adeno-associated viral (AAV) vectors do not efficiently target muscle satellite cells. Mol Ther Methods Clin Dev 2014; 1:14038.

506. Chamberlain JS, Jaynes JB, Hauschka SD. Regulation of creatine kinase induction in differentiating mouse myoblasts. Mol Cell Biol 1985; 5:484–92.

507. Jaynes JB, Chamberlain JS, Buskin JN, Johnson JE, Hauschka SD. Transcriptional regulation of the muscle creatine kinase gene and regulated expression in transfected mouse myoblasts. Mol Cell Biol 1986; 6:2855–64.

508. Guescini M, Guidolin D, Vallorani L, Casadei L, Gioacchini AM, Tibollo P, Battistelli M, Falcieri E, Battistin L, Agnati LF, et al. C2C12 myoblasts release micro-vesicles containing mtDNA and proteins involved in signal transduction. Experimental Cell Research 2010; 316:1977–84.

509. Madison RD, McGee C, Rawson R, Robinson GA. Extracellular vesicles from a muscle cell line (C2C12) enhance cell survival and neurite outgrowth of a motor neuron cell line (NSC-34). Journal of Extracellular Vesicles 2014; 3:22865.

510. Fry CS, Lee JD, Jackson JR, Kirby TJ, Stasko SA, Liu H, Dupont-Versteegden EE, McCarthy JJ, Peterson CA. Regulation of the muscle fiber micro environment by activated satellite cells during hypertrophy. The FASEB Journal 2014; 28:1654–65.

511. Karttunen J, Heiskanen M, Navarro-Ferrandis V, Das Gupta S, Lipponen A, Puhakka N, Rilla K, Koistinen A, Pitkänen A. Precipitation-based extracellular vesicle isolation from rat plasma co-precipitate vesicle-free microRNAs. J Extracell Vesicles 2019; 8:1555410.

512. Van Deun J, Mestdagh P, Sormunen R, Cocquyt V, Vermaelen K, Vandesompele J, Bracke M, De Wever O, Hendrix A. The impact of disparate isolation methods for extracellular vesicles on downstream RNA profiling. J Extracell Vesicles 2014; 3.

513. Whittaker TE, Nagelkerke A, Nele V, Kauscher U, Stevens MM. Experimental artefacts can lead to misattribution of bioactivity from soluble mesenchymal stem cell paracrine factors to extracellular vesicles. Journal of Extracellular Vesicles 2020; 9:1807674.

514. Cvjetkovic A, Lötvall J, Lässer C. The influence of rotor type and centrifugation time on the yield and purity of extracellular vesicles. J Extracell Vesicles 2014; 3.

515. Nordin JZ, Lee Y, Vader P, Mäger I, Johansson HJ, Heusermann W, Wiklander OPB, Hällbrink M, Seow Y, Bultema JJ, et al. Ultrafiltration with size-exclusion liquid chromatography for high yield isolation of extracellular vesicles preserving intact biophysical and functional properties. Nanomedicine 2015; 11:879–83.

516. Chan CYX, Masui O, Krakovska O, Belozerov VE, Voisin S, Ghanny S, Chen J, Moyez D, Zhu P, Evans KR, et al. Identification of differentially regulated secretome components during skeletal myogenesis. Mol Cell Proteomics 2011; 10:M110.004804.

517. Henningsen J, Rigbolt KTG, Blagoev B, Pedersen BK, Kratchmarova I. Dynamics of the skeletal muscle secretome during myoblast differentiation. Mol Cell Proteomics 2010; 9:2482–96.

518. Yaffe D, Saxel O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature 1977; 270:725–7.

519. Kosaka N, Iguchi H, Yoshioka Y, Takeshita F, Matsuki Y, Ochiya T. Secretory Mechanisms and Intercellular Transfer of MicroRNAs in Living Cells*♦. Journal of Biological Chemistry 2010; 285:17442–52.

520. Maguire CA, Balaj L, Sivaraman S, Crommentuijn MH, Ericsson M, Mincheva-Nilsson L, Baranov V, Gianni D, Tannous BA, Sena-Esteves M, et al. Microvesicle-associated AAV Vector as a Novel Gene Delivery System. Molecular Therapy 2012; 20:960–71.

521. Kolb HC, Finn MG, Sharpless KB. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. Angewandte Chemie International Edition 2001; 40:2004–21.

522. Deng Z, Poliakov A, Hardy RW, Clements R, Liu C, Liu Y, Wang J, Xiang X, Zhang S, Zhuang X, et al. Adipose tissue exosome-like vesicles mediate activation of macrophage-induced insulin resistance. Diabetes 2009; 58:2498–505.

523. Coenen-Stass AML, Pauwels MJ, Hanson B, Perez CM, Conceição M, Wood MJA, Mäger I, Roberts TC. Extracellular microRNAs exhibit sequence-dependent stability and cellular release kinetics. RNA Biology 2019; 16:696–706.

524. Stenmark H. Rab GTPases as coordinators of vesicle traffic. Nature Reviews Molecular Cell Biology 2009; 10:513–25.

525. Fukuda M. Regulation of secretory vesicle traffic by Rab small GTPases. Cell Mol Life Sci 2008; 65:2801–13.

526. Bobrie A, Krumeich S, Reyal F, Recchi C, Moita LF, Seabra MC, Ostrowski M, Théry C. Rab27a Supports Exosome-Dependent and -Independent Mechanisms That Modify the Tumor Microenvironment and Can Promote Tumor Progression. Cancer Res 2012; 72:4920–30.

527. Webber JP, Spary LK, Sanders AJ, Chowdhury R, Jiang WG, Steadman R, Wymant J, Jones AT, Kynaston H, Mason MD, et al. Differentiation of tumour-promoting stromal myofibroblasts by cancer exosomes. Oncogene 2015; 34:290–302.

528. Kowal J, Tkach M, Théry C. Biogenesis and secretion of exosomes. Current Opinion in Cell Biology 2014; 29:116–25.

529. Charrasse S, Meriane M, Comunale F, Blangy A, Gauthier-Rouvière C. N-cadherin–dependent cell–cell contact regulates Rho GTPases and β-catenin localization in mouse C2C12 myoblasts. Journal of Cell Biology 2002; 158:953–65.

530. Thermo Fisher Scientific. Opti-MEM™ I Reduced Serum Medium Product Information Sheet [Internet]. [cited 2021 Jun 15]; Available from: https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FOPTI_MEM_I_reduced_serum_PI.pdf&title=UHJvZHVjdCBTaGVldDogT1BUSS1NRU0gSSBSZWR1Y2VkIFNlcnVtIE1lZGl1bQ==

531. Okabe S, Forsberg-Nilsson K, Spiro AC, Segal M, McKay RDG. Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. Mechanisms of Development 1996; 59:89–102.

532. Barberi T, Bradbury M, Dincer Z, Panagiotakos G, Socci ND, Studer L. Derivation of engraftable skeletal myoblasts from human embryonic stem cells. Nature Medicine 2007; 13:642–8.

533. Awaya T, Kato T, Mizuno Y, Chang H, Niwa A, Umeda K, Nakahata T, Heike T. Selective development of myogenic mesenchymal cells from human embryonic and induced pluripotent stem cells. PLoS One 2012; 7:e51638.

534. Chal J, Oginuma M, Al Tanoury Z, Gobert B, Sumara O, Hick A, Bousson F, Zidouni Y, Mursch C, Moncuquet P, et al. Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. Nat Biotechnol 2015; 33:962–9.

535. Borchin B, Chen J, Barberi T. Derivation and FACS-mediated purification of PAX3+/PAX7+ skeletal muscle precursors from human pluripotent stem cells. Stem Cell Reports 2013; 1:620–31.

536. Chal J, Al Tanoury Z, Hestin M, Gobert B, Aivio S, Hick A, Cherrier T, Nesmith AP, Parker KK, Pourquié O. Generation of human muscle fibers and satellite-like cells from human pluripotent stem cells in vitro. Nat Protoc 2016; 11:1833–50.

537. Swartz EW, Baek J, Pribadi M, Wojta KJ, Almeida S, Karydas A, Gao F-B, Miller BL, Coppola G. A Novel Protocol for Directed Differentiation of C9orf72-Associated Human Induced Pluripotent Stem Cells Into Contractile Skeletal Myotubes. Stem Cells Transl Med 2016; 5:1461–72.

538. Smits AM, van Vliet P, Metz CH, Korfage T, Sluijter JP, Doevendans PA, Goumans M-J. Human cardiomyocyte progenitor cells differentiate into functional mature cardiomyocytes: an in vitro model for studying human cardiac physiology and pathophysiology. Nat Protoc 2009; 4:232–43.

539. Hidalgo-Carcedo C, Hooper S, Chaudhry SI, Williamson P, Harrington K, Leitinger B, Sahai E. Collective cell migration requires suppression of actomyosin at cell-cell contacts mediated by DDR1 and the cell polarity regulators Par3 and Par6. Nat Cell Biol 2011; 13:49–58.

540. Ahmad SS, Ahmad K, Lee EJ, Lee Y-H, Choi I. Implications of Insulin-Like Growth Factor-1 in Skeletal Muscle and Various Diseases. Cells 2020; 9:E1773.

541. LeRoith D, McGuinness M, Shemer J, Stannard B, Lanau F, Faria TN, Kato H, Werner H, Adamo M, Roberts CT. Insulin-like growth factors. Biol Signals 1992; 1:173–81.

542. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. Nat Rev Mol Cell Biol 2006; 7:85–96.

543. Cai W, Sakaguchi M, Kleinridders A, Gonzalez-Del Pino G, Dreyfuss JM, O'Neill BT, Ramirez AK, Pan H, Winnay JN, Boucher J, et al. Domain-dependent effects of insulin and IGF-1 receptors on signalling and gene expression. Nat Commun 2017; 8:14892.

544. Yoshida T, Delafontaine P. Mechanisms of IGF-1-Mediated Regulation of Skeletal Muscle Hypertrophy and Atrophy. Cells 2020; 9:1970.

545. Liu D, Moberg E, Kollind M, Lins PE, Adamson U. A high concentration of circulating insulin suppresses the glucagon response to hypoglycemia in normal man. J Clin Endocrinol Metab 1991; 73:1123–8.

546. Ostenfeld MS, Jeppesen DK, Laurberg JR, Boysen AT, Bramsen JB, Primdal-Bengtson B, Hendrix A, Lamy P, Dagnaes-Hansen F, Rasmussen MH, et al. Cellular Disposal of miR23b by RAB27-Dependent Exosome Release Is Linked to Acquisition of Metastatic Properties. Cancer Res 2014; 74:5758–71.

547. Madeo M, Colbert PL, Vermeer DW, Lucido CT, Cain JT, Vichaya EG, Grossberg AJ, Muirhead D, Rickel AP, Hong Z, et al. Cancer exosomes induce tumor innervation. Nature Communications 2018; 9:4284.

548. Zheng Y, Campbell EC, Lucocq J, Riches A, Powis SJ. Monitoring the Rab27 associated exosome pathway using nanoparticle tracking analysis. Experimental Cell Research 2013; 319:1706–13.

549. Sinha S, Hoshino D, Hong NH, Kirkbride KC, Grega-Larson NE, Seiki M, Tyska MJ, Weaver AM. Cortactin promotes exosome secretion by controlling branched actin dynamics. Journal of Cell Biology 2016; 214:197–213.

550. Akat KM, Lee YA, Hurley A, Morozov P, Max KE, Brown M, Bogardus K, Sopeyin A, Hildner K, Diacovo TG, et al. Detection of circulating extracellular mRNAs by modified small-RNA-sequencing analysis. JCI Insight 2019; 5.

551. Chevillet JR, Kang Q, Ruf IK, Briggs HA, Vojtech LN, Hughes SM, Cheng HH, Arroyo JD, Meredith EK, Gallichotte EN, et al. Quantitative and stoichiometric analysis of the microRNA content of exosomes. PNAS 2014; 111:14888–93.

552. Kim S, Lee M-J, Choi J-Y, Park D-H, Kwak H-B, Moon S, Koh J-W, Shin H-K, Ryu J-K, Park C-S, et al. Roles of Exosome-Like Vesicles Released from Inflammatory C2C12 Myotubes: Regulation of Myocyte Differentiation and Myokine Expression. Cell Physiol Biochem 2018; 48:1829–42.

553. Sverdlov ED. Amedeo Avogadro's cry: what is 1 μg of exosomes? Bioessays 2012; 34:873–5.

554. Peltz SW, Morsy M, Welch EM, Jacobson A. Ataluren as an agent for therapeutic nonsense suppression. Annu Rev Med 2013; 64:407–25.

555. Welch EM, Barton ER, Zhuo J, Tomizawa Y, Friesen WJ, Trifillis P, Paushkin S, Patel M, Trotta CR, Hwang S, et al. PTC124 targets genetic disorders caused by nonsense mutations. Nature 2007; 447:87–91.

556. Péladeau C, Adam N, Bronicki LM, Coriati A, Thabet M, Al-Rewashdy H, Vanstone J, Mears A, Renaud J-M, Holcik M, et al. Identification of therapeutics that target eEF1A2 and upregulate utrophin A translation in dystrophic muscles. Nature Communications 2020; 11:1990.

557. Scholz A, Eggenhofer F, Gelhausen R, Grüning B, Zarnack K, Brüne B, Backofen R, Schmid T. uORF-Tools—Workflow for the determination of translation-regulatory upstream open reading frames. PLOS ONE 2019; 14:e0222459.

558. Ji Z, Song R, Regev A, Struhl K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. Elife 2015; 4:e08890.

559. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science 2009; 324:218–23.

560. Ingolia NT, Lareau LF, Weissman JS. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. Cell 2011; 147:789–802.

561. Fresno M, Jiménez A, Vázquez D. Inhibition of Translation in Eukaryotic Systems by Harringtonine. European Journal of Biochemistry 1977; 72:323–30.

562. Robert F, Carrier M, Rawe S, Chen S, Lowe S, Pelletier J. Altering chemosensitivity by modulating translation elongation. PLoS One 2009; 4:e5428.

563. Takahashi H, Miyaki S, Onouchi H, Motomura T, Idesako N, Takahashi A, Murase M, Fukuyoshi S, Endo T, Satou K, et al. Exhaustive identification of conserved upstream open reading frames with potential translational regulatory functions from animal genomes. Scientific Reports 2020; 10:16289.

564. Kozak M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem 1991; 266:19867–70.

565. Sedman SA, Gelembiuk GW, Mertz JE. Translation initiation at a downstream AUG occurs with increased efficiency when the upstream AUG is located very close to the 5' cap. J Virol 1990; 64:453–7.

566. Kozak M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J Mol Biol 1987; 196:947–50.

567. Luukkonen BG, Tan W, Schwartz S. Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. J Virol 1995; 69:4086–94.

568. Kozak M. Determinants of translational fidelity and efficiency in vertebrate mRNAs. Biochimie 1994; 76:815–21.

569. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 1990; 18:6097–100.

570. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. Nature Reviews Cancer 2004; 4:177–83.

571. Yamashita R, Suzuki Y, Nakai K, Sugano S. Small open reading frames in 5' untranslated regions of mRnas. C R Biol 2003; 326:987–91.

572. Lee S, Liu B, Lee S, Huang S-X, Shen B, Qian S-B. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci U S A 2012; 109:E2424-2432.

573. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005; 15:1034–50.

574. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer 2018; 18:696–705.

575. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 2009; 6:343–5.

576. Harding AE. Molecular genetics and clinical aspects of inherited disorders of nerve and muscle. Curr Opin Neurol Neurosurg 1992; 5:600–4.

577. Werdnig G. Two early infantile hereditary cases of progressive muscular atrophy simulating dystrophy, but on a neural basis. 1891. Arch Neurol 1971; 25:276–8.

578. Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M. Identification and characterization of a spinal muscular atrophy-determining gene. Cell 1995; 80:155–65.

579. Kay DM, Stevens CF, Parker A, Saavedra-Matiz CA, Sack V, Chung WK, Chiriboga CA, Engelstad K, Laureta E, Farooq O, et al. Implementation of population-based newborn screening reveals low incidence of spinal muscular atrophy. Genet Med 2020; 22:1296–302.

580. Verhaart IEC, Robertson A, Wilson IJ, Aartsma-Rus A, Cameron S, Jones CC, Cook SF, Lochmüller H. Prevalence, incidence and carrier frequency of 5q-linked spinal muscular atrophy - a literature review. Orphanet J Rare Dis 2017; 12:124.

581. Dabrowski C, Alwine JC. Translational control of synthesis of simian virus 40 late proteins from polycistronic 19S late mRNA. J Virol 1988; 62:3182–92.

582. Kozak M. A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. Gene Expr 1991; 1:111–5.

583. Schuller AP, Green R. Roadblocks and resolutions in eukaryotic translation. Nat Rev Mol Cell Biol 2018; 19:526–41.

584. Oyama M, Itagaki C, Hata H, Suzuki Y, Izumi T, Natsume T, Isobe T, Sugano S. Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs. Genome Res 2004; 14:2048–52.

585. Rock KL, Gramm C, Rothstein L, Clark K, Stein R, Dick L, Hwang D, Goldberg AL. Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. Cell 1994; 78:761–71.

586. Goldberg AL, Stein R, Adams J. New insights into proteasome function: from archaebacteria to drug development. Chemistry & Biology 1995; 2:503–8.

587. Karagyozov L, Godfrey R, Böhmer S-A, Petermann A, Hölters S, Östman A, Böhmer F-D. The structure of the 5′-end of the protein-tyrosine phosphatase PTPRJ mRNA reveals a novel mechanism for translation attenuation. Nucleic Acids Res 2008; 36:4443–53.

588. Raney A, Law GL, Mize GJ, Morris DR. Regulated Translation Termination at the Upstream Open Reading Frame in S-Adenosylmethionine Decarboxylase mRNA*. Journal of Biological Chemistry 2002; 277:5988–94.

589. Wells JA. Systematic mutational analyses of protein-protein interfaces. Methods Enzymol 1991; 202:390–411.

590. Cunningham BC, Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 1989; 244:1081–5.

591. Matthews BW. Structural and genetic analysis of the folding and function of T4 lysozyme. FASEB J 1996; 10:35–41.

592. Poyry TAA. What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? Genes & Development 2004; 18:62–75.

593. Ito K, editor. Regulatory Nascent Polypeptides [Internet]. Tokyo: Springer Japan; 2014 [cited 2021 May 18]. Available from: http://link.springer.com/10.1007/978-4-431-55052-5

594. Reshetnikov R, Golovin A, Spiridonova V, Kopylov A, Šponer J. Structural Dynamics of Thrombin-Binding DNA Aptamer d(GGTTGGTGTGGTTGG) Quadruplex DNA Studied by Large-Scale Explicit Solvent Simulations. J Chem Theory Comput 2010; 6:3003–14.

595. de Mezer M, Wojciechowska M, Napierala M, Sobczak K, Krzyzosiak WJ. Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. Nucleic Acids Res 2011; 39:3852–63.

596. Gillman EC, Slusher LB, Martin NC, Hopper AK. MOD5 translation initiation sites determine N6-isopentenyladenosine modification of mitochondrial and cytoplasmic tRNA. Mol Cell Biol 1991; 11:2382–90.

597. Zhao M, Zhou S, Wu L, Deng Y. Model-driven promoter strength prediction based on a fine-tuned synthetic promoter library in Escherichia coli. bioRxiv 2020; :2020.06.25.170365.

598. McCarty DM, Fu H, Monahan PE, Toulson CE, Naik P, Samulski RJ. Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction in vivo. Gene Ther 2003; 10:2112–8.

599. Sasaki S, Sun R, Bui H-H, Crosby JR, Monia BP, Guo S. Steric Inhibition of 5′ UTR Regulatory Elements Results in Upregulation of Human CFTR. Molecular Therapy 2019; 27:1749–57.

600. Fukuda M, Umeno H, Nose K, Nishitarumizu A, Noguchi R, Nakagawa H. Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. Sci Rep 2017; 7:41478.

601. Wettengel J, Reautschnig P, Geisler S, Kahle PJ, Stafforst T. Harnessing human ADAR2 for RNA repair - Recoding a PINK1 mutation rescues mitophagy. Nucleic Acids Res 2017; 45:2797–808.

602. Heep M, Mach P, Reautschnig P, Wettengel J, Stafforst T. Applying Human ADAR1p110 and ADAR1p150 for Site-Directed RNA Editing-G/C Substitution Stabilizes GuideRNAs against Editing. Genes (Basel) 2017; 8.

603. Kearse MG, Wilusz JE. Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev 2017; 31:1717–31.

604. Ivanov IP, Loughran G, Sachs MS, Atkins JF. Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). Proc Natl Acad Sci U S A 2010; 107:18056–60.

605. Na CH, Barbhuiya MA, Kim M-S, Verbruggen S, Eacker SM, Pletnikova O, Troncoso JC, Halushka MK, Menschaert G, Overall CM, et al. Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. Genome Res 2018; 28:25–36.

606. Menschaert G, Van Criekinge W, Notelaers T, Koch A, Crappé J, Gevaert K, Van Damme P. Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. Mol Cell Proteomics 2013; 12:1780–90.

607. Young SK, Wek RC. Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response. J Biol Chem 2016; 291:16927–35.

608. Gao X, Wan J, Liu B, Ma M, Shen B, Qian S-B. Quantitative profiling of initiating ribosomes in vivo. Nat Methods 2015; 12:147–53.

609. Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, Schreiber S, Platzer M, Krawczak M, Hampe J, et al. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. Genome Res 2012; 22:2208–18.

610. Millay DP, O'Rourke JR, Sutherland LB, Bezprozvannaya S, Shelton JM, Bassel-Duby R, Olson EN. Myomaker is a membrane activator of myoblast fusion and muscle formation. Nature 2013; 499:301–5.

611. Quinn ME, Goh Q, Kurosaka M, Gamage DG, Petrany MJ, Prasad V, Millay DP. Myomerger induces fusion of non-fusogenic cells and is required for skeletal muscle development. Nat Commun 2017; 8:15665.

612. Heim EN, Marston JL, Federman RS, Edwards APB, Karabadzhak AG, Petti LM, Engelman DM, DiMaio D. Biologically active LIL proteins built with minimal chemical diversity. Proc Natl Acad Sci U S A 2015; 112:E4717-4725.

613. Samandi S, Roy AV, Delcourt V, Lucier J-F, Gagnon J, Beaudoin MC, Vanderperre B, Breton M-A, Motard J, Jacques J-F, et al. Deep transcriptome annotation enables the discovery and functional characterization of cryptic small proteins. eLife 2017; 6:e27860.

614. Asano K. Why is start codon selection so precise in eukaryotes? Translation (Austin) 2014; 2:e28387.

615. Zhang H, Wang Y, Lu J. Function and Evolution of Upstream ORFs in Eukaryotes. Trends Biochem Sci 2019; 44:782–94.

616. Erhard F, Halenius A, Zimmermann C, L'Hernault A, Kowalewski DJ, Weekes MP, Stevanovic S, Zimmer R, Dölken L. Improved Ribo-seq enables identification of cryptic translation events. Nature Methods 2018; 15:363–6.

617. Spealman P, Naik AW, May GE, Kuersten S, Freeberg L, Murphy RF, McManus J. Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. Genome Res 2018; 28:214–22.

618. Van Damme P, Gawron D, Van Criekinge W, Menschaert G. N-terminal proteomics and ribosome profiling provide a comprehensive view of the alternative translation initiation landscape in mice and men. Mol Cell Proteomics 2014; 13:1245–61.

619. Willems P, Ndah E, Jonckheere V, Stael S, Sticker A, Martens L, Van Breusegem F, Gevaert K, Van Damme P. N-terminal Proteomics Assisted Profiling of the Unexplored Translation Initiation Landscape in Arabidopsis thaliana. Mol Cell Proteomics 2017; 16:1064–80.

620. van der Horst S, Filipovska T, Hanson J, Smeekens S. Metabolite Control of Translation by Conserved Peptide uORFs: The Ribosome as a Metabolite Multisensor. Plant Physiol 2020; 182:110–22.

621. Bhushan S, Meyer H, Starosta AL, Becker T, Mielke T, Berninghausen O, Sattler M, Wilson DN, Beckmann R. Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. Mol Cell 2010; 40:138–46.

622. Rook F, Gerrits N, Kortstee A, van Kampen M, Borrias M, Weisbeek P, Smeekens S. Sucrose-specific signalling represses translation of the Arabidopsis ATB2 bZIP transcription factor gene. Plant J 1998; 15:253–63.

623. Rahmani F, Hummel M, Schuurmans J, Wiese-Klinkenberg A, Smeekens S, Hanson J. Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. Plant Physiol 2009; 150:1356–67.

624. Yamashita Y, Takamatsu S, Glasbrenner M, Becker T, Naito S, Beckmann R. Sucrose sensing through nascent peptide-meditated ribosome stalling at the stop codon of Arabidopsis bZIP11 uORF2. FEBS Lett 2017; 591:1266–77.

625. Mize GJ, Ruan H, Low JJ, Morris DR. The inhibitory upstream open reading frame from mammalian S-adenosylmethionine decarboxylase mRNA has a strict sequence specificity in critical positions. J Biol Chem 1998; 273:32500–5.

626. Johansson M, Ieong K-W, Trobro S, Strazewski P, Åqvist J, Pavlov MY, Ehrenberg M. pH-sensitivity of the ribosomal peptidyl transfer reaction dependent on the identity of the A-site aminoacyl-tRNA. Proc Natl Acad Sci USA 2011; 108:79–84.

627. Rychlík I, ČErná J, Chládek S, Pulkrábek P, ŽEmlička J. Substrate Specificity of Ribosomal Peptidyl Transferase. European Journal of Biochemistry 1970; 16:136–42.

628. Ito K, Chiba S. Arrest peptides: cis-acting modulators of translation. Annu Rev Biochem 2013; 82:171–202.

629. Wang Y, Zhang H, Lu J. Recent advances in ribosome profiling for deciphering translational regulation. Methods 2020; 176:46–54.

630. Ivanov IP, Shin B-S, Loughran G, Tzani I, Young-Baird SK, Cao C, Atkins JF, Dever TE. Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. Mol Cell 2018; 70:254-264.e6.

631. Hayes CS, Bose B, Sauer RT. Proline residues at the C terminus of nascent chains induce SsrA tagging during translation termination. J Biol Chem 2002; 277:33825–32.

632. Sunohara T, Abo T, Inada T, Aiba H. The C-terminal amino acid sequence of nascent peptide is a major determinant of SsrA tagging at all three stop codons. RNA 2002; 8:1416–27.

633. Gamble CE, Brule CE, Dean KM, Fields S, Grayhack EJ. Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. Cell 2016; 166:679–90.

634. Fijałkowska D, Verbruggen S, Ndah E, Jonckheere V, Menschaert G, Van Damme P. eIF1 modulates the recognition of suboptimal translation initiation sites and steers gene expression via uORFs. Nucleic Acids Research 2017; 45:7997–8013.

635. Acevedo JM, Hoermann B, Schlimbach T, Teleman AA. Changes in global translation elongation or initiation rates shape the proteome via the Kozak sequence. Sci Rep 2018; 8:4018.

636. Van Alstyne M, Tattoli I, Delestrée N, Recinos Y, Workman E, Shihabuddin LS, Zhang C, Mentis GZ, Pellizzoni L. Gain of toxic function by long-term AAV9-mediated SMN overexpression in the sensorimotor circuit. Nature Neuroscience 2021; :1–11.

637. Geisler A, Jungmann A, Kurreck J, Poller W, Katus HA, Vetter R, Fechner H, Müller OJ. microRNA122-regulated transgene expression increases specificity of cardiac gene transfer upon intravenous delivery of AAV9 vectors. Gene Therapy 2011; 18:199–209.

638. Cawood R, Chen HH, Carroll F, Bazan-Peregrino M, Rooijen N van, Seymour LW. Use of Tissue-Specific MicroRNA to Control Pathology of Wild-Type Adenovirus without Attenuation of Its Ability to Kill Cancer Cells. PLOS Pathogens 2009; 5:e1000440.

639. Fischer A, Manske K, Seissler J, Wohlleber D, Simm N, Buerck LW, Knolle P, Schnieke A, Fischer K. Cytokine-inducible promoters to drive dynamic transgene expression: The "Smart Graft" strategy. Xenotransplantation 2020; 27:e12634.

640. Joshi J, Rubart M, Zhu W. Optogenetics: Background, Methodological Advances and Potential Applications for Cardiovascular Research and Medicine. Front Bioeng Biotechnol 2019; 7:466.

641. Li T, Chen X, Qian Y, Shao J, Li X, Liu S, Zhu L, Zhao Y, Ye H, Yang Y. A synthetic BRET-based optogenetic device for pulsatile transgene expression enabling glucose homeostasis in mice. Nat Commun 2021; 12:615.

642. Moncion A, Harmon JS, Li Y, Natla S, Farrell EC, Kripfgans OD, Stegemann JP, Martín-Saavedra FM, Vilaboa N, Franceschi RT, et al. Spatiotemporally-controlled transgene expression in hydroxyapatite-fibrin composite scaffolds using high intensity focused ultrasound. Biomaterials 2019; 194:14–24.

643. Das AT, Tenenbaum L, Berkhout B. Tet-On Systems For Doxycycline-inducible Gene Expression. Curr Gene Ther 2016; 16:156–67.

644. Wittmann A, Suess B. Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. Mol Biosyst 2011; 7:2419–27.

645. Beilstein K, Wittmann A, Grez M, Suess B. Conditional control of mammalian gene expression by tetracycline-dependent hammerhead ribozymes. ACS Synth Biol 2015; 4:526–34.

646. Yokobayashi Y. Aptamer-based and aptazyme-based riboswitches in mammalian cells. Current Opinion in Chemical Biology 2019; 52:72–8.

647. Liu Y, Li J, Chen Z, Huang W, Cai Z. Synthesizing artificial devices that redirect cellular information at will. eLife 2018; 7:e31936.

648. Forand A, Muchir A, Mougenot N, Sevoz-Couche C, Peccate C, Lemaitre M, Izabelle C, Wood M, Lorain S, Piétri-Rouxel F. Combined Treatment with Peptide-Conjugated Phosphorodiamidate Morpholino Oligomer-PPMO and AAV-U7 Rescues the Severe DMD Phenotype in Mice. Mol Ther Methods Clin Dev 2020; 17:695–708.

649. Figueiredo M. SMA Trial to Test Spinraza in Children Previously Given Zolgensma [Internet]. [cited 2021 Jun 19]; Available from: https://smanewstoday.com/news-posts/2020/07/22/biogen-to-launch-first-phase-4-trial-spinraza-children-previously-given-zolgensma/

650. Mendell JR, Sahenk Z, Lehman K, Nease C, Lowes LP, Miller NF, Iammarino MA, Alfano LN, Nicholl A, Al-Zaidy S, et al. Assessment of Systemic Delivery of rAAVrh74.MHCK7.micro-dystrophin in Children With Duchenne Muscular Dystrophy: A Nonrandomized Controlled Trial. JAMA Neurol 2020; 77:1122–31.

651. Alhaji SY, Ngai SC, Abdullah S. Silencing of transgene expression in mammalian cells by DNA methylation and histone modifications in gene therapy perspective. Biotechnology and Genetic Engineering Reviews 2019; 35:1–25.

652. Chen F, Alphonse M, Liu Q. Strategies for nonviral nanoparticle-based delivery of CRISPR/Cas9 therapeutics. Wiley Interdiscip Rev Nanomed Nanobiotechnol 2020; 12:e1609.

653. Padhy SK, Takkar B, Narayanan R, Venkatesh P, Jalali S. Voretigene Neparvovec and Gene Therapy for Leber's Congenital Amaurosis: Review of Evidence to Date. Appl Clin Genet 2020; 13:179–208.

654. Maeder ML, Stefanidakis M, Wilson CJ, Baral R, Barrera LA, Bounoutas GS, Bumcrot D, Chao H, Ciulla DM, DaSilva JA, et al. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. Nat Med 2019; 25:229–33.

655. de Jong OG, Murphy DE, Mäger I, Willms E, Garcia-Guerra A, Gitz-Francois JJ, Lefferts J, Gupta D, Steenbeek SC, van Rheenen J, et al. A CRISPR-Cas9-based reporter system for single-cell detection of extracellular vesicle-mediated functional transfer of RNA. Nat Commun 2020; 11:1113.

656. Wiklander OPB, Brennan MÁ, Lötvall J, Breakefield XO, El Andaloussi S. Advances in therapeutic applications of extracellular vesicles. Sci Transl Med 2019; 11:eaav8521.

657. O'Brien K, Breyne K, Ughetto S, Laurent LC, Breakefield XO. RNA delivery by extracellular vesicles in mammalian cells and its applications. Nat Rev Mol Cell Biol 2020; 21:585–606.

658. Elsharkasy OM, Nordin JZ, Hagey DW, de Jong OG, Schiffelers RM, Andaloussi SE, Vader P. Extracellular vesicles as drug delivery systems: Why and how? Advanced Drug Delivery Reviews 2020; 159:332–43.

659. Shao H, Im H, Castro CM, Breakefield X, Weissleder R, Lee H. New Technologies for Analysis of Extracellular Vesicles. Chem Rev 2018; 118:1917–50.

660. Winkelsas AM, Grunseich C, Harmison GG, Chwalenia K, Rinaldi C, Hammond SM, Johnson K, Bowerman M, Arya S, Talbot K, et al. Targeting the 5′ untranslated region of SMN2 as a therapeutic strategy for spinal muscular atrophy. Molecular Therapy - Nucleic Acids 2021; 23:731–42.

661. Qu L, Yi Z, Zhu S, Wang C, Cao Z, Zhou Z, Yuan P, Yu Y, Tian F, Liu Z, et al. Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. Nat Biotechnol 2019; 37:1059–69.

662. Merkle T, Merz S, Reautschnig P, Blaha A, Li Q, Vogel P, Wettengel J, Li JB, Stafforst T. Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. Nat Biotechnol 2019; 37:133–8.

663. Skelton MR, Schaefer TL, Graham DL, Degrauw TJ, Clark JF, Williams MT, Vorhees CV. Creatine transporter (CrT; Slc6a8) knockout mice as a model of human CrT deficiency. PLoS One 2011; 6:e16187.

664. Ghosh A, Khandelwal N, Kumar A, Bera AK. Leucine-rich repeat-containing 8B protein is associated with the endoplasmic reticulum Ca2+ leak in HEK293 cells. J Cell Sci 2017; 130:3818–28.

8 Appendix

8.1 Manuscripts accepted for publication

Hanson B, Wood MJA, Roberts TC. Molecular correction of Duchenne muscular dystrophy by splice modulation and gene editing. RNA Biol 2021; 20:1–15. DOI: 10.1080/15476286.2021.1874161.

Marsh S, **Hanson B**, Wood MJA, Varela MA, Roberts TC. Application of CRISPR-Cas9-Mediated Genome Editing for the Treatment of Myotonic Dystrophy Type 1. Molecular Therapy 2020; 28:2527–39. DOI: 10.1016/j.ymthe.2020.10.005.

van Westering TLE, Johansson HJ, **Hanson B**, Coenen-Stass AML, Lomonosova Y, Tanihata J, Motohashi N, Yokota T, Takeda S, Lehtiö J, et al. Mutation-independent Proteomic Signatures of Pathological Progression in Murine Models of Duchenne Muscular Dystrophy. Molecular & Cellular Proteomics 2020; 19:2047–68. DOI: 10.1074/mcp.RA120.002345.

Coenen-Stass AML, Pauwels MJ, **Hanson B**, Perez CM, Conceição M, Wood MJA, Mäger I, Roberts TC. Extracellular microRNAs exhibit sequence-dependent stability and cellular release kinetics. RNA Biology 2019; 16:696–706. DOI: 10.1080/15476286.2019.1582956.

8.2 Conferences attended

Virtual Genome Engineering in Oxford seminar held by the University of Oxford, Oxford, UK. December 2020. (Co-organiser and chair)

Genome Engineering in Oxford seminar held at the Dunn School of Pathology, University of Oxford, UK. October 2019. (Oral presentation)

Virtual Genome Engineering: Frontiers of CRISPR/Cas meeting held by Cold Spring Harbour, New York, USA. August 2020. (Poster presentation)

MDUK Oxford Neuromuscular Centre 1st Annual Meeting held at St Peter's College, University of Oxford, UK. February 2020. (Poster presentation)

Genome Engineering: Frontiers of CRISPR/Cas meeting in Cold Spring Harbour, New York, USA. October 2019. (Poster presentation)

3rd Annual Genome Editing Congress held at the Novotel London West Hotel, London, UK. November 2017. (Attendee)