The role of MBNL1 and MBNL2 in nuclear foci formation and repeat expansion transcript retention in Myotonic Dystrophy 1

Xiaomeng Xing; BSc

School of Life Sciences

University of Nottingham

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

March 2023

Principal supervisor – Prof. J David Brook

Table of Contents

Table of Contents i
List of Figures v
List of Tables viii
List of Abbreviations ix
Acknowledgements xiii
Abstract xv
Chapter 1 Introduction 1
1.1 History of Myotonic Dystrophy 1
1.2 Multisystem effects of DM 2
1.3 Genetic basis of myotonic dystrophy 5
1.3.1 The genetic basis of DM type 1 5
1.3.2 The genetic basis of DM type 2 7
1.4 The role of MBNL, CELF and other AS factors involved in DM pathogenesis 9
1.4.1 MBNL domains important for RNA binding and splicing regulation 11
1.4.2 Altered MBNL and CELF1 functions in DM 15
1.4.3 $[C(C)UG]^{exp}$ RNA-protein Interactions in RNA Foci Formation 17
1.5 Therapies for Myotonic Dystrophy 22
1.5.1 Targeting the expanded CTG repeats in the *DMPK* gene 22
1.5.2 Targeting the toxic RNA 24
1.5.3 Targeting protein-RNA binding 29
1.5.4 Altering levels of RNA binding proteins 30
1.5.5 Targeting the mis-spliced genes 32
1.5.6 Summary 35
1.6 RNA decay 36
1.6.1 Mechanisms of normal mRNA degradation 36
1.6.2 Mechanisms of aberrant mRNA degradation in the cytoplasm 40
1.6.3 Nuclear RNA decay 43
1.6.4 UPF1: beyond nonsense-mediated decay 46

1.7 Super-Resolution Microscopy 48
1.7.1 Fluorophores for SMLM 49
1.7.2 Optics of SMLM 49
1.7.3 Acquiring data for SMLM 50
1.7.4 Cluster analysis of data in SMLM 51
1.8 Aim of study ..... **Error! Bookmark not defined.**
Chapter 2 Materials and Methods 56
2.1 Tissue culture 56
2.2 Cell lines 56
2.3 Media, solutions and antibiotics 56
2.3.1 Cell culture 56
2.3.2 Cells passaging 56
2.3.3 Cell count 57
2.3.4 Cryopreservation of cells 57
2.3.5 Defrosting cells from liquid nitrogen 58
2.4 RNA Fluorescence In Situ Hybridisation (FISH) 58
2.4.1 FISH using 96-well Plates and Celldiscoverer 7 System 58
2.4.2 FISH using 6-well plates and Super Resolution Microscopy 59
2.5 Immunocytochemistry (ICC) 61
2.6 Combined immunocytochemistry and in-situ hybridisation (ICC + FISH) 62
2.7 Molecular Biology Techniques 63
2.7.1 PCR Methods 63
2.7.2 Gel Electrophoresis 65
2.7.3 Gel Extraction and Purification 65
2.7.4 DNA Sequencing 65
2.7.5 DNA Extraction 66
2.7.6 Total RNA Extraction 66
2.7.7 Nuclear and Cytoplasmic RNA Extraction 68
2.7.8 Poly(A)$^+$ mRNA Extraction 68
2.7.9 5' cap-dependent RNA Extraction 69
2.7.10 Quantification of DNA and RNA 69

2.7.11 RNA quality assessment using the Bioanalyzer 69
2.7.12 Reverse Transcriptase PCR 69
2.8 Protein Analysis 70
2.8.1 Solutions, Buffers and Antibodies 70
2.8.2 Total Protein Extraction 73
2.8.3 Nuclear and Cytoplasmic Protein Extraction 73
2.8.4 Protein Quantification 73
2.8.5 Western Blot 74
2.9 Genome Engineering by CRISPR/Cas9 System 75
2.9.1 Target Genes 75
2.9.2 gRNAs Design 75
2.9.3 *In vitro* Assessment of the gRNA and Cas9 Activities 75
2.10 shRNA lentiviral transduction 77
2.11 Microscopy 80
2.11.1 Micro Widefield Plate Reader 80
2.11.2 Zeiss Celldiscoverer 7 Microscope 80
2.11.3 Confocal Microscopy Zeiss LSM880F 80
2.11.4 Zeiss Elyra Super Resolution Microscope 80
2.12 Electronic Databases 82
2.13 Nanopore Sequencing 82
Chapter 3 Double knock-out of *MBNL1* and *MBNL2* in DM1 fibroblast cells 84
3.1 MBNL gRNAs 85
3.2 Transfection of cells and expansion of clones 87
3.3 Determining the efficiency of genome editing 87
3.4 *MBNL1* and *MBNL2* Knock-out assessment by Western blotting 88
3.5 Effect of *MBNL1* and *MBNL2* double knock-out on nuclear foci 92
3.6 Effect of *MBNL* knock-out on the abundance of $(CUG)^{exp}$ *DMPK* transcripts 93
3.6.1 Method 1: BpmI restriction enzyme digestion 98
3.6.2 Method 2: Nanoplate Digital PCR 99
3.6.3 Discussion: 101

3.7 Effect of *MBNL1* and *MBNL2* double knock-out on the distribution of residual MBNL proteins in DM1 cells 102
3.7.1 Results 103
3.7.2 Discussion 105
3.8 Effect of *MBNL1* and *MBNL2* double knock-out on Alternative splicing and gene expression 111
3.9 Re-expression of MBNL1 and 2 135
3.10 Selective Nanopore sequencing of human *DMPK* by Cas9-assisted targeting of chromosome segments (CATCH) 138
3.11 Discussion and conclusions 142
Chapter 4 Detection of *DMPK* mRNA using Super-Resolution Single Molecule Localisation Microscopy (SMLM) 145
4.1 CUG repeat single-labelling 147
4.2 Dual labelling the CUG repeats and single copy 5' or 3' sequences 153
4.3 Discussion and conclusions 167
Chapter 5 Identification of RNA decay factors involved in the degradation of $(CUG)^{exp}$ *DMPK* transcripts 170
5.1 Poly(A) tail and 5' cap structure status of the $(CUG)^{exp}$ *DMPK* transcripts 172
5.2 Knock-down of RNA decay factors 176
5.3 UPF1 over-expression 184
5.4 Discussion and conclusions 188
Chapter 6 Final Conclusions 193
References 201
Appendix 220

List of Figures

Figure 1-1 Myotonia is a characteristic skeletal muscle feature of DM patients. 2
Figure 1-2 Myotonic dystrophy is a multisystemic disorder. 4
Figure 1-3 DM1- and DM2-associated gene loci. 6
Figure 1-4 DM1 pedigree highlights genetic anticipation. 7
Figure 1-5 Clinical manifestations and disease stages in DM1, DM2, and CDM. 8
Figure 1-6 Mis-regulated alternative splicing and mis-regulated alternative polyadenylation model in human DM1. 10
Figure 1-7 MBNL functions in RNA processing. 11
Figure 1-8 MBNL gene and ZnF structures. 18
Figure 1-9 MBNL sequestration on the CUG expansion transcript. 20
Figure 1-10 RNAi-based therapeutic strategies. 25
Figure 1-11 An overview of cytoplasmic poly(A) + mRNA decay pathways. 39
Figure 1-12 Surveillance for faulty mRNA processing that occurs in the cytoplasm. 42
Figure 1-13 Principle of single-molecule localisation microscopy (SMLM). 52
Figure 3-1 Overview of the CRISPR/Cas9 in vitro assay protocol. 86
Figure 3-2 Genome Editing of the Human MBNL Locus. 90
Figure 3-3 Assessment of double knock-out of MBNL1 and 2 by Western blotting 91
Figure 3-4 Effect of MBNL1 and 2 double knock-out on RNA foci. 92
Figure 3-5 Mechanisms of detecting DMPK transcripts. 95
Figure 3-6 Assessment of nuclear and cytoplasmic RNA quality. 96
Figure 3-7 Assessment of the effect of MBNL1 and 2 double knock-out on the $(CUG)^{exp}$ DMPK transcript via BmpI digestion method. 98
Figure 3-8 Assessment of the effect of MBNL1 and 2 double knock-out on the $(CUG)^{exp}$ DMPK transcript via dPCR method. 100
Figure 3-9 Gene structures of MBNL1 and MBNL2. 102
Figure 3-10 Sub-cellular localisation of MBNL1 and 2 proteins. 104
Figure 3-11 Effect of MBNL1 and 2 double knock-out on the MBNL1 and 2 mRNA compositions 107
Figure 3-12 Speculation of the MBNL mRNA compositions. 110
Figure 3-13 Effect of MBNL1 and 2 double knock-out on the INSR and NFIX alternative splicing. 114

Figure 3-14 Effect of MBNL1 and 2 double knock-out on the BIN1 alternative splicing. 118

Figure 3-15 Schematic representation of MTMR1 alternative splicing in skeletal muscle..... 123

Figure 3-16 Aberrant gene expression identified in KB, P4B6 and P4C1 via RNA-Seq data analysis. 127

Figure 3-17 Vitamin D deficiency in DM1. 134

Figure 3-18 Re-expression of MBNL1 and 2. 136

Figure 3-19 Assessment of the sgRNA activity by Cas9 in vitro assays. 139

Figure 3-20 Distribution of the length of CTG repeats in DM1 cell lines. 140

Figure 4-1 Scheme of the probe design for the stochastic optical reconstruction microscopy (STORM). 147

Figure 4-2 Detecting the CUG repeat expansion in the DMPK transcript with STROM. 149

Figure 4-3 Quantification of the number of localisations in $(CUG)^{exp}$ clusters. ... 151

Figure 4-4 Quantification of the number of micro $(CUG)^{exp}$ foci in KB, SB, P4C1 and P4B6 (n=6). 153

Figure 4-5 Detecting the CUG repeat region and its 5' end flanking regions in the DMPK transcript with STORM. 154

Figure 4-6 Detecting the CUG repeat region and its 3' end flanking regions in the DMPK transcript with STORM. 155

Figure 4-7 Quantification of the number of localisations in $(CUG)^{exp}$ clusters and in 5' end clusters. 158

Figure 4-8 Quantification of the number of localisations in $(CUG)^{exp}$ clusters and in 3' end clusters. 160

Figure 4-9 Quantification of the number of clusters identified with 5' or 3' probe sets in KB, SB and P4C1 (n=6). 163

Figure 4-10 Detection of $(CUG)^{exp}$ cluster and 5' or 3' end cluster co-localisations. 165

Figure 5-1 Assessment of poly(A) tail and 5' cap structure status of the $(CUG)^{exp}$ DMPK transcripts via the BpmI digestion method. 174

Figure 5-2 Assessment of poly(A) tail and 5' cap structure status of the $(CUG)^{exp}$ DMPK transcripts via the dPCR method. 175

Figure 5-3 Assessment of the RNA decay factor knock-down by western blotting. 176

Figure 5-4 Quantification of the protein levels and knock-down efficiency. 177

Figure 5-5 Effect of the knock-down (KD) of decay factors on RNA foci. 179

Figure 5-6 Effect of the knock-down of RNA decay factors on the mutant DMPK transcripts. 183

Figure 5-7 Assessment of the UPF1 over-expression by western blotting. 184

Figure 5-8 Effect of UPF1 over-expression (OEx) on RNA foci. 186

Figure 5-9 Effect of the UPF1 over-expression (OEx) on the mutant DMPK transcripts. 188

Figure 5-10 Schematic illustration depicting the possible mechanisms involved in the nuclear export and degradation of mutant DMPK mRNA. 192

Figure 0-1 Assessment of double knock-out of MBNL1 and 2 by Western blotting in triplicate. 220

Figure 0-2 Assessment of the effect of MBNL1 and 2 double knock-out on the $(CUG)^{exp}$ DMPK transcript via BmpI digestion method (in duplicate). 221

Figure 0-3 Subcellular Localisation of MBNL1 and 2 proteins detected by Western blotting in triplicate. 222

Figure 0-4 Signal detection in STORM microscopy by using the scramble probe. 223

Figure 0-5 Assessment of the knock-down efficiency of RNA decay factors by Western blotting in triplicate. 224

Figure 0-6 Assessment of the efficiency of UPF1 over-expression by Western blotting in triplicate. 226

Figure 0-7 Assessment of nuclear and cytoplasmic RNA quality for RNA samples utilised in Chapter 5. 227

Figure 0-8 Assessment of poly(A) tail and 5' cap structure status of the $(CUG)^{exp}$ DMPK transcript via the BpmI digestion method. 231

Figure 0-9 Probes positioning in the DMPK cDNA. 232

List of Tables

Table 2-1 List of probes used in the super-resolution microscopy experiments. .. 60

Table 2-2 dPCR setup components. 64

Table 2-3 DNase treatment components. 67

Table 2-4 List of primary and secondary antibodies used in this thesis 71

Table 2-5 List of lentiviral particles used in this thesis. 78

Table 3-1 gRNA sequences targeting MBNL1 and MBNL2. 86

Table 3-2 List of the colonies screened and the mutations present in each of them. 88

Table 3-3 Aberrantly spliced exons of MBNL1 and 2 identified in P4C1 and P4B6. 109

Table 3-4 Aberrantly spliced exons and introns identified in KB, P4B6 and P4C1 via RNA-Seq data analysis. 123

List of Abbreviations

ABLIM1	Actin-binding LIM protein 1
Ago2	Argonaute 2
ASOs	Antisense oligonucleotides
ATP11C	ATPase phospholipid transporting 11C
BIN1	Bridging integrator 1
biNLS	Bipartite nuclear localisation signal
BSA	Bovine serum albumin
CACNA1S	Calcium channel voltage-dependent L type alpha 1S subunit
CATCH	Cas9-assisted targeting of chromosome segments
CDK	Cyclin D3–dependent kinase
CDM	Congenital myotonic dystrophy
CELF1	Elav-like family member 1
circRNA	Circular RNA
CLCN1	Chloride channel protein 1
CNBP	CCHC-type zinc finger nucleic acid binding protein
conNLS	Conformational nuclear localisation signal
$(CUG)^{exp}$ *DMPK*	exp*DMPK*= mutant *DMPK*
DIS3	Exosome complex exonuclease RRP44
DM	Myotonic dystrophy
DMD	Duchenne muscular dystrophy
DMEM	Dulbecco's modified Eagle medium
DMPK	Dystrophia myotonica protein kinase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dPCR	Digital PCR
DSB	Double-strand break

EDTA	Ethylenediaminetetraacetic acid
EJC	Exon junction complex
EXOSC10	Exosome component 10
FISH	Fluorescence in situ hybridisation
FRET	Fluorescence resonance energy transfer
FXS	Fragile X syndrome
HD	Huntington disease
HDR	Homology directed repair
ICC	Immunocytochemistry
INSR/IR	Insulin receptor
KD	Knock-down
KO	Knock-out
LNA	locked nucleic acid
MBNL	Muscleblind-like
MEF	Myocyte enhancer binding factor
miRNA	microRNA
MTMR1	Myotubularin-related 1
NC1	Nucleocytoplasmic index
NF1 gene	Neurofibromin 1
NFIX	Nuclear factor IX
NHEJ	Non-homologous end joining
NLS	Nuclear localisation signal
NMD	Nonsense-mediated decay
NPC	Nuclear pore complex
NSD	Non-stop decay
OEx	Over-expression
PABP	Poly(A)-binding protein
PALM	Photoactivated localisation microscopy
PAM	Protospacer adjacent motif
pAs	Poly(A) signals

PBS	Phosphate buffered saline
PBS-T	PBS + 0.1% Tween-20
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PKC	Protein kinase C
PSF	Point spread function
PTC	Premature termination codon
qPCR	Quantitative PCR
RAN translation	Repeat associated non-AUG translation
RBFOX	RNA binding fox-1
RBP	RNA binding protein
RIN	RNA integrity number
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
RNP	Ribonucleoprotein
RT	Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
SCA	Spinocerebellar ataxia
SDS	Sodium dodecyl sulphate
SFPQ	Splicing factor proline and glutamine rich
sgRNA	Single guide ribonucleic acid
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SMD	Staufen-Mediated mRNA Decay
SMLM	Single molecule localisation microscopy
SNP	Single nucleotide polymorphism
SSC	Saline sodium citrate
STAU1	Staufen 1
STORM	Stochastic optical reconstruction microscopy

STR	Short tandem repeat
SURF	SMG-1–UPF1–eRF1–eRF3 complex
TALEN	Transcription activator-like effector nuclease
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline with 1% Tween 20
UPF1	Regulator of nonsense transcripts 1
UTR	Untranslated region
VDR	Vitamin D receptor
VRC	Vanadyl ribonucleoside complex
WT/wt	Wild-type
XRN2	5'-3' Exoribonuclease 2
ZnF	Zinc finger

Acknowledgements

Completing a PhD is undoubtedly an incredible achievement, and it would not have been possible without the help and support of numerous individuals throughout my four-year journey. I am sincerely grateful to my supervisor, Prof. David Brook, for giving me the opportunity to join his lab and for providing unwavering support and guidance from day one. His expertise and mentorship have been invaluable throughout my research, and I feel fortunate to have had such an incredible supervisor who is always available to help and provide encouragement.

I would also like to express my heartfelt gratitude to all the past and current members of Brook's lab who have been an integral part of my academic journey. Your friendship, knowledge sharing, and encouragement have undoubtedly made me a better scientist and a better person. Special thanks to Tushar, Sarah, Jose, Tasabeeh, Marzena, Philippine, Anjani, Reem, Gareth, Jessy, Ayushri, and Erik for being such great colleagues and friends.

I am deeply indebted to Dr. Robert Markus, whose outstanding teaching and expertise have made a significant impact on this thesis. His insights and advice have been invaluable in shaping the direction of my research. I would also like to extend my thanks to Tushar again for his invaluable help and advice on western blots, and to Sarah for always ensuring the smooth and successful execution of my experiments. Additionally, I am grateful to the colleagues from the University of Nottingham Next Generation Sequencing Facility for their invaluable assistance with sequencing services and data analysis.

It cannot be expressed with words how grateful I am to Jacob, my boyfriend. He is the best partner someone could ever have and an endless source of encouragement, resilience and positivism on all

those difficult moments I have had to go through along the years. I also greatly appreciate my parents who have unconditionally supported me all these years. Regardless of the distance, they have always been there for me whenever I needed them.

Last but not least, I would like to express my gratitude to The University of Nottingham, The Myotonic Dystrophy Foundation, The Alan Webster Legacy Fund and The Myotonic Dystrophy Support Group which have funded my work during these 4 years.

Abstract

Myotonic dystrophy (*dystrophia myotonica*, DM) is a multisystemic disorder that is dominantly inherited. There are two types of DM: DM1, which is caused by a CTG repeat expansion in the 3' UTR of the *DMPK* gene, and DM2, which is caused by a CCTG repeat expansion in intron 1 of the *CNBP* gene. *In vitro* experiments have shown that once transcribed, expanded C(C)UG repeats fold into very stable and long RNA hairpin structures that sequester MBNLs and other nuclear factors to form ribonuclear foci, in both types of DM. It is important to note that MBNLs are not just passively trapped within $[C(C)UG]^{exp}$, but also actively participate in foci formation and cause nuclear retention of $[C(C)UG]^{exp}$ transcripts. The sequestration of MBNLs directly impairs the splicing of several key regulatory pre-mRNAs in muscles and neural cells. Missplicing of muscle chloride channel (CLCN1), calcium channel voltage-dependent L type alpha 1S subunit (CACNA1S), bridging integrator 1 (BIN1) and insulin receptor (INSR), results in myotonia, muscle weakness and insulin resistance, respectively, which are key hallmarks of DM.

This study aimed to investigate the dynamics of MBNL1 and MBNL2 co-localisation with $(CUG)^{exp}$ *DMPK* transcripts to shed light on the mechanisms underlying the aggregation, nuclear retention and degradation of mRNA containing CUG-triplet repeats. To achieve this, I conducted the double knock-out of *MBNL1* and *MBNL2* in DM1 fibroblast cells using the CRISPR/Cas9 system, which led to the loss of full-length MBNL1 and 2 proteins. The effects of *MBNL1* and *2* double knock-out on DM1 are complex and multi-faceted. It exacerbated many mis-splicing events and the mis-regulation of expression of many genes implicated in the disease. Additionally, it led to the emergence of novel aberrant splicing events and dysregulated gene expression. Furthermore, the double-inactivation

of MBNL1 and 2 caused the instability of RNA foci formed by the mutant *DMPK* transcript, facilitating the dissipation of RNA foci and the degradation or nuclear export of the mutant transcript.

To gain a better understanding of the dynamics of $(CUG)^{exp}$ RNA foci, I utilised super-resolution single-molecule localisation microscopy (SMLM) to detect *DMPK* transcripts. I designed probes from three intervals of the *DMPK* mRNA which allowed me to determine signals from the regions 5' and 3' to the repeat expansion and from the repeat itself. My results uncovered several key findings. First, I discovered large $(CUG)^{exp}$ foci are primarily formed by multiple *DMPK* transcripts, while micro $(CUG)^{exp}$ foci are mainly formed by single *DMPK* transcripts, suggesting that micro $(CUG)^{exp}$ foci may be the intermediate degradation products of large $(CUG)^{exp}$ foci. Second, I found that in the absence of MBNL, mutant *DMPK* transcripts can only form micro $(CUG)^{exp}$ foci, which rapidly dissipate or export to the cytoplasm. Third, I observed that micro $(CUG)^{exp}$ foci with or without 5' or 3' flanking regions are present in the DM1 cytoplasm, potentially leading to the source of RAN translation in DM1.

I also examined the participation of various RNA decay factors in DM1 cells and their derivatives through shRNA lentiviral transduction. The successful knock-down of *XRN2*, *EXOSC10*, *UPF1* and *STAU1* produced a slight but significant increase in the foci area and the abundance of mutant *DMPK* mRNA in DM1 cells. Notably, a more pronounced increase was observed in MBNL-deficient DM1 cells. These findings highlight the critical role played by various RNA decay pathways in reducing the accumulation of mutant *DMPK* RNA. Furthermore, the sequestration of MBNLs to the mutant RNA may impede its degradation process, as the absence of MBNLs appears to enhance the efficiency of these pathways.

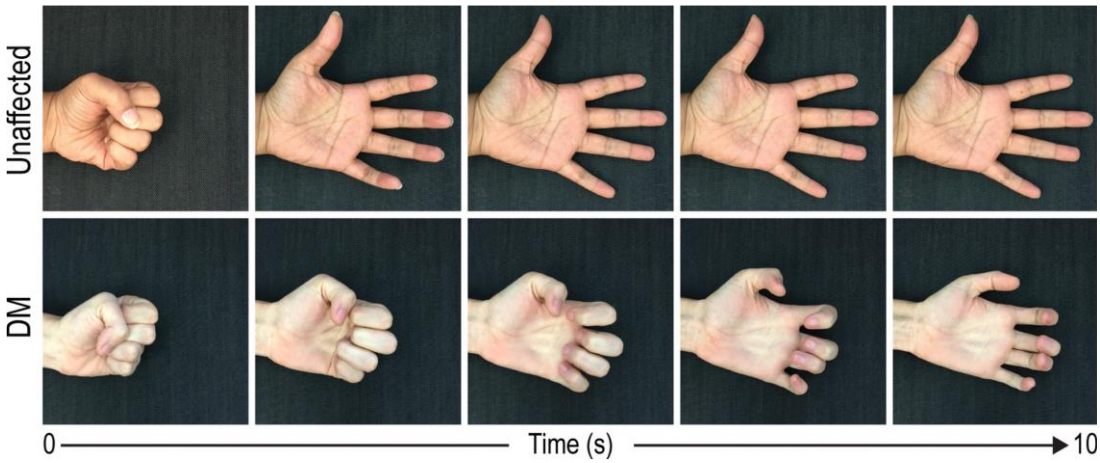
In summary, my investigations uncovered valuable data on RNA foci dynamics in DM1, revealing the intricate mechanisms that underlie their formation, stability, and turnover. In addition, I contributed to delineate the complex pathways involved in the transportation and degradation of the mutant mRNA and provided insights into the critical role played by MBNL proteins in these processes.

Chapter 1 Introduction

1.1 History of myotonic dystrophy (DM)

In 1909, Hans Steinert reported on a group of patients who displayed myotonia (Figure 1-1) and progressive muscle wasting with involvement of multiple systems. This led to the designation of the disorder as Steinert's disease [1, 2]. Despite the autosomal dominant inheritance pattern, the genetic basis of the disease was not clear for several decades due to its variable presentation. This was due in part because, unlike disorders such as Duchenne muscular dystrophy (DMD) which is associated with direct links between protein loss-of-function and disruption of tissue homeostasis, consolidating the molecular mechanisms governing the pleiotropy of Steinert's disease was, and remains a complex task [3]. Furthermore, the variable types of symptoms, observation of increased symptom severity and decreased age-of-onset in successive generations of affected families, underscored the complex and variable nature of this disease [2, 4, 5]. As research on Steinert's disease continued throughout the 20th century, the disorder was eventually renamed myotonic dystrophy, or DM, after its hallmark muscle symptoms. DM is one of the most striking examples of anticipation, whereby affected individuals in successive generations display earlier onset and increased severity of symptoms. For example, a mutation-harbouring grandparent may be primarily asymptomatic while her daughter presents with adult-onset disease that is not diagnosed as DM until the birth of her severely affected congenital infant, the proband. Indeed, the birth of a child with congenital myotonic dystrophy (CDM) is a common impetus for the evaluation of other members of affected families [5].

Figure 1-1 Myotonia is a characteristic skeletal muscle feature of DM patients.



Normally, grip relaxation is unencumbered and accompanied by muscle repolarization to resting potential (upper panels). For DM patients, loss of ion homeostasis causes delayed muscle relaxation (lower panels). Figure adapted from [6].

1.2 Multisystem effects of DM

Myotonic dystrophy affects both sexes and can manifest at any age, from infancy to old age. The multisystemic effects of DM are numerous (Figure 1-2), and their severity and frequency can vary depending on the subtype and age of onset of the disease.

Muscular System

The muscular system is the most commonly affected system in DM. Progressive muscle weakness and wasting are the hallmark features of the disease [7, 8]. The muscle weakness typically starts in the distal extremities and progresses to the proximal muscles. Patients with DM can experience significant muscle atrophy and contractures, which can lead to joint stiffness and limited mobility [9]. Myotonia, which is the inability to relax the muscles after contraction, is another hallmark feature of DM [8]. Myotonia can affect the muscles of the face, neck, and limbs, causing difficulty with speech, swallowing, and fine motor tasks.

Cardiovascular System

Cardiac abnormalities in DM and can result in various complications, including arrhythmias, conduction defects, and cardiomyopathy [10-12]. The incidence of sudden death due to cardiac arrhythmias is significantly increased in patients with DM, particularly those with DM1 [13, 14]. The presence of conduction defects, such as atrioventricular block, can also increase the risk of sudden death [12]. Cardiomyopathy, which is the abnormal thickening or thinning of the heart muscle, is another common complication in DM [15]. Cardiomyopathy can lead to heart failure, which is a significant cause of morbidity and mortality in DM [15, 16].

Respiratory System

Respiratory involvement in DM can manifest as respiratory muscle weakness, which can lead to hypoventilation, respiratory failure, and pneumonia [17, 18]. The severity of respiratory involvement can vary depending on the subtype of DM and the age of onset of the disease [19]. Patients with DM1, which is the more severe subtype, are at a higher risk of respiratory complications [20].

Gastrointestinal System

Gastrointestinal symptoms are common in DM and can include dysphagia, constipation, and abdominal pain [21-23]. Dysphagia, or difficulty swallowing, can result from muscle weakness in the muscles of the throat and oesophagus [24, 25]. Constipation and abdominal pain can result from muscle weakness in the abdominal muscles and smooth muscle of the gastrointestinal tract [21].

Endocrine System

Endocrine abnormalities in DM can include insulin resistance, hypogonadism, and thyroid dysfunction [26-28]. Insulin resistance can lead to glucose intolerance and diabetes mellitus, which is a

common complication in DM [29]. Hypogonadism, or the underproduction of sex hormones, can result in decreased libido, erectile dysfunction, and infertility [30, 31]. Thyroid dysfunction can result in hypothyroidism, which can cause fatigue, weight gain, and other symptoms [26, 32].

Central Nervous System

Central nervous system involvement in DM can result in various cognitive and behavioural abnormalities [33, 34]. Cognitive impairment is common in DM, particularly in patients with DM1 [35, 36]. Behavioural abnormalities, such as apathy, disinhibition, and emotional lability, can also occur in DM [34, 37].

Figure 1-2 Myotonic dystrophy is a multisystemic disorder.

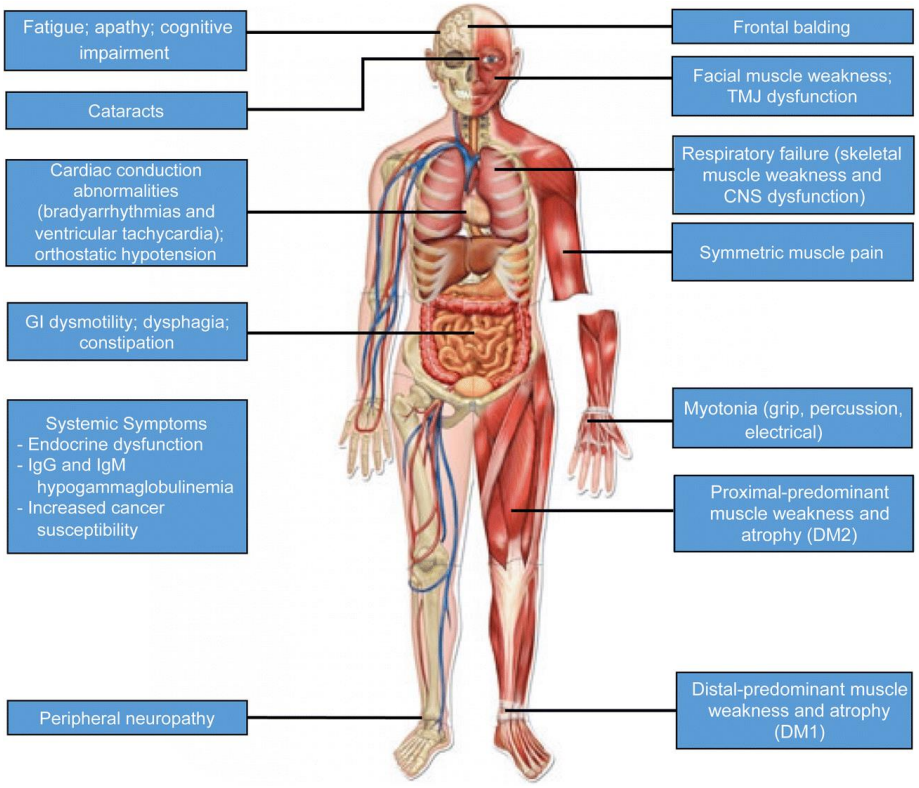


Figure adapted from [38].

Myotonic dystrophy is a multisystemic disorder that can affect various organ systems, including the muscular, cardiovascular, respiratory, gastrointestinal, endocrine, and central nervous systems. The severity and frequency of multisystemic involvement can vary depending on the subtype and age of onset of the disease. The management of DM requires a multidisciplinary approach, with the involvement of various specialists, including neurologists, cardiologists, pulmonologists, gastroenterologists, endocrinologists, and psychiatrists.

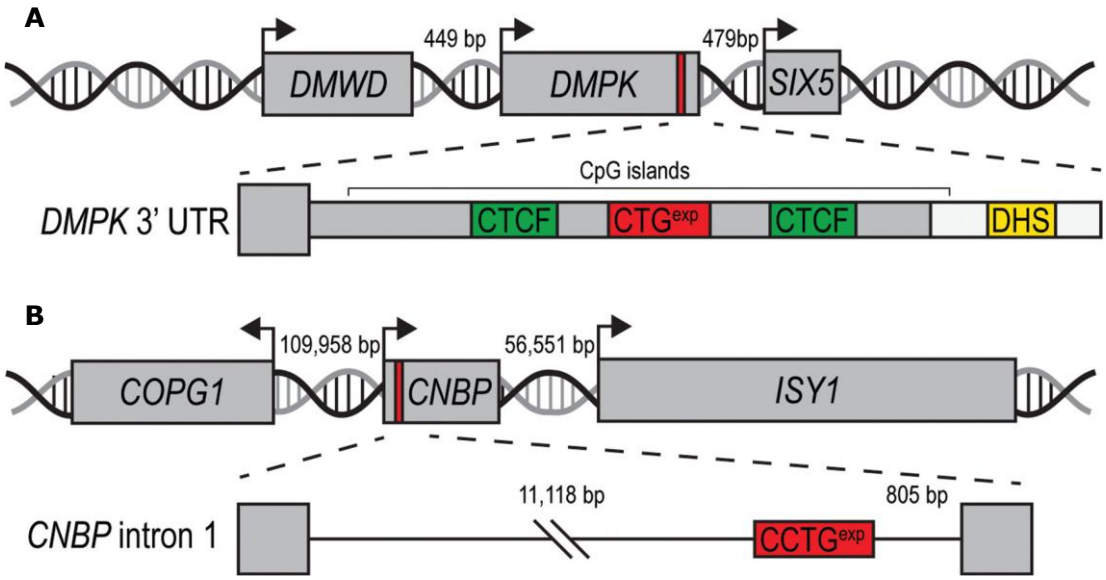
1.3 Genetic basis of myotonic dystrophy

1.3.1 The genetic basis of DM type 1

In 1992, the genetic basis of DM type 1 (DM1) was uncovered, revealing an unstable CTG triplet-repeat expansion in the 3' UTR of *DMPK* located on chromosome 19q13.3. This discovery dispelled the prior belief that DM1 was simply a result of ascertainment bias (Figure 1-3 A) [4, 39, 40]. Typically, individuals who are unaffected by DM1 have CTG repeats ranging from 5 to 37. However, once the $(CTG)^{exp}$ exceeds 50 repeats, disease symptoms arise [6]. DM1 was one of the first three diseases to be identified as being associated with unstable nucleotide expansions, the others being fragile X syndrome (FXS) and X-linked spinal-bulbar muscular atrophy (Kennedy's disease). Repeat tracts in these diseases become unstable and are susceptible to both intergenerational and somatic expansions beyond a disease-specific threshold [6]. This discovery of microsatellite expansions has provided a mechanistic understanding of genetic anticipation, which is associated with several microsatellite expansion disorders, including FXS, Huntington disease (HD), and various types of spinocerebellar ataxia (SCA) [41, 42]. The length of $(CTG)^{exp}$ repeats in DM1 generally correlates with the earlier age-of-onset and increased severity of

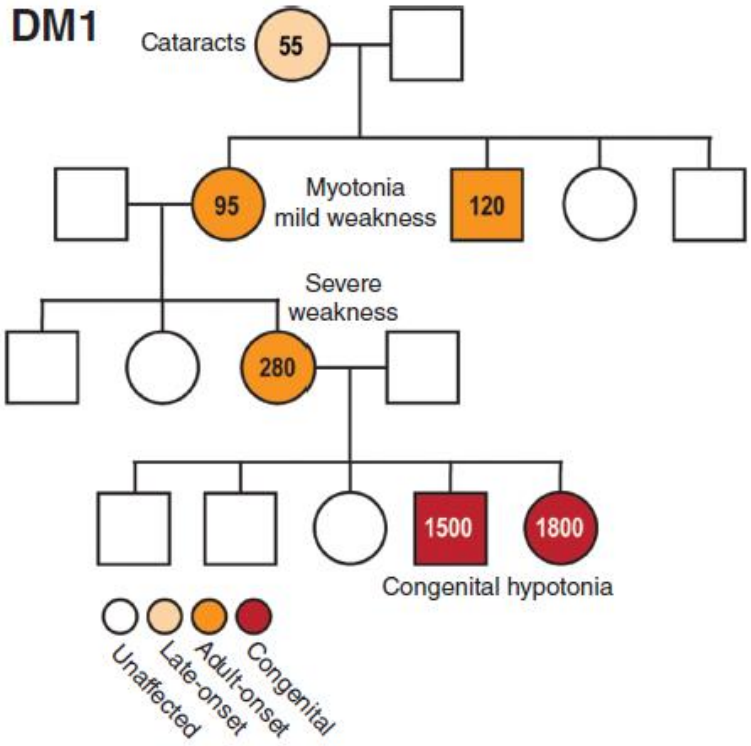
symptoms, with longer repeats associated with more severe disease (Figure 1-4) [43]. The polymorphism of CTG expansion in length is associated with the clinically defined DM manifestations, ranging from mild/asymptomatic (~50-<150), classic (~50-<1000) to congenital (>1000) (Figure 1-5) [44]. Moreover, expanded DM1 alleles are somatic instable, which means the size of CTG repeats expands over time and the rates of expansion are variable between affected individuals and between different tissues in the same individual [45, 46]. However, factors that regulate individual- or tissue-specific dynamics of somatic instability remain unknown [43].

Figure 1-3 DM1- and DM2-associated gene loci.



(A) The *DMPK* $(CTG)^{exp}$ (red box) is situated in the 3' untranslated region (UTR) of the gene, in close proximity to two neighbouring genes, *DMWD* and *SIX5* (transcription start sites are indicated by arrows). The $(CTG)^{exp}$ is flanked by CTCF binding sites (green boxes) and a downstream DNase hypersensitivity site (DHS, yellow box), which may play a role in regulating the epigenetic features of this genomic region. (B) The $(CCTG)^{exp}$ associated with DM2 is located within the first intron of *CNBP*. Figure adapted from [6].

Figure 1-4 DM1 pedigree highlights genetic anticipation.



Hypothetical pedigree of a DM1 family with males (boxes) and females (circles) and mutant allele CTG repeat lengths indicated. Figure adapted from [6].

1.3.2 The genetic basis of DM type 2

After the discovery of the *DMPK*-linked $(CTG)^{exp}$ mutation, patients who resembled DM1 patients but lacked a DM1 mutation were reported to have a preference for proximal rather than distal muscle involvement [47]. This condition was initially referred to as proximal myotonic myopathy (PROMM). In the late 1990s, the genetic mutation responsible for PROMM was mapped to the chromosomal region 3q21, and it was later discovered to be an unstable CCTG quadruplet-repeat expansion in the first intron of *CNBP* (previously known as zinc finger 9, *ZNF9*) (Figure 1-3 B) [48, 49]. Subsequently, the *DMPK*-linked disease was renamed DM type 1 (DM1) and PROMM) [50] or proximal myotonic dystrophy (PDM) [51] was designated as

DM type 2 (DM2), based on their phenotypic similarity but distinct aetiology and presentation. Individuals without DM2 typically have less than 30 CCTG repeats, whereas disease symptoms have been observed in patients with as few as 55 CCTG repeats (231). DM2 is generally a later onset and less severe disease compared to DM1, and a congenital form of DM2 has not been reported (Figure 1-5). Furthermore, the prevalence of DM1 (1 in approximately 8000) is higher and more widespread than DM2 (approximately 3% of DM cases worldwide), except in some regions of Northern Europe where the epidemiology is more similar [52]. However, DM2 may be underdiagnosed since it is often a late-onset disease and may be confused with aging-associated diseases [53].

Figure 1-5 Clinical manifestations and disease stages in DM1, DM2, and CDM.

Disease	Subtype	Repeat #	Skeletal muscle	Other tissues
DM1	CDM	> 1000	Hypotonia, muscle immaturity, respiratory distress, polyhydraminos, dysphagia, talipes	Heart defects, intellectual disabilitiy
DMPK CTG^{exp}	Juvenile-onset	50 to 1000	Myotonia (grip, percussion, electrical), weakness, distal muscle wasting, insulin resistance	Heart defects, hypersomnia, cataracts, endocrine dysfunction
	Adult-onset	50 to 1000		
	Late-onset	50 to 1000		
	Premutation	38 to 49	Normal	Normal
	Normal	5 to 37	Normal	Normal
DM2	Affected	> 75 to 11,000	Myotonia (grip, percussion), proximal muscle wasting, muscle hypertrophy, myalgia	Heart defects, cataracts, endocrine dysfunction
CNBP CCTG^{exp}	Normal	10 to 35	Normal	Normal

In DM1, different subtypes are identified based on clinical criteria and their associated symptoms. Although juvenile-, adult-, and late-onset DM1 subtypes are characterized by 50 to 1000 CTG repeats, a larger CTG expansion size is typically observed in individuals with earlier age-of-onset and more severe disease manifestations. This relationship between CTG expansion size and disease severity is not as strong in DM2. Figure adapted from [6].

1.4 The role of MBNL, CELF and other alternative splicing factors involved in DM pathogenesis

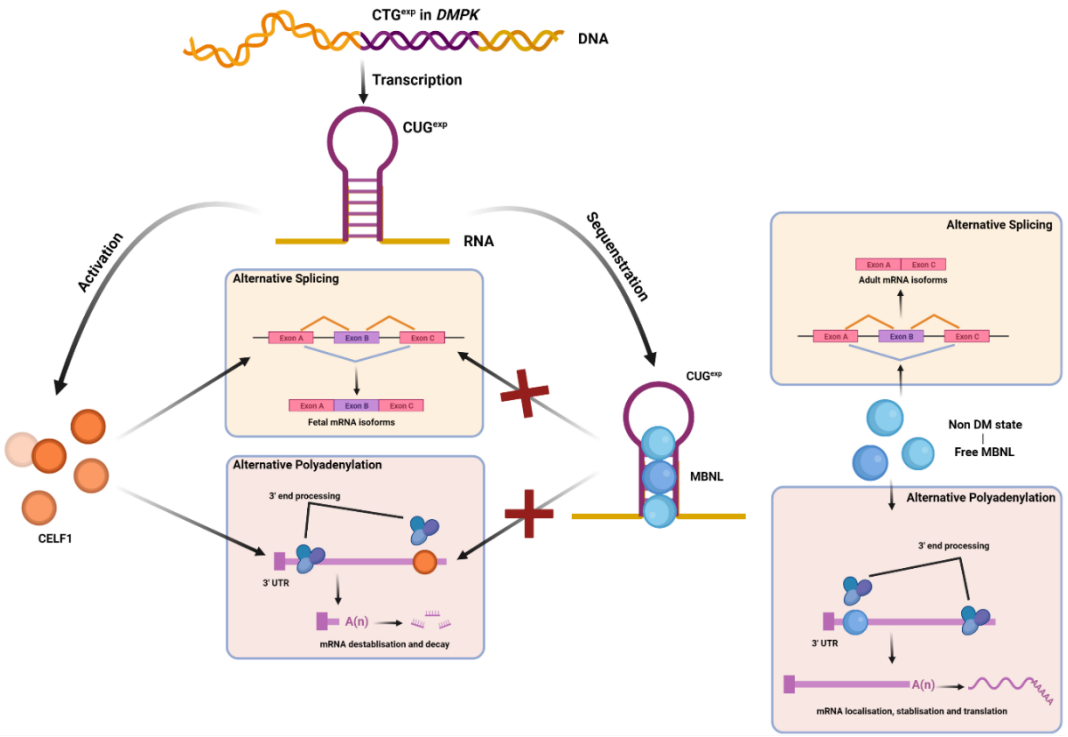
One of the best studied aspects of the molecular pathway underlying DM is that the nuclear retained C(C)UG expansion transcript causes the sequestration of MBNL proteins, formation of RNA foci and activation of CELF1 to produce RNA mis-splicing in several different transcripts [54, 55] (Figure 1-6).

Alternative splicing (AS) is a fundamental process in eukaryotic gene expression that allows the generation of multiple protein isoforms from a single gene [56, 57]. The process involves the differential inclusion or exclusion of exons during pre-mRNA splicing, resulting in the production of different mRNA transcripts and ultimately different protein products [57]. Alternative splicing is regulated by a complex interplay of RNA-binding proteins, splicing factors, and other regulatory molecules [58, 59].

MBNL proteins and CELF1 play a critical role in regulating alternative splicing in normal cells by binding to specific RNA sequences in pre-mRNA transcripts and promoting exon inclusion or exclusion [60-62]. However, in DM1, sequestration of the MBNL proteins in RNA foci and activation of CELF1 (discussed in more detail below), results in depletion of MBNLs and elevated steady-state level of CELF1 with a shift in their target mRNAs from adult isoforms to fetal isoforms and DM disease manifestations [55]. For example, aberrant splicing of the chloride channel (*CLCN1*) gene can result in the loss of muscle excitability and myotonia [62], a hallmark symptom of DM. Similarly, aberrant splicing of the insulin receptor (*INSR*) gene can lead to insulin resistance and glucose intolerance [63], common features of DM. Other downstream effects of aberrant splicing in DM can include altered ion channel function, abnormal mitochondrial function, and dysregulation of various signalling pathways [64-67].

The loss-of-function of MBNL proteins has also been linked to several other CTG expansion diseases beyond myotonic dystrophy (DM), including Fuchs endothelial corneal dystrophy (FECD) [68] and spinocerebellar ataxia type 8 (SCA8) [69]. Finally, MBNL proteins may be indirectly sequestered by other RNA binding proteins (RBPs) or by other types of repeats, such as $(CGG)^{exp}$, which is associated with Fragile X-associated tremor/ataxia syndrome (FXTAS), and $(CAG)^{exp}$, which is linked to polyglutamine diseases [70, 71].

Figure 1-6 Mis-regulated alternative splicing and mis-regulated alternative polyadenylation model in human DM1.

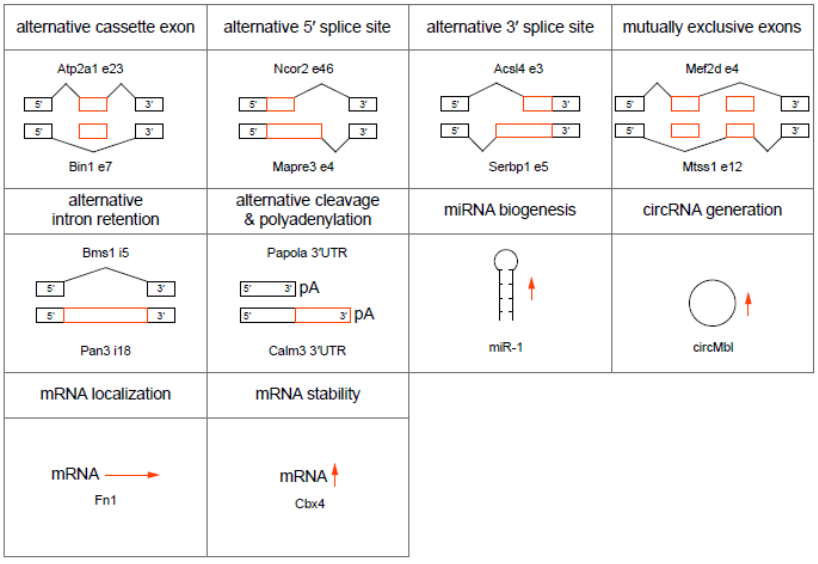


CUG expansion *DMPK* mRNA sequesters Muscleblind-like (MBNL) and activates the expression of CUGBP Elav-Like Family Member 1 (CELF1) to a steady-state level through protein kinase C (PKC)-mediated hyperphosphorylation. These changes cause spliceopathy and fetal splicing shifts in adult tissues. In Addition, CELF1 binding to 3’ UTRs leads to mRNA repression and decay while MBNL binding to 3’ UTRs promotes mRNA stabilisation, localisation to membrane compartments and translation.

1.4.1 MBNL domains important for RNA binding and splicing regulation

According to current pathomechanistic models, the onset and progression of DM1 and DM2 diseases are largely influenced by the MBNL family of RNA processing factors. These factors, which are trans-acting and approximately 37-43 kDa in size, play a significant role in alternative pre-mRNA splicing, pre-mRNA 3' end cleavage/ polyadenylation, mRNA localisation, mRNA stability, micro RNA biogenesis, and circular RNA generation during both embryonic and postnatal development (Figure 1-7) [72-77]. While all three mammalian Mbnl/MBNL paralogs are important, mouse Mbnl1 and Mbnl2 are primarily involved in postnatal development, with their RNA targets shifting to adult expression patterns. However, *Mbnl1* and *2* double knock-out mice are embryonic lethal, indicating essential roles during fetal development [78]. Mbnl3, on the other hand, is mainly expressed during embryonic development and adult tissue regeneration [79].

Figure 1-7 MBNL functions in RNA processing.



MBNL controls diverse alternative splicing patterns such as cassette exon, 5’ splice site, 3’ splice site, mutually exclusive exons, intron retention [76], and alternative 3’ end formation through alternative cleavage and polyadenylation (pA) [77]. Additionally, MBNL plays a regulatory role in the biogenesis of microRNA (miRNA) [80], the formation of circular RNA (circRNA) [81], mRNA localisation (horizontal arrow) [76], and mRNA stability enhancement (vertical arrow) [74]. These examples highlight MBNL-mediated processes, with specific targeted RNAs being identified. Figure adapted from [82].

MBNL1 is the most extensively studied member of the MBNL protein family, primarily due to its predominant expression in muscle tissue. The human MBNL1 gene consists of 10 exons, of which 8 correspond to the coding sequence (labelled 1-8 in this study). Six of these exons (3, 5, 6, 7, 8) undergo alternative splicing (Figure 1-8 A), leading to the expression of at least seven different mRNA variants of MBNL1 [83, 84]. Such extensive alternative splicing is also observed for other MBNL proteins, expanding the functional repertoire of the MBNL protein family. MBNL1 variants containing exons 5 and 7 are mainly expressed during early differentiation stages and in adult DM tissues [85]. Notably, these exons encode nuclear localisation signals that enhance the sequestration of MBNLs in the nuclei of DM cells, thereby contributing to the severity of the DM phenotype.

MBNL1 has two classes of nuclear localisation signal (NLS) (Figure 1-8 C). The first is a bipartite NLS that consists of two highly conserved motifs, KRPALE, which are embedded in exons 5 and 6 [86]. The second is a novel conformational NLS (conNLS) that involves the N-terminus region and either one of the conserved motifs in exon 5 or exon 6 [86]. Similarly, MBNL2 has an alternative exon corresponding to MBNL1 exon 5, which is associated with nuclear localisation and stronger splicing regulation [86]. Exon 7 on the other hand, is involved in determining MBNL1 dimerization. Its

presence induces the formation of multimeric ring-like structures upon binding to toxic RNA hairpins and induces MBNL1 self-association in yeast-two-hybrid assays [84, 87].

MBNL proteins bind to toxic RNAs containing CUG and CCUG repeats through specific recognition of the Watson-Crick face of the GC dinucleotide [88], binding to the YGCY motifs (where Y indicates pyrimidines) with their four CCCH Zinc fingers (ZnFs). The four ZnFs form two tandem ZnF pairs (ZnF1–2 and ZnF3–4) which differ in their ability to bind target RNAs and regulate splicing events. ZnF3–4 shows lower affinity binding to target RNA than ZnF1–2. Hale, M.A., et al. (2018) designed two synthetic MBNL proteins with duplicate ZnF1–2 or ZnF3–4 domains which brought the splicing activity up to 5-fold and down to 25%, respectively, compared to a natural MBNL protein [89]. Their findings indicate that ZnF1–2 drives alternative splicing activity through specific recognition of YGCY motifs, whilst ZnF3–4 serves as an aspecific RNA binding domain which assists MBNL to recognise a wide range of target mRNAs by binding to a broader region of motifs [89]. These results are consistent with previous studies showing that truncated MBNL proteins containing only one of the two ZnF pairs bind to YGCY motifs impotently, suggesting that the tandem ZnFs work collaboratively [90]. Their results also suggested that ZnF3–4 domain may confer stability to MBNL as the expression of MBNL with two ZnF1-2 is reduced by 0.5-fold while the expression of MBNL with two ZnF3-4 is increased by 2.5-fold, compared with wild-type MBNL [89].

However, other regions of the proteins may also interact with RNA and play an important role in regulating gene expression and RNA splicing [91]. Regulatory regions of MBNL for alternative splicing are located within an 80-amino-acid segment, downstream of the N-terminal zinc-finger pair [91]. These domains possess features in

common with the splicing regulatory regions within CELF proteins [91].

Studies by Kino, Y., et al. (2015) and Lee, K.Y., et al. (2013) suggested cross-regulation and splicing compensation among MBNL proteins [78, 86]. Overexpression of any MBNL paralogs repressed the inclusion of exon 5 (54 nt) which is essential for nuclear localisation and splicing activity of MBNL. A reduction in one of the MBNL paralogs would be compensated by the increased nuclear isoforms of the other MBNL protein. Lee, K.Y., et al. (2013) also found, *in vivo* experiments, that MBNL2 can functionally compensate for splicing activity due to the loss of MBNL1 by increased binding to MBNL1-related splicing targets resulting in partial preservation of the adult splicing pattern [78].

Although nuclear localisation of MBNL is crucial for its alternative splicing activity in the nucleus, Pei- Wang, P.Y., et al. (2018) showed that it is the MBNL1 cytoplasmic isoform, not the nuclear isoform, that boosts neurite morphogenesis and reverses the morphological defects caused by CUG triplet repeat RNA [92]. Expanded CUG RNA induces de-ubiquitination of cytoplasmic MBNL1 which leads to nuclear translocation and defects in neurite outgrowth and differentiation. Inhibition of MBNL1 de-ubiquitination with o-PA (1,10-phenanthroline), an inhibitor protecting K63-linked polyubiquitin chains from decay, reduces RNA foci formation and remarkably preserves the morphological features of neurons expressing CUG repeat mRNA [92]. Moreover, MBNL1 exon 7 (36 nt) is fundamental for MBNL dimerization [61, 93]. MBNL1 +ex7 proteins are required for the abundance and splicing of specific mRNAs involved in cell migration and division and DNA repair [93].

1.4.2 Altered MBNL and CELF1 functions in DM

CELF1 and MBNL proteins are antagonistic regulators of alternative splicing. In DM1, MBNL protein activity is repressed due to its sequestration by toxic RNAs, while CELF protein is stabilised and its steady-state level is increased by protein kinase C (PKC)-mediated CELF1 hyperphosphorylation (Figure 1-6) [6, 94].

CELF and MBNL proteins antagonistically regulate alternative splicing of over 200 exons [95]. For instance, MBNL1 is responsible for the regulation of *ABLIM1* (actin-binding LIM protein 1) exon 11 and *CLCN1* (muscle chloride channel) exon 7a splicing which are dysregulated due to sequestration of MBNL1 in the expansion RNAs in DM1. The CELF family also contributes to the aberrant *ABLIM1* and *CLCN1* splicing by antagonizing the effect of MBNL1 [96]. In both cases, aberrant splicing is associated with muscle symptoms of DM1. Many other alternative splicing events have been well validated to be MBNL-dependent, including those affecting *MAPT, NUMA1, MBNL2, NCOR2 and LDB3* [97]. The binding position of MBNL to its target RNAs determines the consequence of splicing events: exon inclusion or exon exclusion. MBNL binding within alternative exons and upstream intronic regions usually promotes exon skipping, while binding to downstream intronic regions promotes exon inclusion. This splicing regulation mechanism resembles that of other factors, such as neurooncological ventral antigen (NOVA), forkhead box (FOX) or heterogeneous nuclear ribonucleoprotein L (hnRNP L) [98].

In addition to a role in alternative splicing regulation, it has been shown that MBNLs and CELF1 play a role in poly A site selection (Figure 1-6). Studies using HITS-CLIP and minigene reporter assays revealed that MBNLs binding to the 3’ untranslated regions (3’ UTRs) of target RNAs has a direct impact on polyadenylation site switching.

MBNLs binding sites are frequently concentrated proximal poly(A) signals (pAs). Depletion of MBNLs resulted in silencing of distal pAs while proximal pAs were boosted [99]. Thus, loss of MBNL mis-regulates alternative polyadenylation affecting multiple pre-mRNA processing procedures in DM [99]. The up-regulation of many alternatively spliced last exons in DM skeletal muscle-related transcripts were closely correlated with estimated MBNL function and consequently muscle strength [97]. MBNL protein sequestration and dysfunction causes a reversion to fetal polyadenylation patterns in adult muscle which may induce protein synthesis inhibition and promote protein catabolism leading to muscle atrophy [55]. Likewise, CELF1 proteins bind to the 3’ UTRs of mRNAs and increase the density of upstream cleavage and polyadenylation sites by recruiting cytoplasmic deadenylases [95]. Down-regulation of CELF1-bound mRNAs is dose-dependent on CELF1 binding to 3’ UTRs [95]. CELF1 binding to alternative 3’ UTRs and altering 3’ UTR isoform abundance is associated with isoform-specific repression. The possible consequence is the altered regulation of cleavage and polyadenylation to produce isoforms with different pAs [95].

CELF1 and MBNL1 have similar binding sites in 3’ UTRs and tend to cluster together, where they compete to bind their co-target mRNAs and lead them to different fates. The extent of CELF1 and MBNL1 binding to 3' UTRs correlates with the degree of mRNA suppression or stabilisation, respectively. Greater MBNL binding prevents mRNAs from CELF-mediated inhibition. However, this antagonism effect by MBNL1 and CELF1 is also dependent on binding site proximity. Wang *et al* (2015) showed that a distal MBNL1 binding site in 3’ UTRs was sufficient to prevent mRNA repression in the presence of a CELF1 site, but this protective response was not observed in the presence of a proximal MBNL1 site [95]. mRNAs with more proximal CELF1/MBNL1 pairs are more strongly down-regulated than those

with distal CELF1/MBNL1 pairs [95]. But no evidence has shown direct antagonism of binding between CELF1 and MBNL proteins.

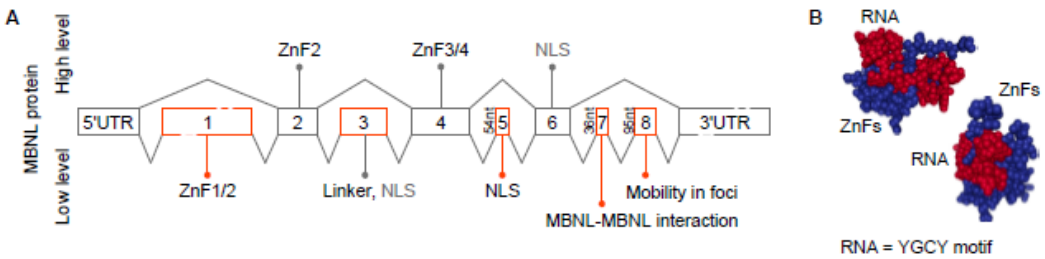
CELF1 is also an important skeletal muscle translational regulator. Its translational function is important for normal myogenesis and is activated by phosphorylation at Ser-302 by cyclin D3–dependent kinase 4 (CDK4) [100]. Interestingly, it has been reported that CELF1 is under-phosphorylated at Ser-302 in DM1 muscle cells compared to that in normal muscle cells [101]. This reduction of phosphorylation of CELF1 is due to the suppression of CDK4 signalling which is caused by the increased stability and activity of glycogen synthase kinase 3β (GSK3β) [101]. However, the mechanism of the elevated level of GSK3β in DM1 remains to be investigated. Inhibition of CELF1 translational activity is coupled with myogenic defects and muscle weakness which are the hallmark characteristics of DM1 [101, 102].

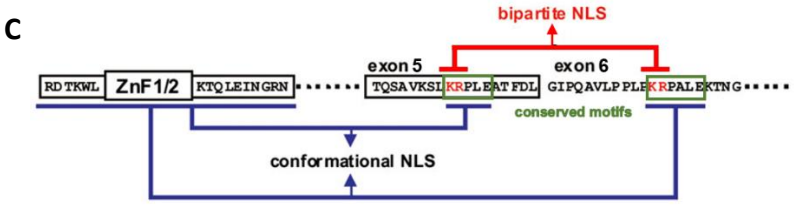
1.4.3 $[C(C)UG]^{exp}$ RNA-protein Interactions in RNA Foci Formation

RNA fluorescence *in situ* hybridisation (RNA-FISH) using a fluorescent probe complementary to expanded CUG repeats in DM1 nuclei revealed unusual focal concentrations of these repeats, just three years after the discovery of the DM1 mutation [103]. This was reported by Singer and colleagues. Subsequently, RNA foci were also discovered in other diseases caused by short tandem repeat (STR) or microsatellite expansions [82, 104]. The discovery of RNA foci has changed our understanding of these structures from insoluble RNA clusters to dynamic ribonucleoprotein (RNP) complexes. RNA foci are a type of non-membrane bound RNP complex, similar to other complexes found in the nucleus, including nucleoli, paraspeckles, nuclear speckles, and Cajal bodies, as well as in the cytoplasm, such as P bodies, stress granules, and neuronal and germ granules [105].

The formation of RNA foci in DM1 and DM2 occurs due to co-transcriptional recruitment of MBNL proteins to [C(C)UG]^{exp} repeats [106]. The exact role of MBNL proteins in RNA foci assembly is not completely understood, but studies have shown that the size of RNA foci increases with higher MBNL concentrations, while depletion of these proteins results in a decrease in foci area/size. This suggests that MBNL plays a role in promoting foci assembly and/or stability [106-109]. Furthermore, MBNL proteins have two tandem ZnF pairs (Figure 1-8 A), which could enable a single protein to interact with two RNA molecules. ZnF 1/3 and ZnF 2/4 have similar motifs or CX7CX6CX3H and CX7CX4CX3H, X = amino acid, respectively and a single ZnF pair is sufficient to bind to target RNA (Figure 1-8 B)[110]. Protein-protein interactions might also enhance RNA foci formation, as homotypic interactions are mediated by the MBNL1 C-terminal region, which includes alternatively spliced exon 7 [84, 87]. Other proteins, including hnRNP H, H2, H3, F, A2/B1, K, L, DDX5, DDX17, and DHX9, have been proposed to be sequestered within RNA foci, but their precise role in foci formation is uncertain [111]. These proteins could exacerbate (e.g., DDX5) or ameliorate (e.g., DDX6) the toxicity of these complexes by either promoting or blocking foci formation [112, 113]. Despite the detection of several other proteins within foci, including hnRNPs, their impact on the DM pathomechanism is less clear [111, 114].

Figure 1-8 MBNL gene and ZnF structures.



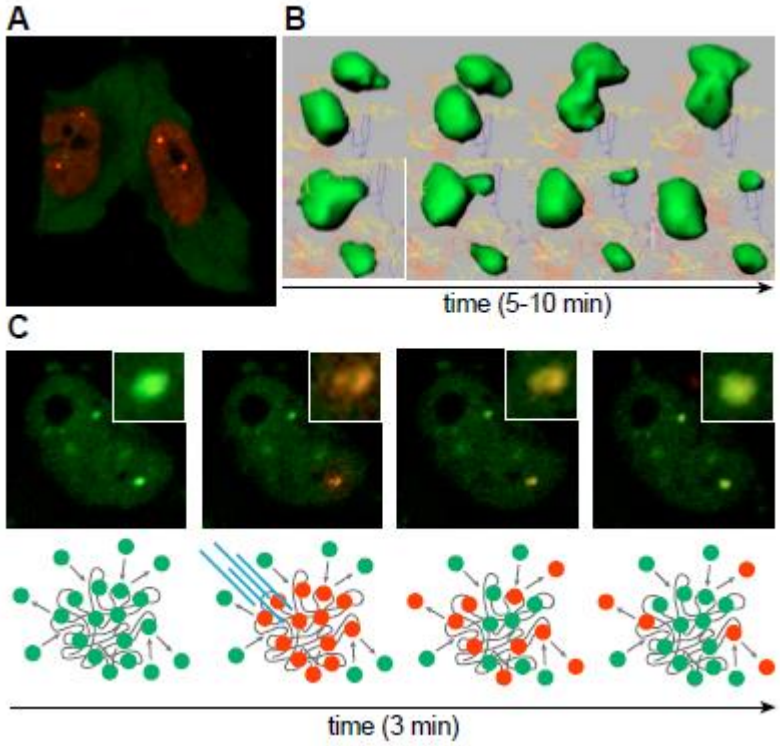


(A) Exonic structure and function of MBNL paralogs [86, 107]. The enumeration of exons is based on previous studies [83], and the zinc finger domain pairs are labelled ZnF1/2 and ZnF3/4, corresponding to pairs 1/2 and 3/4, respectively. NLS, nuclear localisation signal. The MBNL protein levels are shown as Low (bottom) and High (top), which trigger the splicing events indicated by black line. (B) A structural model of MBNL1 zinc fingers (ZnF) shown in blue in complex with RNA shown in red, from different views [115]. Figure adapted from [82]. (C) The nuclear localisation mechanism of MBNL1.

Reports have indicated the dynamic nature of interactions between RBPs and expanded RNAs in foci, particularly for MBNL and RBFOX paralogs in various length STR RNAs [107, 108, 114]. Live-cell imaging has captured the dynamics of RNP foci (Figure 1-9 A), demonstrating that foci can merge upon direct interaction or break down into smaller units (Figure 1-9 B). The characteristics of foci, such as their number, shape, and volume, are susceptible to changes over time and in different cellular states [107, 108, 116]. RNA foci can remain stable for minutes or hours, while MBNL proteins, which are densely packed in foci, move dynamically between RNA binding sites within the foci and the pool of unbound MBNL proteins in the surrounding nucleoplasm (Figure 1-9 C). Interestingly, a small amount of expanded CUG RNA can quickly become saturated with MBNL proteins, which can then exchange dynamically with the unbound MBNL in the nucleoplasmic pool, while higher CUG loads significantly deplete this pool [107]. Thus, the assembly of RNA foci *in vivo* is likely modulated not only by the length of the expansion repeat but also by the spatiotemporal

pattern of *DMPK* and *CNBP* expression. For example, *DMPK* is expressed at significantly higher levels during fetal muscle development than in mature tissues, and its expression varies considerably among different tissues [117].

Figure 1-9 MBNL sequestration on the CUG expansion transcript.



(A) Live-cell imaging performed on cells transfected with CUG repeats, mCherry-MBNL1, and GFP-MBNL3. MBNL1 was found mainly in the nucleus, while MBNL3 was predominantly in the cytoplasm. (B) Time-lapse images of $(CUG)^{exp}$-GFP-MBNL1 complexes (green) captured and analysed using Imaris software. (C) Time-lapse sequence images of a cell with transient transfection of CUG repeats and Dendra2-MBNL1 fusion protein, which accumulated in two distinct nuclear foci. Laser stimulation induced photoconversion of Dendra2-MBNL, changing its emission from green (507nm) to red (573nm) in a single focus, revealing the dynamic exchange of MBNL between the focus and surrounding nucleoplasm over the course of several minutes. The focus core appeared to contain less-exchangeable MBNL proteins, indicated by its yellowish centre. The experimental setup is schematically represented below each image. Figure adapted from [107].

Although DM1 and DM2 are both myotonic dystrophies with similar clinical features such as myotonia, progressive myopathy, and multiorgan involvement, they exhibit distinct differences [44]. DM2 typically manifests later in life and lacks a congenital form, and the *CNBP* gene is expressed at considerably higher levels than the *DMPK* gene [48, 118]. In DM1, MBNL proteins are primarily sequestered, while in DM2, both MBNL and RBFOX proteins can be sequestered. The sequestration of these RNA-binding proteins disrupts their normal functions and leads to mis-regulation of alternative splicing events [114].

Despite the fact that the affinity of MBNL proteins for $(CCUG)^{exp}$ is higher than for $(CUG)^{exp}$ [87, 119], mis-splicing is usually less severe in DM2 than in DM1. This difference in severity implies that MBNL proteins may have a more profound impact on alternative splicing than RBFOX proteins in the context of these diseases. RBFOX proteins bind to $(CCUG)^{exp}$ but not $(CUG)^{exp}$ RNAs, competing with MBNL for RNA binding sites. This selective reduction of MBNL sequestration in DM2 promotes more adult-like splicing patterns [114].

While both protein families are involved in alternative splicing regulation, the specific contribution of each protein to the disease phenotype may vary. MBNL proteins have been extensively studied in the context of DM1 and are known to play a crucial role in regulating alternative splicing [61, 77, 86]. The sequestration of MBNL proteins by expanded repeat RNA leads to widespread changes in alternative splicing patterns, contributing to the pathogenesis of DM1 [62, 73]. The mis-regulation of specific genes involved in muscle function and other processes contributes to the severity of the disease. On the other hand, the exact role of RBFOX proteins in DM2 and their contribution to the disease phenotype is still being investigated. While it is known that RBFOX proteins can

be sequestered by the expanded CCTG repeats, the extent of their sequestration and the downstream effects on alternative splicing are not yet fully understood [114, 120]. It is possible that the competition between RBFOX and MBNL proteins for binding to the expanded repeats in DM2 may partially mitigate the impact on alternative splicing compared to DM1.

Overall, the interplay between MBNL and RBFOX proteins and their differential sequestration in DM1 and DM2 likely contributes to the varying disease severities. However, further research is needed to elucidate the specific roles of these proteins and the underlying molecular mechanisms that determine the phenotypic differences between DM1 and DM2.

DM2 also exhibits distinct cell-specific molecular signatures. For instance, mis-splicing of MBNL1 RNA targets is absent in DM2 iPSC-derived cardiomyocytes, and there is retention of *CNBP* intron 1 with nuclear export of the incompletely processed mRNA. This retention results in enhanced expression of aberrant tetrapeptide repeats through RAN translation, which is a major pathological factor in DM2 across multiple tissues [121, 122].

1.5 Therapies for Myotonic Dystrophy

There is currently no curative treatment for either form of myotonic dystrophy, but with improved understanding of the disease mechanism, many disease intermediates have become potential therapeutic targets, such as the expanded CTG repeats, the toxic mRNAs, RNA-binding proteins, splicing regulators and mis-spliced transcripts. There are many levels at which the therapy can be directed, and different strategies have been developed and tested.

1.5.1 Targeting the expanded CTG repeats in the *DMPK* gene

The repeat expansion is highly unstable in the germline and in

somatic cells. To date, targeting the expanded CTG repeats has been difficult because the expression of the repeats is regulated by cis- and trans-acting gene modifiers. Nonetheless, elimination of the repeat expansions or reductions in repeat length should be curative. A few studies have focused on targeting repeat expansions. CRISPR/Cas9-mediated therapeutic gene editing has been developed as a powerful tool to correct genetic defects *in vitro* and *in vivo*. The CRISPR system requires a short RNA of around 20-bp (guide RNA or gRNA) and an endonuclease, usually Cas9, to identify the target site in the genome and create a single or double-strand break, which can then be repaired by NHEJ, introducing random mutations, or by the introduction of precise mutations using a repair template. Recently this technology has been employed to target *DMPK* 3' UTR sequence upstream and downstream the CTG repeat region and led to efficient deletions of CTG repeats and subsequently foci disappearance and correction of aberrant alternative splicing in myogenic cells derived from DM1 patients [123-126]. In one of the most recent studies, the expression of Cas9 nuclease was driven by a muscle-specific promoter and the expression guide RNAs was time-controlled by tetracycline-inducible system, so that the protential occurance of off-target events was minimised [125]. Elimination of the repeats did not affect the transcription of *DMPK* or its neighbour gene *SIX5* [123, 124].

Another approach to target the DNA repeat expansion is to inhibit its transcription and thus the production of the toxic RNA. Pentamidine, a small molecule composed of two phenymamidine groups joined by a five-carbon methylene linker has been identified to inhibit the transcription of $(CUG)^{exp}$ *DMPK* gene by binding to the CTG·CAG repeat DNA directly [127, 128]. Actinomycin D is another small molecule with high affinity for GC-rich double stranded DNA, which has been identified to inhibit the transcription of CTG repeats,

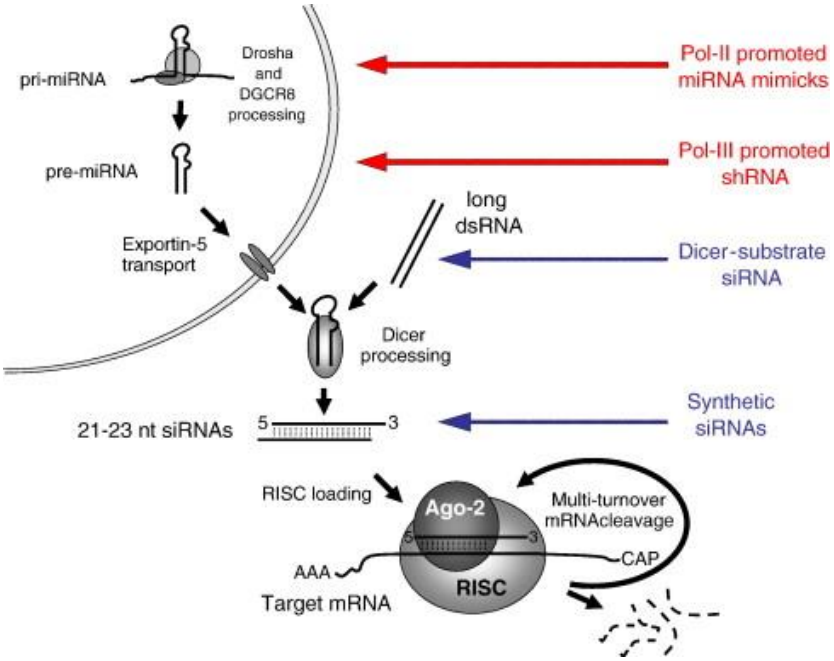
producing a reduction in the number of foci [129]. Inhibition of the transcription of CTG repeats can also be achieved by using the CRISPR system with a deactivated version of the Cas9 (dCas9). The dCas9 binds to the CTG repeats, blocks the action of RNA polymerase II and impedes the transcription of CTG repeats in a length-, PAM- and strand-dependent manner [130]. Transcriptional Inhibition of the CTG repeats by small molecules or CRISPR/dCas9 system is also shown to be able to rescue several missplicing events and reduce the rate of RAN translation [128-130].

1.5.2 Targeting the toxic RNA

Targeting toxic RNA in DM1 therapy can be accomplished by utilising RNA interference (RNAi), which makes use of the endogenous cellular machinery. RNAi can be triggered through two different approaches (Figure 1-10), each with its own unique mechanism. The first approach is RNA-based, where short double-stranded RNA molecules known as effector siRNAs are delivered to target cells. Once inside the cell, each siRNA is unwound into two single-stranded RNAs, the sense (passenger) strand and the antisense (guide) strand [131]. The antisense strand is then incorporated into the RNA-induced silencing complex (RISC), which subsequently binds to and degrades the target mRNA. At the same time, the sense strand is cleaved by the protein Argonaute 2 (Ago2) and degraded [132]. The second approach is DNA-based and involves the production of siRNA effectors through intracellular processing of longer RNA hairpin transcripts [131] (Figure 1-10). This process is primarily based on the nuclear synthesis of short hairpin RNAs (shRNAs), which are then transported to the cytoplasm via the miRNA machinery. Once in the cytoplasm, shRNAs are cleaved into siRNAs by the ribonuclease Dicer.

Interestingly, the expanded C(C)UG repeats are frequently formed into a hairpin structure which is a endogenous substrate for Dicer [133]. However, MBNLs, rbFOXs and other proteins that are sequestered by the C(C)UG repeats may block Dicer binding to the hairpin structures. Therefore, researchers have engineered siRNAs or shRNAs to target the $(CUG)^{exp}$ *DMPK* transcript directly in DM1 cells and in HSA^{LR} mice, which has been successful in reducing the abundance of $(CUG)^{exp}$ RNAs as well as nuclear foci formation, and rescuing several aberrant alternative splicing events [134, 135]. HSA^{LR} transgenic mice carry a fragment of human skeletal actin (*HSA*) gene, containing an expanded 250-CTG repeat sequence inserted in its final exon [136, 137].

Figure 1-10 RNAi-based therapeutic strategies.



This figure depicts two approaches for utilising the RNA interference (RNAi) pathway to induce gene silencing. The left-hand side shows a schematic diagram of the endogenous RNAi pathway, which involves the formation of small interfering RNAs (siRNAs) that bind to and degrade target mRNA in a homology-dependent manner in the cytoplasm via the RNA-induced silencing complex (RISC). The right-hand side shows two synthetic siRNA molecules, one DNA-

based (marked in red) and the other RNA-based (marked in blue), which are introduced at different entry points to activate RISC for gene silencing. These siRNAs are designed to mimic the endogenous siRNA molecules and trigger the RISC pathway for targeted gene silencing. Figure adapted from [131].

Direct use of siRNAs is simple and result in potent gene silencing. However the effect is transient, which means in a clinical setting, repeated treatments would have to be administered. DNA-based shRNA strategy on the other hand is a gene-editing therapy and has the potential of being stably introduced, allowing in principle a single treatment of shRNA gene delivery.

Therapies using oligos, such as antisense oligonucleotides (ASOs) or oligomers containing all-LNA units, also show particular promise in treating DM1 by targeting toxic CUG expansions. An antisense 25-nucleotide morpholino oligonucleotide termed CAG25, and a 2'-O-methyl-phosphorothioate-modified $(CAG)_7$ ASO have been designed to compete against MBNLs to bind to the expanded CUG repeats, as a consequence it releases MBNL from sequestration as well as triggers the degradation of the toxic RNA [138, 139]. However, it is still unknown that what degradation pathway the transcript undergoes when bound to the ASOs as they are predicted to be incompatible with RNase-H degradation [138, 139]. Different to ASOs, LNA oligomers can stably bind to the expanded CUG repeats and prevent the MBNL sequestration without triggering the degradation of toxic RNAs [140].

Other therapeutic strategies include the direct use of the RNase-H degradation pathway. RNase H is a non-sequence-specific enzyme that recognises DNA-RNA heteroduplexes and cleaves the RNA strand. Lee et al. [141] generated a neotype ASO, termed gapmer to target the toxic RNAs. The stability and efficiency of the gapmers are enhanced by using locked nucleic acids (LNA) or 2'-O-

Methoxyethyl (MOE) nucleic acids and a centre "gap" region with 7–10 nucleotides containing phosphorothioate (PS) DNA can trigger RNase H-induced RNA degradation [141]. Administration of the gapmers to cultured DM1 cells and DM1 mouse models preferentially targeted long CUG repeats as the $(CUG)^{exp}$ *DMPK* transcripts are predominantly retained in the nucleus where RNase-H degradation is mainly carried out [141]. Later, S.K. Pandey et al [142] modified the structure of gapmers that the new gapmer termed ISIS 486178 contained 2',4'-constrained ethyl-modified (cEt) residues, which are structural analogs of LNAs with the addition of a single methyl group on the ethyl bridge and it exhibited a significantly increased RNA binding affinity and *in vivo* potency. ISIS 486178 significantly reduced levels of the toxic RNAs in heart and skeletal tissues via targeting the 3' UTR region of mutant *DMPK* mRNA instead and was well tolerated by the DMSXL mice [142]. DMSXL transgenic mice carry a large fragment of human genomic DNA containing the three genes of the DM1 locus (*DMWD*, *DMPK* and *SIX5*) and >1000 CTG repeats [136, 137]. The safety of ISIS 486178 was also verified in wild-type mice and cynomolgus monkeys that both species tolerated it well in spite of the reduction of endogenous *DMPK* mRNA [142]. Thus, the use of RNase-H activated antisense technology could provide a therapeutic advantage in reducing the toxic RNAs.

However, further optimization of the delivery technologies is still required to circumvent the poor efficacy of antisense drugs in patients and the challenge of tissue-specific drug delivery.

The CRISPR system can also be used to target the repeat expansions at the RNA level. A modified version of the Cas9, an RNA-targeting Cas9 (RCas9) effectively induced degradation of the repeat expansion in different microsatellite expansion disorders models, including myotonic dystrophy type 1 and 2. Elimination of the repeats reduced foci in the cells, the levels of polyGln proteins and

rescued alternative splicing of several genes such as *INSR*, *ATP2A1*, *SERCA1*, *MBNL1* and *MBNL2* [143]. Thus, the CRISPR system might be used as a therapy for myotonic dystrophy and can be directed to different levels of targets involved in the pathogenesis of the disease.

Ketley et al. recently discovered that cyclin-dependent kinase 12 (CDK12) could serve as a potent target to develop drug treatments for DM1 [144]. The protein level of CDK12 is elevated DM1 cell lines and patient muscle biopsies [144]. Inhibition of CDK12 through treatment with small compounds reduced RNA foci and had beneficial effects in a mouse model of DM1 [144]. The mechanism-of-action remains unclear, but it is hypothesised that the inhibition of CDK12 might reduce the production of the mutant *DMPK* transcript or promote its degradation.

Additionally, there are two particularly promising drugs currently undergoing clinical trials: AOC 1001 and Dyne 101. AOC 1001 is an investigational drug being developed by Avidity Biosciences for the treatment of DM1. It is a small molecule compound designed to target the toxic DMPK RNA responsible for DM1 [145, 146]. AOC 1001 aims to reduce the levels of the toxic RNA by selectively binding to it and promoting its degradation [145, 146]. The drug is intended to alleviate the symptoms of DM1 by addressing the underlying molecular cause of the disease. Dyne 101 is an investigational antisense oligonucleotide therapy for DM1 [145, 147]. It is being developed by Dyne Therapeutics. Dyne 101 is designed to target the toxic expanded CUG repeats in the DMPK gene [147]. By binding to the expanded CUG repeats, Dyne 101 aims to alter the RNA structure, preventing the formation of toxic RNA foci and restoring normal cellular function.

Both AOC 1001 and Dyne 101 demonstrate exciting potential in providing much-needed therapeutic options for DM1 patients. As

they progress through clinical trials, researchers and clinicians eagerly anticipate their success in improving the quality of life for individuals affected by this complex and challenging genetic disorder. The hope is that these groundbreaking drugs will pave the way for a new era of targeted treatments, bringing us one step closer to effectively managing and ultimately curing DM1.

1.5.3 Targeting protein-RNA binding

The C(C)UG expansion mRNAs sequester MBNL proteins and cause MBNLs loss-of-function in alternative splicing which contributes to the disease pathogenesis. Thus, the protein-RNA complex could be a prospective target to develop a therapy for myotonic dystrophy. Oligomeric compounds and other small molecules have been applied to specifically target the DM1 pathobiology at the DNA or RNA level via selective recognition of the repeating array of 1 × 1 nucleotide U/U internal loops formed by $(CUG)^{exp}$ RNA. It has been shown that a dimeric 2,9-diamino-1,10-phenanthroline derivative (DDAP) designed by Li, *et al*. (2018) [148] and monomeric bioactive ligands designed by Angelbello, *et al*. (2021) [149] enable interaction with $(CUG)^{exp}$ RNA and inhibit the formation of $(CUG)^{exp}$ RNA-MBNL1 complexes. Those molecules are dimers that contain two 5'CUG/3'GUC-binding modules. DDAP binding to $(CUG)^{exp}$ RNA is enchanced through Π–Π stacking interaction with neighboring guanines [148]. Amongst the ligands, the doubly guanidinylated molecules are the most effective inhibitors of the $(CUG)^{exp}$ RNA-MBNL1 complex by efficient interaction with the phosphodiester backbone [149]. Some of such compounds significantly ameliorated DM1-associated molecular defects including toxic CUG repeat RNA aggregation into foci and alternative splicing deficiency [148, 149]. Similarly, Lee, J., *et al*. (2019) designed multivalent ligands which effectively blocked transcription of $d(CTG{\cdot}CAG)^{exp}$ and inhibited the $(CUG)^{exp}$ RNA-MBNL1 interaction [150, 151].

Previous investigations have explored the potential of targeting the C(C)UG repeat sequence directly using antisense oligonucleotides (ASOs). For example, phosphoroamidate morpholino oligomers (PMO) which comprise of a short segment of CA(G)G repeats that bind to expanded C(C)UG RNA directly [152], enabled nuclear redistribution of MBNLs and other RNA binding proteins and induce the degradation of $[C(C)UG]^{exp}$ RNA through the RNase-H mechanism [153, 154]. However, those PMOs also target *DMPK* transcripts from the wild-type *DMPK* allele or other alleles containing C(C)TG repeat expansions which may induce loss-of-function phenotypes, such as myopathy [155].

Chakraborty, *et al*. (2018) identified a drug, daunorubicin hydrochloride which was reported to interact with double-stranded RNA [156]. Their research indicates that daunorubicin directly binds to CUG RNA repeats and alleviates MBNL sequestration. Similarly, erythromycin which is a widely used antibiotic to treat various bacterial infections and two antifungal antibiotics, Lomofungin and dilomofungin, have been shown to dissipate RNA foci and rescue mis-splicing in DM1 cell models by displacing sequestered MBNL [157, 158]. Another potential approach to alleviate MBNL sequestration is by using oligopeptides. Cpy(a/t)(q/w)e D-Hexapeptides compete with MBNL binding to CUG repeat RNAs as well as doubling the transcription level of MBNL1. As a consequence, MBNL-dependent mis-splicing events are rescued. Intriguingly, those peptides share some similarities in structure and sequence with conserved amino acids within ZnFs of MBNL [60].

1.5.4 Altering levels of RNA binding proteins

The aberrant alternative splicing of genes in myotonic dystrophy is caused by the MBNL loss-of-function and the over-activated expression of CELF1. Thus, regulating the expression MBNL and

CELF1 intrinsically could ameliorate DM1 symptoms.

Overexpressing MBNL1 improves the proliferation capability of skeletal muscle satellite cells (SSCs) in DM1 by inhibiting autophagy via the mTOR pathway [159]. Treatment with the anti-autophagic drug chloroquine, was shown to up-regulate MBNL1 and 2 proteins in animal DM models and patient-derived myoblasts, alleviating myotonic dystrophy phenotypes [160]. It was also found that MBNL2 expression was significantly upregulated in breast and lung cancer cells following treatment with a natural compound, Neobractatin (NBT) and overexpression of MBNL2 inhibited tumour metastasis [161].

Louis, J.M., et al. recently discovered that tumour necrosis factor-alpha (TNF-α) which is a pro-inflammatory cytokine and a crucial contributor to various metabolic syndromes, modulates the expression of MBNL [162, 163]. MBNL1 and 3 exhibited a significant reduction at the transcriptional level in cells under TNF-a-induced metabolic stress [163]. Previous studies showed High glucose induces TNF-α, expression [164] and MBNL1 expression was reduced under high glucose conditions [165] which strengthen the role of TNF-α in regulating MBNL expression. Concordantly, the activation of Cyclooxygenase (COX) stimulates intracellular signals via NFκB, p38, or MAPKs signalling pathway which modulates pro-inflammatory cytokine expression such as TNF-α, interleukin 1 beta (IL1β) and interleukin 6 (IL6) [166]. It was found that Phenylbutazone and other COX-1-selective non-steroidal anti-inflammatory drugs (NSAIDs) could up-regulate MBNL1 expression in myogenic cells via COX-1 inhibition [167, 168]. COX-1 inhibition elevated MBNL1 transcription by suppressing methylation of a specific region in MBNL1 intron 1.

Other small molecules, A plant-derived alkaloide, Harmine and

histone deacetylase inhibitors (HDAC), ISOX and Vorinostat were also identified to upregulate MBNL1 and partially rescue splicing defects in human DM1 myoblasts or fibroblasts [169, 170]. However, the mechanism of action of these small molecules in modulating MBNL expression remains unclear.

On the other hand, down-regulation of CELF1 has been achieved by using two ATP-site-directed PKC (protein kinase C) inhibitors, the imidazolo-oxindole inhibitor C16 and the pyrimidine-based inhibitor C51, which impede phosphorylation of CELF1 and thus lead to a reduction in steady-state levels of the protein. Treatment with either of the two compounds successfully reduced RNA foci and rescued several mis-splicing events in DM1 myoblasts [171]. Another small molecule PKC inhibitor, Ro 31-8220 which appears to have a PKC-independent mechanism-of-action in DM1 is also shown to affect phosphorylation of CELF1 and ameliorate the cardiac phenotype in a DM1 mouse model [172, 173].

Treatment with GSK3β inhibitors (detailed introduction in Section 1.4.2), Tideglusib and lithium has been found to restore normal phosphorylation of Celf1 by correcting cyclin D3 levels in a DM1 mouse model [101, 174]. Phase II trials of orally administered Tideglusib has been completed on patients with congenital DM1 (CDM) and childhood-onset DM1, where it was reported that the drug significantly reduced neuromuscular and cognitive abnormalities in patients [175].

1.5.5 Targeting the mis-spliced genes

Aberrant alternative splicing of genes is a major contributor to the pathogenesis of DM1, and is caused by a loss-of-function of MBNL and a gain-of-function of CELF. This splicing abnormality, known as spliceopathy, leads to the mis-regulation of alternative splicing in hundreds to thousands of genes in both muscle and neural cells.

Notably, Some of the most notable mis-spliced mRNAs include those encoding proteins involved in muscle function and maintenance, such as miss-splicing of specific genes including the muscle chloride channel (*CLCN1*) [62], calcium channel voltage-dependent L type alpha 1S subunit (*CACNA1S*) [64], cardiac troponin T (*cTNT*) [94, 176], bridging integrator 1 (*BIN1*) [177] and insulin receptor (*INSR*) [63] are associated with myotonia, muscle weakness, and insulin resistance, respectively, which are hallmark symptoms of DM1. Hence, reversal of these mis-spliced genes could have great significance to therapeutic interventions for DM. Very few studies have focused on this approach solely because correction of the alternative splicing can be achieved by other means, as a subsequent result of MBNL releasing, $(CUG)^{exp}$ *DMPK* RNA degradation, etc (discussed previously).

For targeting the spliceopathy, T.M. Wheeler, et al. used a morpholino and K. Oana, et al. used Manumycin A to target the *CLCN1* RNA directly and induce suppression of exon 7a during the alternative splicing process [178, 179]. Manumycin A is a natural antibiotic and regarded as an inhibitor of Ras farnesyltransferase [180]. It reduces the activity of H-Ras which is associated with the alternative splicing of *CLCN1* [179]. Delivery of the two drugs in DM1 mouse models rescued the mis-splicing of *CLCN1*, improving the chloride channel function and ameliorating myotonia [178, 179]. These two studies provide evidence that the mis-spliced events can constitute another potent target to develop direct potential therapies for myotonic dystrophy.

However, in addition to the mis-splicing of individual mRNAs has been well-documented in DM, recent research has shed light on the cumulative effect of multiple splice variants on the disease's pathophysiology [72, 181]. The dysregulation of splicing in various genes can disrupt multiple cellular pathways, giving rise to a wide

spectrum of symptoms that manifest across different tissues and cell types. This multisystem nature of DM poses significant challenges when considering targeting mis-spliced genes as a therapeutic strategy.

The heterogeneity and severity of the disease phenotype in DM patients can be attributed to the cumulative impact of these mis-spliced mRNAs and their downstream effects on cellular functions [55, 119]. The intricacies of these intricate cellular dysregulations make it difficult to pinpoint a single target or pathway for therapeutic intervention. Moreover, the diversity of splice variants complicates the development of a one-size-fits-all approach, requiring personalized treatment strategies tailored to individual patients' unique splicing profiles.

Another limitation of targeting mis-spliced genes as a therapeutic strategy lies in the intricate web of interactions between splicing factors and their target genes. RNA-binding proteins, particularly members of the MBNL family, play a central role in DM pathogenesis by sequestering expanded CUG or CCUG repeats in the mRNA [73, 77]. While targeting these mis-spliced mRNAs directly seems like a promising approach, the interconnected network of splicing regulators and their pleiotropic effects on gene expression can create unforeseen consequences. Attempting to modulate splicing in one gene may inadvertently impact other splicing events, leading to unintended outcomes.

Furthermore, the vast number of genes mis-spliced in DM presents a daunting challenge for therapeutic intervention. Identifying and targeting each individual mis-spliced mRNA is a laborious task, and achieving comprehensive correction of all aberrant splicing events may be impractical. Even if successful, there is no guarantee that such correction would translate into meaningful clinical benefits, as

the molecular underpinnings of the disease are multifaceted and not solely dependent on splicing abnormalities.

To address these limitations, a multifaceted approach that targets various aspects of the disease pathology may be necessary. Combining splicing modulation with other therapeutic strategies, such as gene editing, gene silencing, or small molecule interventions, could potentially provide more comprehensive and effective treatments for DM. Additionally, exploring alternative approaches that focus on restoring normal cellular functions rather than specifically targeting mis-spliced genes may offer promising avenues for therapeutic development.

1.5.6 Summary

Despite ongoing efforts to develop therapeutic strategies for DM1, there are limitations to the current approaches that have been developed. One major limitation is that most current therapies focus on managing the symptoms of the disease rather than targeting the underlying genetic defect. While some promising therapeutic strategies are being developed, such as antisense oligonucleotide (ASO) therapy, these approaches are still in the early stages of development and their long-term efficacy and safety have yet to be fully established. Additionally, ASO therapy may not be effective for all patients with DM1, and there may be challenges in delivering the therapy to affected tissues and cells.

Another limitation is the complexity of DM1 itself. The disease is caused by a mutation in the *DMPK* gene that leads to the formation of toxic RNA molecules that disrupt normal cellular processes. However, the exact mechanisms by which these toxic RNA molecules cause muscle weakness, cardiac abnormalities, and cognitive impairments are still not fully understood. As a result, it can be

challenging to develop targeted therapies that address the specific molecular pathways involved in DM1 pathogenesis.

Furthermore, DM1 is a highly variable disease, with a wide range of symptoms and severity between individuals. This heterogeneity can make it difficult to develop a one-size-fits-all treatment approach. Instead, a personalised medicine approach may be necessary, which takes into account the unique genetic, clinical, and environmental factors that contribute to the disease.

Overall, while there has been progress in developing therapeutic strategies for DM1, there are still significant limitations and challenges that need to be overcome in order to develop effective treatments for this complex and variable disease.

1.6 RNA decay

1.6.1 Mechanisms of normal mRNA degradation

Eukaryotic mRNAs are produced with two integral stability determinants, the 5′ 7-methylguanosine cap and the 3′ poly(A) tail. These two parts interact with eIF4E and the poly(A)-binding protein (PABP) and therefore form a closed-loop mRNA structure to protect the transcript and to enhance translation initiation. To initiate decay, either one of these two structures must be compromised or the mRNA must be cleaved internally by endonucleases.

Deadenylation-dependent mRNA decay. In eukaryotes, for the bulk of mRNAs, decay is initiated by poly(A) tail shortening (deadenylation). This is an event signalling that mRNA should be destined to degradation. Shortened 3′-terminal adenosine tails serve as a landing pad for the Lsm1–7/Pat1 complex, which stimulates cap hydrolysis (decapping) and allows Xrn1-mediated decay in the 5′→3′ direction. Alternatively, unprotected 3′ ends can be targeted by the multi-subunit RNA exosome and degrade mRNA in the 3′→5′ direction (Figure 1-11 A). These two pathways are not mutually

exclusive and the relative contribution of each mechanism remains a subject of debate.

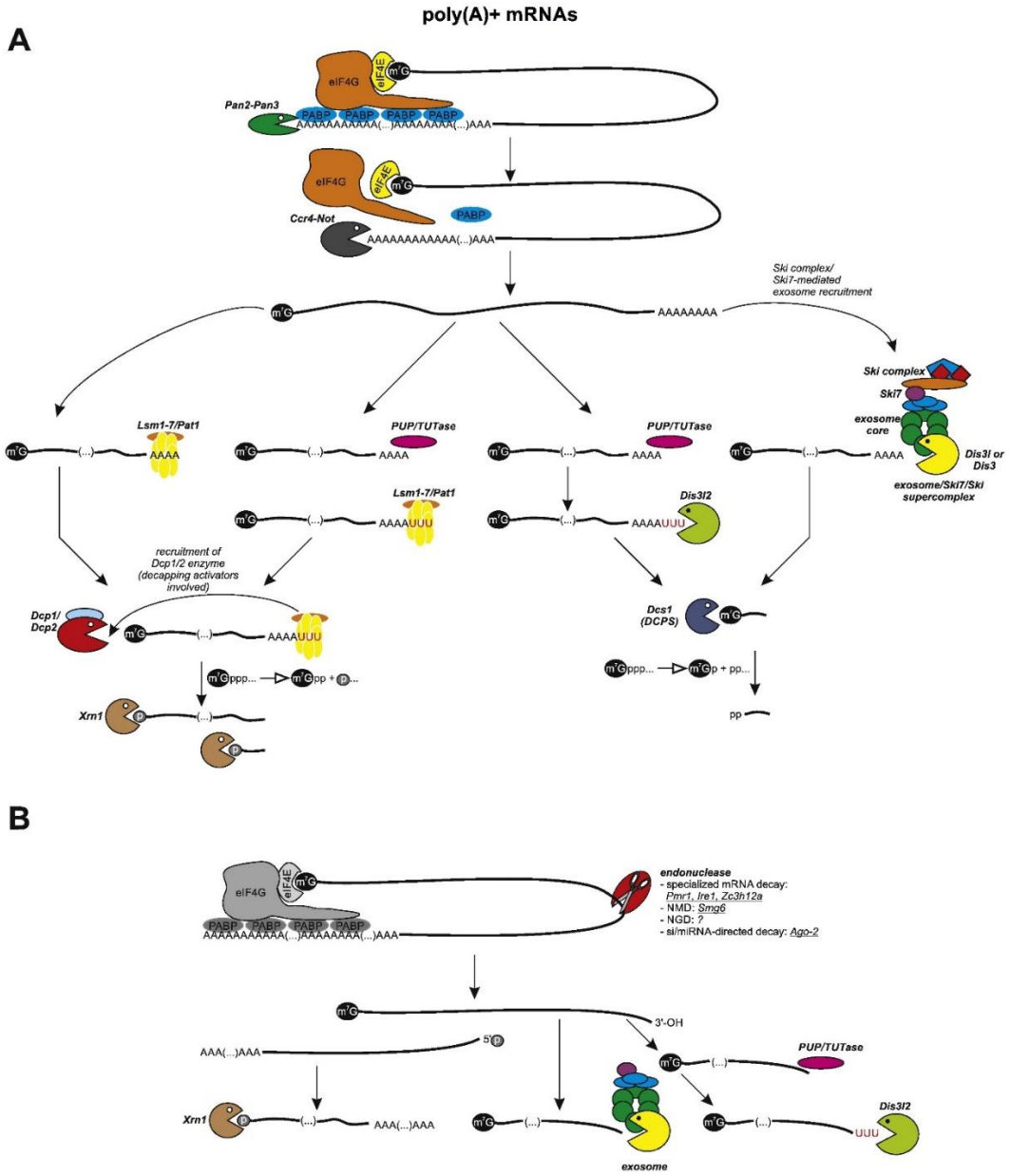
The 5′→3′ mRNA-decay components are concentrated in cytoplasmic RNP granules known as processing bodies (P bodies) or GW bodies [182, 183]. Factors that are involved in translation initiation [184], deadenylation [185], decapping [186], 5′→3′ exonucleolytic decay [187], nonsense-mediated decay [188] and microRNA (miRNA)-mediated RNA decay [189] have been shown to colocalise in P bodies, and the additional observation of mRNA decay intermediates in P bodies support the hypothesis that P-bodies were subcellular sites of mRNA decay and storage. However, it was subsequently demonstrated that macroscopically observable P-bodies are not required for mRNA decay to occur and that mRNAs can be recycled from P bodies and move to translating polysomes [190, 191], suggesting a broader, regulatory role for P bodies in RNA sorting and storing rather than simply acting as sites of mRNA decay.

P bodies assemble when RNA substrates overload the 5′→3′ decay system or when the mRNA decay pathway is impaired, for example, when decapping factors or the XRN1 exonuclease are mutated [185, 186]. Hence, P bodies are dynamic structures that are likely to form when required to regulate the interplay between translation and mRNA decay, and substantially dissipate or disappear when the amount of mRNA waiting to be decayed is reduced.

Endoribonucleolytic decay. Endonucleases initiate degradation through internal cleavage within the mRNA and produce two fragments that are susceptible to exonucleases (Figure 1-11 B). Several cellular endonucleases have been characterized, including such as Pmr1, Ire1, Zc3h12a (MCPIP1), Smg6, the siRNA RISC (RNA-induced silencing complex) and the predominantly nucleolar ribosomal RNA (rRNA)-processing enzyme RNase MRP), which are

involved in the decay of specific mRNA. Furthermore, unidentified endonucleases have been implicated in the decay of aberrant mRNAs [192, 193], and short interfering RNAs (siRNAs) initiate mRNA decay through endonucleolytic cleavage mediated by Argonaute protein-2 (Ago2) [194, 195].

Figure 1-11 An overview of cytoplasmic poly(A) + mRNA decay pathways.



(A) Disruption of the closed circle mRNA structure and deprotection of mRNA termini can trigger various degradation mechanisms. (B) In addition to the specialised pathways, mRNA degradation can also be initiated by endonucleolytic cleavage within the body of the transcript, which is a quality control mechanism. Figure adapted from [196].

1.6.2 Mechanisms of aberrant mRNA degradation in the cytoplasm

Nonsense-mediated decay (NMD) is the most well-studied RNA surveillance mechanism [197] (Figure 1-12 a). NMD detects and degrades transcripts that contain premature termination codons (PTCs). PTCs can arise from mutations, frame-shifts, inefficient processing, leaky translation initiation and extended 3' UTRs. mRNA that contain an extended 3' untranslated region (UTR) can cause normal termination codons to be recognised as premature [198-200]. Thus, *DMPK* mRNA with a long CUG repeat expansion in the 3'UTR is likely to be a target of NMD.

Normally, the ribosome complex would dissociate from the mRNA and release a completed protein. However, in the presence of a PTC, and the ribosome remains stalled at the PTC. At this point, the SMG1 kinase and UPF1 are recruited to the release factors, leading to the formation of a complex known as the SMG-1-UPF1-eRF1-eRF3 complex (SURF) on the stalled ribosome [201]. The UPF1 interacts with the EJC by binding to UPF2, bridging SURF at the ribosome and the EJC. This interaction leads to the phosphorylation of UPF1 by SMG1 and dissociation of the peptide-release factors [201]. This dissociation is necessary to allow other downstream factors, such as SMG5-7 (14-3-3-like proteins), to associate with the complex and activate the mRNA decay machinery [202] (Figure 1-12 a).

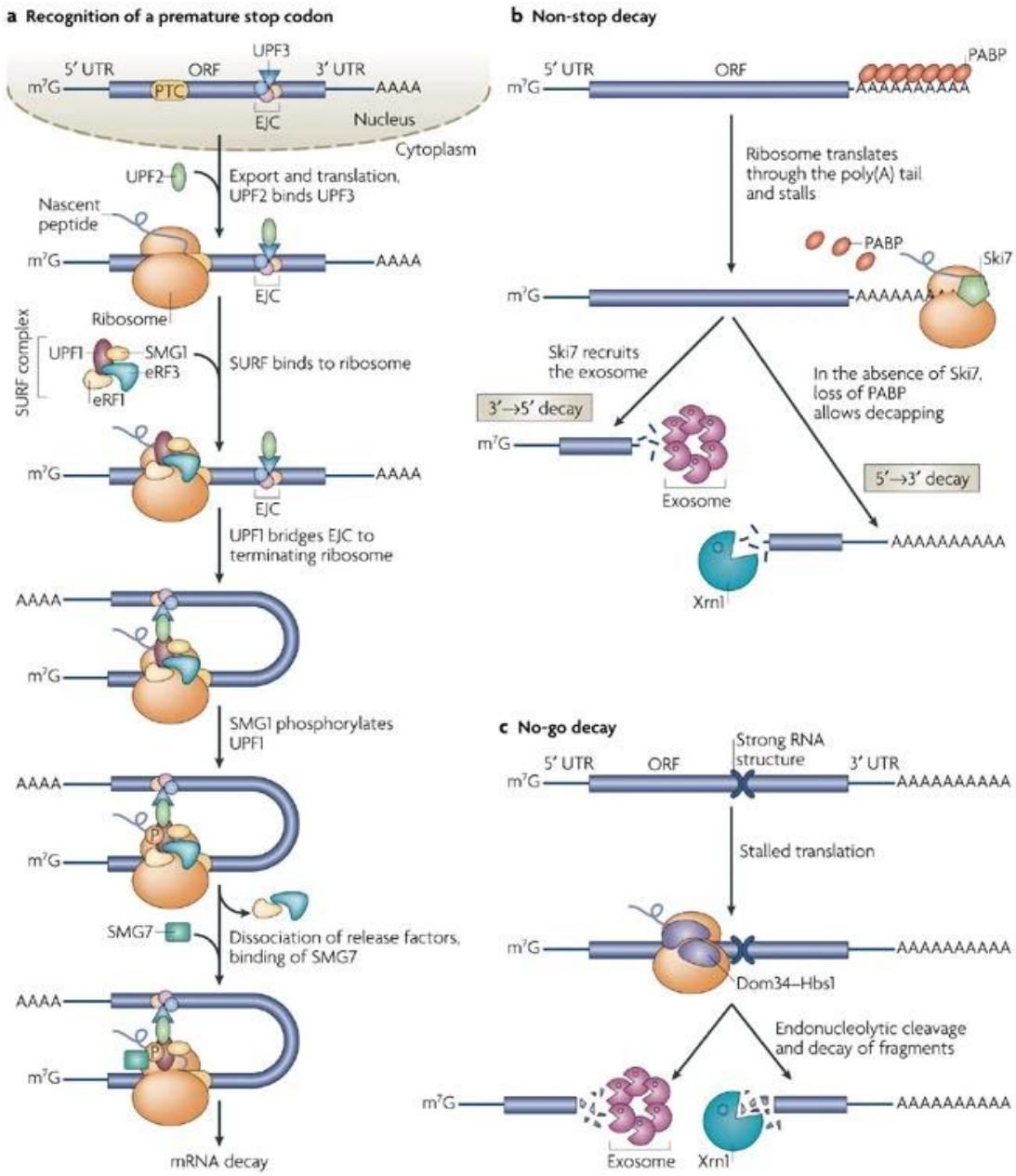
However, an EJC in an inappropriate position is not always the trigger for mRNA decay. In some cases, such as in *D. melanogaster* and *S. cerevisiae* [203], as well as in mammalian immunoglobulin μ-transcripts [204], the EJC is not involved and instead the prolonged distance between the stop codon and the poly(A) tail is sufficient to trigger mRNA decay. The extension of the 3' UTR by the PTC increases the distance from the terminating ribosome to the poly(A)

tail, resulting in an abnormal messenger ribonucleoprotein (mRNP) conformation is then recognised as a signal for faulty RNA [203-205].

It is important to note that the NMD pathway is not limited to the degradation of defective transcripts, but also plays a critical role in regulating normal gene expression. Microarray analyses of the transcriptome in UPF1-depleted mammalian cells have revealed the upregulation of numerous normal transcripts, some of which contain upstream ORFs, 3’ UTR introns, frameshifts, or other features that resemble a PTC [206]. Furthermore, mRNAs lacking any feature resembling a PTC, such as histone mRNAs [207] and the *ARF1* mRNA [208], are also degraded in a UPF1-dependent manner. Intriguingly, in these cases, UPF1 is recruited to the mRNA by a factor bound to the mRNA 3’ UTR: the stem-loop-binding protein (SLBP) in the case of histone mRNAs [207] and STAU1 (Staufen 1) in the case of *ARF1* mRNA [208]. These pathways do not require UPF2 or UPF3, suggesting that they are distinct from classic NMD.

There are two additional mRNA surveillance mechanisms besides NMD: non-stop decay (NSD) and no-go decay (NGD). NSD (Figure 1-12 b) targets mRNAs without a stop codon, which can arise from breaks or the lack of an in-frame stop codon, allowing translation to proceed along the poly(A) tail. Premature polyadenylation may also generate NSD substrates. NSD not only prevents the synthesis of abnormal proteins, but also facilitates ribosome release. NGD is the least understood of these mechanisms and was first identified in yeast [209] (Figure 1-12 c). It recognises ribosomes that have stalled on mRNA and prevents the sequestration of translation factors to faulty transcripts. To achieve this, NGD endonucleolytically cleaves the mRNA near the stalled ribosome, freeing it and the mRNA fragments for decay by the exosome and XRN1 [209].

Figure 1-12 Surveillance for faulty mRNA processing that occurs in the cytoplasm.



(a) Nonsense-mediated decay (NMD) is a mRNA surveillance process that begins with splicing in the nucleus. The exon junction complex (EJC), including UPF3, a core NMD component, is bound to the mRNA and transported to the cytoplasm, where UPF2 joins the complex. As ribosomes translate the mRNA, premature termination codons (PTCs) cause ribosomes to stall, leading to recruitment of the SURF complex, consisting of SMG1, UPF1, and eRF1/3 release factors. UPF1 also binds to UPF2, connecting the EJC to the PTC. SMG1 phosphorylates UPF1, leading to dissociation of eRF1/3 and binding of SMG7. The

final steps of mRNA decay, which remain under investigation, involve various pathways. (b) Non-stop decay occurs when an mRNA lacks a stop codon, causing ribosomes to stall at the 3' end of the mRNA after reaching the poly(A) tail and displacing poly(A)-binding protein (PABP). (c) No-go decay takes place when ribosomes stall within the open reading frame (ORF) upon encountering a robust secondary RNA structure. It triggers an endonucleolytic cleavage event near the stall site, releasing the ribosome and generating two mRNA fragments, each with a free end for decay by the exosome and XRN1 exonucleases, respectively. Figure adapted From [210].

1.6.3 Nuclear RNA decay

The major 5’ exonucleases in eukaryotes are the related proteins XRN1, which is predominantly cytoplasmic, and XRN2/Rat1, which is predominantly nuclear [211]. The activity of XRN2 is largely blocked by the presence of a 5’ triphosphate and the further protection of a (m7G) cap to 5’ end of the RNA. Transcripts that fail to be capped, or on which the cap is not 7-methylated, are rapidly decapped and dephosphorylated by Rai1 and are subsequently exposed to be degraded by XRN2.

In eukaryotic cells, the nuclear exosome complex is the primary 3' exonuclease responsible for RNA surveillance. Originally characterized for its involvement in the processing of pre-stable RNAs, such as rRNA, snoRNA, and snRNA, the exosome has been found to degrade various transcripts including defective pre-rRNAs, pre-tRNAs, aberrant mRNAs, and transcripts from pervasive transcription [212, 213].

Structurally, the exosome is composed of a hexameric ring surrounded by a barrel-shaped structure with nine proteins, and a central channel that allows for the passage of single-stranded RNA. Within this structure, the 3' to 5' exonuclease DIS3/Rrp44 is located at the base of the barrel and is responsible for the catalytic activity required for RNA degradation [214, 215]. DIS3 also has an

endonuclease domain, although its *in vivo* targets are not fully understood. Another exonuclease, EXOSC10/Rrp6, can also associate with the exosome through its cofactor Rrp47 (C1D in humans) [216] and together with DIS3, they function in eliminating faulty RNA in the nucleus [217]. While there is some specialisation between the two exonucleases, there is significant overlap in their target sets [218, 219].

Nuclear-retained RNA is subject to turnover, as demonstrated by the highly unstable nature of pervasive transcripts in both yeast and human nuclei [220, 221]. However, long-lived functional non-coding RNAs, including snRNAs, snoRNAs, NEAT1, and MALAT1, persist in the nucleus for extended periods of time, likely due to their tightly packaged structures and specialised protective features. While protein-coding RNAs and short-lived pervasive transcripts share some characteristics, only protein-coding RNAs are generally resistant to nuclear decay [222]. This is because the 5’ 7-methylguanosine cap and the 3’ poly(A) tail only offer short-term protection, and therefore nuclear residence time appears to be a key factor in determining RNA stability [223, 224]. Specifically, mRNA is rapidly exported from the nucleus, allowing it to escape the degradative environment, while non-coding RNAs lacking export competence remain in the nucleus for a longer period of time and are therefore more prone to decay [225]. This notion was evidenced by the observation that inhibition of yeast mRNA export factors triggers the nuclear degradation of highly expressed heat-shock mRNA [226]. Consistently, Tudek et al. (2018) demonstrated that newly synthesized protein-coding transcripts are sufficiently degraded as a result of nuclear export inhibition via depletion of main mRNA export adaptors or disruption of the nuclear pore structure [227]. It is worth noting that prolonged nuclear residence is a key feature of the $(CUG)^{exp}$ *DMPK* transcript, therefore it may be

subject to nuclear RNA surveillance.

Nuclear poly(A)-binding proteins (PABPs) have dual functions in regulating RNA stability and decay. Upon transcription, mRNAs are quickly polyadenylated, a process coupled to packaging by PABPs. If poly(A) binding is delayed, the 3’ end becomes vulnerable to exosome-mediated degradation before the transcript can be exported to the cytoplasm. Research by Schmid et al. (2015) found that depletion of Nab2, a *S. cerevisiae* nuclear PABP, led to extensive post-transcriptional decay of newly produced poly(A)-tailed RNAs, indicating the critical role of PABPs in immediate 3’ -end protection [228]. However, nuclear PABPs can also recruit the exosome. Transcripts with significantly delayed nuclear export will eventually degrade, albeit at a slower rate. [211]. The human nuclear protein PABPN1 targets full-length and polyadenylated transcripts, and it stimulates decay while bound to the poly(A) tail,[211], possibly by creating a free and accessible 3’ end for the exosome. Consistently, in human cells, sensitivity of transcripts to PABPN1-mediated decay has been shown to be correlated with prolonged nuclear retention [223, 224].

A ‘timed nuclear decay’ model has therefore been invoked to justify the paradoxical activity of PABPs: it is protective of newly produced poly(A)$^+$ RNA, while inducing decay of transcripts with prolonged nuclear residence by interacting with slow-acting nuclear deadenylases [229-231]. However, the mechanism underlying this timed nuclear decay pathway still needs to be further investigated.

It is well-established that most mRNAs, presumably including the wild type *DMPK* mRNA, undergo deadenylation-dependent decay in the cytoplasm [232, 233]. However, due to the presence of extraordinarily long CUG repeat tract in the 3’ UTR and MBNL sequestration by the CUG repeats, mutant *DMPK* mRNAs fail to be

exported to the cytoplasm. Instead, they may principally undergo nuclear mRNA decay pathway. A study on C9orf72 FTLD/ALS which is caused by a (C_4G_2) hexanucleotide repeat expansion in the *C9orf72* gene demonstrated that EXOSC10 can degrade *C9orf72*-derived repeat RNA and EXOSC10 repression can enhance RNA foci formation and (C_4G_2) repeat transcripts retention in patient-derived cells [234]. Analogously, I predict XRN2 and EXOSC10 participate in the degradation of mutant *DMPK* mRNA in DM1.

1.6.4 UPF1: beyond nonsense-mediated decay

Examination of the subcellular localisation of UPF1 in *Drosophila* showed that UPF1 constantly shuttles between the nucleus and cytoplasm.

The hypothesis of a SURF complex that forms at ribosomes terminating prematurely remains valid, but recent discoveries have shown that UPF1 has some level of binding throughout the length of all transcripts. UPF1 bound to the coding region of an mRNA is dislodged by ribosomes during translation. As a result, UPF1 is primarily found on the 3' UTRs in a way that is dependent on their length, suggesting that UPF1 plays a role in sensing 3' UTR length. [235].

Although UPF1 can bind to all 3’ UTRs, its ATPase activity has been reported to disengage UPF1 preferentially from non-target mRNAs [236]. UPF1 binding to ATP reduces its affinity for RNA and activates its helicase activity [237, 238]. Mutated UPF1 proteins that lose the ability of binding or hydrolysing ATP display enhanced RNA binding and are hyperphosphorylated and locked on target and non-target mRNAs by SMG5/7 [236, 237, 239]. While UPF1 mutants with hyperactive ATPase activity lose their RNA binding competence.

Recently, it has been reported that PTBP1, as well as hnRNP L and PABP which were previously known to protect mRNAs from

degradation, promote the preferential ATPase-dependent release of UPF1 from non-target RNAs [240-242]. However, while incremental hyperphosphorylation of UPF1 bound to NMD targets and preferential ATPase-dependent release from non-targets can distinguish between these transcripts, these processes alone are insufficient to trigger or prevent mRNA decay [236, 243]. Finally, a recent report provided evidence that in *D. melanogaster*, UPF1 is co-transcriptionally loaded onto mRNAs and is required for releasing mRNAs from transcription sites and their export from the nucleus to the cytoplasm, confirming that UPF1 recruitment to target RNAs is not specific to NMD [244]. It remains to be determined whether this is also the case in mammals.

Additionally, UPF1 is involved in the Staufen-Mediated mRNA Decay (SMD). In Mammalia, two *D. melanogaster* Staufen paralogs, STAU1 and STAU2, regulate the expression of partially overlapping sets of target mRNAs [245]. In the SMD pathway, STAU binding sites are double-stranded RNA structures in the 3′ UTR, usually generated by intramolecular base-pairing [246]. Similar to NMD, SMD requires translation termination and the recruitment and phosphorylation of UPF1 [247]. In the SMD pathway, UPF2 binds to a STAU1 dimer and recruits UPF1 to the STAU1 mRNP, which promotes the phosphorylation of UPF1. This, in turn, results in the recruitment of RNA degrading enzymes, including SMG5-7, and ultimately leads to mRNA degradation. Notably, UPF2 binds to UPF3 in the classic NMD pathway and recruits UPF1 to the UPF3 mRNP. Therefore, STAU1 and UPF3 may compete for binding to UPF2 since they both contact the MIF4G domain of UPF2. This observation provides a potential explanation for the competition between SMD and NMD [248].

Studies by Aymeric et al. (2012) showed that Staufen1 interacts directly with *DMPK* mRNA and binds to it in a manner that depends on the length of CUG repeats [249]. This suggests that in DM1, the

RNA hairpin structures formed by the expanded CUG repeats in the 3'UTR of *DMPK* mRNA are very likely to be the binding sites of STAU proteins. Consequently, $(CUG)^{exp}$ *DMPK* mRNA may be subject to SMD. Additionally, studies by Garcia et al. (2014) demonstrated that UPF1 plays a conserved role in regulating CUG repeat RNA transcript levels and toxicity [250]. For example, inactivating UPF1 protein by RNAi in human DM1 patient fibroblasts causes an increase in the number of toxic RNA foci [250]. Taken together, I hypothesise that multiple RNA decay pathways are involved in the degradation of mutant *DMPK* mRNA.

1.7 Super-Resolution Microscopy

Fluorescence microscopy has revolutionized our understanding of biological systems by enabling the direct visualisation of molecular interactions with high specificity and multicolour imaging capabilities. However, a major challenge of this technique has been the limited spatial resolution imposed by the diffraction of light [251], as described by Ernst Abbe in 1873. This resolution limit restricts the smallest objects that can be resolved by conventional optical microscopes to approximately half of the wavelength of visible light, i.e., ~200 to 300 nm [251]. Consequently, many biological structures occur on shorter-length scales, making it difficult to study them using diffraction-limited microscopy. To overcome this limitation, there has been a strong scientific motivation to develop methods that allow for super-resolution imaging.

In recent years, super-resolution single-molecule localisation microscopy (SMLM) has emerged as a widely used technique for achieving resolutions beyond the diffraction limit. Popular SMLM techniques include photoactivated localisation microscopy (PALM) and stochastic optical reconstruction microscopy (STORM), which can achieve resolutions in the 20- to 30-nm range by exploiting the

temporal separation of individual fluorescing molecules [252, 253]. The key principle of SMLM is the stochastic activation of a sparse subset of fluorophores in each frame, which ensures that their associated point spread functions (PSFs) do not overlap on the detector, allowing for precise localisation of each emitter at the centre of its PSF [254]. After this subset is bleached, a new random subset is activated and imaged in the next frame, and this process is repeated until the full fluorophore population distribution is acquired. The resulting reconstructed image data consists of a list of localisation coordinates from all data frames, along with the associated localisation uncertainties.

1.7.1 Fluorophores for SMLM

The utility of SMLM hinges on the choice of appropriate fluorophores, which must be able to switch reversibly between an active and inactive state under imaging conditions, and possess high contrast ratios [255]. Specifically, they should emit a large number of photons relative to the background, ensuring a high signal-to-noise ratio and allowing for precise localisation [256, 257]. While some commercial fluorophores, such as cyanine-based dyes, Alexa Fluor® dyes, and CF® dyes, are well-suited for SMLM applications, other commonly used dyes, such as fluorescein, lack the necessary properties to be effectively employed in SMLM. As such, careful consideration of fluorophore properties is essential for successful SMLM imaging.

1.7.2 Optics of SMLM

To perform SMLM, a wide-field microscope is required, equipped with standard continuous wave lasers for excitation and activation and a highly sensitive camera capable of detecting single molecules, such as an EM-CCD or an sCMOS camera [258]. EM-CCDs have the advantage of being able to detect single photons with minimal read-

out noise and are well-suited for detecting low photon counts [259, 260]. However, modern sCMOS cameras, while somewhat less sensitive to weak signals and producing inhomogeneous read-out noise that requires correction, have higher frame rates and allow for a larger field of view (FOV) than EM-CCDs, while still achieving similar signal-to-noise ratios for bright dyes [259, 260]. Most commercial or custom microscopes used for SMLM incorporate automated systems to maintain focus on the specimen.

1.7.3 Acquiring data for SMLM

In SMLM, a small subset of fluorophores in the sample are activated by a laser, and they emit photons that are detected by a camera. The activation and emission process is repeated multiple times, with different subsets of fluorophores activated each time [252, 253, 258]. By analysing the positions of the individual fluorophores, a high-resolution image of the sample can be reconstructed. The key to SMLM's high resolution is the ability to precisely localise the activated fluorophores. This is accomplished by using a combination of techniques to determine the position of each individual fluorophore, including single-molecule fitting, sub-pixel localisation, and drift correction [252, 253]. This results in a final image that has a resolution of tens of nanometers or even better.

To minimize the acquisition time in SMLM, it is crucial to have a high number of active fluorophores in each frame, as long as it does not result in PSF overlaps. The density of active fluorophores for photoswitching dyes is regulated by a balance between excitation and activation laser intensities [252, 253]. Moreover, the exposure time of each frame should align with the ON-state lifetime of the fluorophore, typically in the range of 10-100 ms. The number of frames required for image reconstruction depends on the structure being imaged and the desired resolution, and hence can vary

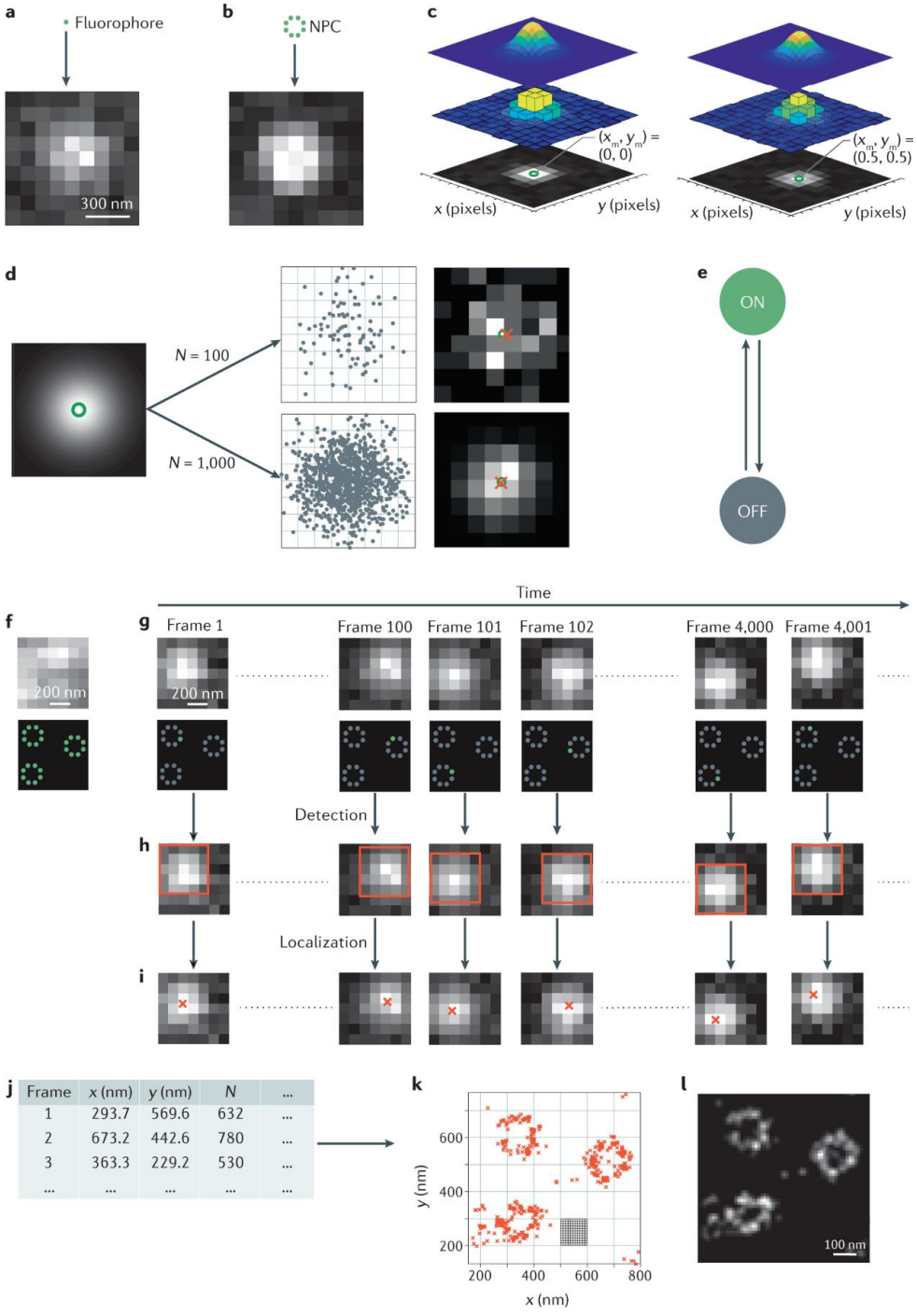
significantly [258, 261]. Here, an nuclear pore complex (NPC) was used as an example to elucidate the principle of SMLM (Figure 1-13). SMLM has been used to image a variety of cellular structures, including the cytoskeleton, endoplasmic reticulum, and synaptic vesicles, and has provided new insights into the organization and dynamics of these structures [261]. However, the technique requires careful optimization and can be time-consuming, and the imaging depth is limited to the top few micrometers of the sample, making it less suitable for imaging deep tissues.

1.7.4 Cluster analysis of data in SMLM

The organization of molecules into clusters is often crucial for their biological function, making cluster analysis of microscopic data a valuable analytical tool. However, conventional microscopes produce images in which each pixel has a numerical value proportional to the intensity at that location, whereas SMLM generates pointillist data consisting of a set of x–y (and possibly z) coordinates with associated localisation uncertainties [258]. As a result, traditional image analysis tools are not suitable for SMLM data, which requires spatial point pattern analysis. Furthermore, there is typically a non-negligible background density of un-clustered points that must be accounted for [262], and each localisation has an associated uncertainty in its position. Recently, a model-based Bayesian statistical approach has been developed to address these issues in SMLM data analysis [263]. This technique generates multiple cluster proposals for each region of interest and scores them against a Bayesian model, taking into account the localisation uncertainty (σ_x, σ_y, σ_z) of each point and avoiding user-supplied analysis parameters. Prior probabilities, which have well-defined statistical interpretations, are required instead. The proposal mechanism is based on a localised version of Ripley's K-function and a topographic prominence thresholding approach, which has been shown to be

advantageous for cluster analysis in SMLM [264]. The highest-scoring cluster proposal is then identified, and key cluster descriptors can be extracted.

Figure 1-13 Principle of single-molecule localisation microscopy (SMLM).



Diffraction-limited microscopy produces fuzzy spots (a) when imaging single fluorophores and causes blurry images of structures such as nuclear pore complexes (NPCs) (b,f). However, SMLM utilises the stochastic switching of fluorophores between active and inactive states (e) to precisely detect and locate single molecules. By analysing subpixel displacements in pixel values of a series of diffraction-limited images, x and y coordinates of a single molecule can be computed with high precision (c, g-i). Higher photon counts improve the signal-to-noise ratio and facilitate more precise localisations (d). The resulting localisation table provides x, y coordinates, and additional information such as the number of photons and localisation uncertainty (j). Accumulated localisation events can be visualised as a scatter plot or a 2D histogram, enabling resolution of NPC structures (k-l). Figure adapted from [258].

Super-resolution imaging techniques surpass the diffraction limit of light and enable the visualisation of structures and interactions that were previously beyond reach. These techniques have greatly expanded our understanding of the molecular organization and dynamics of biological systems. With the advent of super-resolution imaging, I can now directly visualise structures and interactions on a scale of tens of nanometers, providing unprecedented insights into the workings of living systems.

1.8 Hypotheses and Aims

Although our knowledge of the molecular mechanisms underlying DM1 has been greatly extended in recent years, little is known about the degradation of the causal nuclear retained CUG expansion *DMPK* mRNA, and what enzymes are responsible. One of my goals is to address this knowledge gap.

It is widely reported that the fundamental mechanism underlying myotonic dystrophy pathophysiology is the sequestration of MBNL proteins by repeat expansion *DMPK* transcripts. The consensus view is that mutant transcripts sequester MBNL1 and 2, but my preliminary data indicate that it is MBNL1 and 2 that hold or retain the mutant RNA in the nucleus. Although several therapeutic approaches have appeared to eliminate nuclear foci when visualised using in situ hybridisation, they have not successfully reversed the entire pathogenesis of DM1, as it is not clear what happens to the mutant RNA transcripts after the dissolution/disappearance of nuclear foci and. It is important to keep in mind that there is no evidence showing that the foci are the sites of mutant RNA decay. Thus, having a better understanding of the metabolism of mutant *DMPK* mRNA and the effect of $(CUG)^{exp}$ RNA -MBNL protein complex on mutant *DMPK* mRNA decay could have important therapeutic implications for DM.

Based on these considerations, I propose the following hypotheses:

1. Investigating the dynamics of MBNL1 and MBNL2 co-localization with $(CUG)^{exp}$ *DMPK* transcripts will provide insights into the mechanisms underlying the aggregation, nuclear retention, and degradation of mRNA containing CUG-triplet repeats.
2. Utilising super-resolution single-molecule localization microscopy (SMLM) and specific probes designed for different

regions of the *DMPK* mRNA will uncover key findings regarding the formation, stability, and turnover of $(CUG)^{exp}$ foci, including the existence of large and micro foci and the potential role of micro foci as intermediate degradation products.

3. Investigating the participation of RNA decay pathways, including XRN2, EXOSC10, UPF1, and STAU1, in DM1 cells and their derivatives will highlight the critical role played by these pathways in reducing the accumulation of mutant *DMPK* RNA, particularly in the context of MBNL deficiency.

The main aims of my research were:

- To determine the effect of MBNL1 and MBNL2 double knock-out on the nuclear foci formation and the abundance of $(CUG)^{exp}$ *DMPK* transcripts in DM cells.
- To investigate the RNA foci dynamics and identify potential intermediate degradation products of the mutant *DMPK* mRNA by using super-resolution single-molecule localisation microscopy (SMLM).
- To study the contribution of different RNA decay pathways to the degradation of mutant *DMPK* mRNA.
- To unravel the role of MBNL proteins in the mutant *DMPK* mRNA decay.

Chapter 2 Materials and Methods

2.1 Tissue culture

Tissue culture methods were carried out inside a Walker lamina flow safety cabinet (type II) at room temperature unless stated otherwise.

2.2 Cell lines

Cell lines used in this study include human fibroblasts from a non-DM (SBTeloMyoD) person, a DM1 (KBTeloMyoD - 400 CTG repeats) patient. All the cell lines were previously telomerised to allow indefinite growth in culture.

2.3 Media, solutions and antibiotics

The medium used was Dulbecco's Modified Eagle's Medium (DMEM) from Gibco, containing 4.5 g/l glucose and L-glutamine. Other solutions include Fetal Bovine Serum (FBS) from Sigma-Aldrich; Penicillin/Streptomycin from Gibco (10,000 units of Penicillin G and 10,000 mg/ml Streptomycin in 0.9% NaCl); 0.05% Trypsin-EDTA from Gibco; Dulbecco's Phosphate Buffer Solution (DPBS) from Gibco; Dimethyl Sulphoxide (DMSO) from Sigma-Aldrich; Hygromycin B [50 mg/ml] from Thermo Fisher Scientific; and G418 solution (neomycin) [50 mg/mL] from Roche.

2.3.1 Cell culture

All cells were grown and maintained in a humidified incubator at 37 ℃ with 5% CO_2. The media used to grow them was Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 15% Fetal Bovine Serum (FBS) and 2.5% of the antibiotics penicillin (10,000 units of Penicillin G)/streptomycin (10,000 mg/ml).

2.3.2 Cell passaging

Cells were visualised regularly under the microscope and sub-cultured after reaching 80—100% confluency. For passaging, media

was removed from the T75 flask using serological pipettes and cells were briefly washed with 6 ml of prewarmed DPBS, followed by the addition of 2.75 ml of 0.05% Trypsin-EDTA. Flasks were kept for 5 minutes in the incubator at 37 ℃ to allow the dissociation of the cell monolayer from the bottom of the flask. Confirmation of full dissociation of the cells was obtained by looking at them under the microscope, and 8.25 ml of fresh DMEM media were added to the flasks. Each flask was subcultured into two or three T75 flasks, depending on the required cell confluency for future experiments, and fresh media was added to a final volume of 18 ml.

2.3.3 Cell count

Cell number was measured using a haemocytometer, which contains two sets of 16 squares, each of which is divided into 16 smaller squares. 0.5—1 ml of cell suspension was placed into the haemocytometer. Number of cells within five of these 16 squares was counted and an average taken. To estimate the number of cells per ml of suspension, the average number of cells was multiplied by 5,000, as the depth of the haemocytometer was 0.2 mm.

2.3.4 Cryopreservation of cells

Ten percent of DMSO was used as a cryopreservant for human fibroblasts. Media was removed from confluent T75 flasks with serological pipettes and cells were washed with PBS prior to the addition of 2.75 ml of 0.05% Trypsin-EDTA. Flasks were incubated at 37 ℃ for about 5 minutes and complete dissociation of the monolayer of cells was confirmed by visualisation of the cells under the microscope. Cells were resuspended in 10 ml of fresh media and transferred to a 15 ml falcon tube to pellet them by centrifugation at 2,500 g for 10 min. The cell pellet was resuspended in 1 ml of 10% DMSO and the solution containing the cells was transferred to a

cryotube and stored at -80 ℃ for 24 hours. The next day the cells were moved into liquid nitrogen for long term storage.

2.3.5 Defrosting cells from liquid nitrogen

Cryotubes containing the frozen cells were removed from liquid nitrogen and thawed by immediate incubation in the water bath at 37 ℃. 1 ml of fresh media was added and cells were transferred to a 15ml falcon tube containing 10 ml of fresh media. Cells were pelleted by centrifugation at 1000 g for 10 minutes and gently resuspended in 1 ml of media. Finally cells were transferred to a T75 flask containing 18 ml of pre-warmed media and placed in an incubator at 37 ℃ and 5% CO_2.

2.4 RNA Fluorescence In Situ Hybridisation (FISH)

2.4.1 FISH using 96-well Plates and Celldiscoverer 7 System

RNA FISH was performed in all cell lines in order to visualise bright foci which contain mutant RNA transcripts with expanded CUG repeats. Cells were grown in 96 well-plates for 72 hours before fixation. Media was removed from cells and they were washed thrice with DPBS. The fixation step was performed by incubating cells in 4% PFA for 30 minutes at room temperature. Cells were washed again thrice with PBS and then permeabilised in 80% ethanol for 3 hours at room temperature. After another three washes with PBS, cells were incubated in the pre-hybridisation solution for 15 minutes at room temperature. The pre-hybridisation solution contains: 40% formamide, 10% 20x SSC and 50% DEPC water. The prehybridisation solution was then replaced with the hybridisation solution. Each 10ml of hybridisation solution contained: 40% formamide, 10% 20x SSC, 50% DEPC water, 10 µl of 200mM vanadyl ribonucleoside complex (VRC, RNase inhibitor), 100 µl of 10 mg/ml ssDNA (sperm salmon DNA), 20 µl of 20 mg/ml BSA and 500ng of Cy3-labelled $(CAG)_{10}$ probes. Cells were incubated

overnight at 37 ℃ in the incubator. The next day, cells were washed thrice with PBS + 5mM $MgCl_2$, stained for 40 minutes at room temperature with Hoechst 33342 (fluorescent stain for DNA) and finally left in PBS +$MgCl_2$. Cells were then ready for imaging using the ZEISS Celldiscoverer 7 microscope.

2.4.2 FISH using 6-well plates and Super Resolution Microscopy

In order to visualise not only bright RNA foci, but also micro RNA foci and intermediate degradation products of the $(CUG)^{exp}$ *DMPK* mRNA in cells, I employed stochastic optical reconstruction microscopy (STORM), a type of super-resolution microscopy which surpasses the diffraction limit and achieves resolutions in the 20- to 30-nm range. Detection using STORM requires superbly high sensitivity, which means fluorescent contaminants or unspecific binding of probes to non-target structures is much more problematic than for conventional light microscopy. Therefore, the sample preparation should be as clean as possible. All the tools used in STORM-related experiments were cleaned with distilled water three times and sterilised in autoclaves beforehand.

Cells were seeded in a 6-well plate with a cover glass in each well and grown for 72 hours before fixation. Media was removed from cells and they were washed thrice with DPBS. The fixation step was performed by incubating cells in 4% PFA for 30 minutes at room temperature. Cells were washed again thrice with PBS and then permeabilised in 80% ethanol for 3 hours at room temperature. After another three washes with PBS, cells were incubated in the pre-hybridisation solution for 15 minutes at room temperature. The pre-hybridisation solution contained: 40% formamide, 10% 20x SSC and 50% DEPC water. The prehybridisation solution was then replaced with the hybridisation solution. Each 10 ml of hybridisation solution

contained: 40% formamide, 10% 20x SSC, 50% DEPC water, 10 µl of 200mM vanadyl ribonucleoside complex (VRC, RNase inhibitor), 100 µl of 10 mg/ml ssDNA (sperm salmon DNA), 20 µl of 20 mg/ml BSA and 50ng of Atto647n-labelled or Atto550-labelled probes (Table 2-1). Cells were incubated overnight at 37℃ in the incubator. The next day, cells were washed thrice with PBS + 5mM $MgCl_2$, stained for 40 minutes at room temperature with Hoechst 33342 (fluorescent stain for DNA) and finally left in PBS +$MgCl_2$. Cells were then ready for imaging using the Zeiss Elyra Super Resolution Microscope.

Table 2-1 List of probes used in the super-resolution microscopy experiments.

		ID	Probe sequence 5'-3'	Tm (℃)
Single-labelling		Atto647n-$(CAG)_8$	Atto647n-CAGCAGCAGCAG CAGCAGCAGCAG	75.1
Dual-labelling	Set 1	Atto550-$(CAG)_8$	Atto550-CAGCAGCAGCAG CAGCAGCAGCAG	75.1
		5'-DMPK-1	Atto647n-CGCTCGGAGCGGTTGTGAACTG	66.2
		5'-DMPK-2	Atto647n-ACAGAACAACGGCGAACAGGAGCA	64.7
		5'-DMPK-3	Atto647n-ACTGCCACTTCAGCTGTTTCATCCTGT	65.6
		5'-DMPK-4	Atto647n-CCTTCCCGAATGTCCGACAGTGTCT	64.3
		5'-DMPK-5	Atto647n-AGAGGTGCTCCTTGTAGTGGACGAT	64.2
		5'-DMPK-6	Atto647n-GGTCCCCATTCACCAACACGTCC	64.4
		5'-DMPK-7	Atto647n-AGTCGGACCTCCTTAAGCCTCACC	64.4

		5′-DMPK-8	Atto647n-ACATGTTGGACAGGCAGCACCATG	65.2
	Set 2	Atto550-$(CAG)_8$	Atto550-CAGCAGCAGCAG CAGCAGCAGCAG	75.1
		3′-DMPK-1	Atto647n-GGGCCTTTTATTCGCGAGGGTC	65.4
		3′-DMPK-2	Atto647n-CGGGGTCTCAGTGCATCCAAAACG	65.3
		3′-DMPK-3	Atto647n-GGGTCCTGTAGCCTGTCAGCGA	64.8
		3′-DMPK-4	Atto647n-TAAATATCCAAACCGCCGAAGCGGGC	64.6
		3′-DMPK-5	Atto647n-CCCCAGAGCAGGGCGTCATGCA	65.2
		3′-DMPK-6	Atto647n-GGGGTGCGTGGAGGATGGAACAC	65.5
		3′-DMPK-7	Atto647n-GCAGTTTGCCCATCCACGTCAGG	66.1
		3′-DMPK-8	Atto647n-GCCTCAGCCTGGCCGAAAGAAAGA	64.9

2.5 Immunocytochemistry (ICC)

Cells were seeded onto 96-well plates allowed to grow for 72 hours.

_Fixation

Media was removed and cells were washed thrice with DPBS, 5 minutes for each wash and then fixed and permeabilised with ice cold methanol: acetone (50:50) solution for 30 minutes at room temperature. Cells were washed again with DPBS before proceeding with the blocking step.

_Blocking

Cells were blocked in 5% BSA + 5% sheep serum in DPBS for 1 hour at room temperature. The blocking step decreases the nonspecific interactions between proteins and antibodies, thus reduces background.

_Antibody staining

The blocking solution was removed and the cells were incubated with 1% BSA in DPBS with the antibody at the appropriate dilution (mouse MBNL1 1:2,000, mouse MBNL2 1:1,000) overnight at 4℃. The next day the cells were washed thrice in DPBS and incubated with the secondary antibody diluted 1:500 in 1% BSA in PBS for 1 hour at room temperature. The secondary antibodies used were AlexaFluor 488 or 647 goat anti-mouse. The cells were rinsed again in DPBS and stained with Hoechst 33342 for 40 minutes at room temperature, to stain the nuclei of the cells. Finally Hoechst 33342 was replaced with PBS and the cells were imaged using the Widefield plate reader.

2.6 Combined immunocytochemistry and in-situ hybridisation (ICC + FISH)

To study the co-localisation of RNA foci with MBNL proteins, cells were seeded in a 6-well plate with a cover glass in each well and grown for 72 hours, after which they were fixed and permeabilised with 50:50 ice cold methanol: acetone for 30 minutes at room temperature. Cells were blocked with 5% BSA and 5% sheep serum in PBS for 1 hour at room temperature. Anti-MBNL1 or anti-MBNL2 mouse monoclonal antibody was added to the cells at a dilution 1:1, 000 and incubated overnight at 4℃, followed by staining with the secondary antibody AlexaFluor647 goat anti-mouse (1:500) for 30 minutes. Cells were fixed again in 4% PFA for 30 minutes and incubated with the pre-hybridisation solution (50% DEPC water, 40% formamide and 10% 20xSSC) for 15 minutes. Finally cells were incubated with a Atto550-labelled $(CAG)_8$ probe overnight at 37℃ and the next day stained with Hoechst 33342 for 40 minutes and imaged using the Zeiss Elyra Super Resolution Microscope. Images were analysed using the Zen Black software.

2.7 Molecular Biology Techniques

2.7.1 PCR Methods

PCR amplifications were performed in either a Bio-Rad or a G-Storm Thermocycler in 20, 25 or 50 µl reactions. All reactions consisted of 10-20 ng/µl of cell DNA or 1—2 ng/µl of plasmid DNA. The primers used were 20—28 nucleotides long with a GC content between 35—65%.

_Standard PCR

MegaMix 2X (Microzone) Taq polymerase was used for standard PCR amplification. The next reagents were added to the previous amounts of DNA: 2 µl of 10 mM forward and reverse primers, 12.5 µl of MegaMix 2X and water up to a final volume of 25 µl. PCR conditions were as follows: 1 cycle of 5 minutes at 95 ℃, 30-32 cycles of 30 seconds at 95 ℃, 30 seconds at appropriate annealing temperature (53—70 ℃) and 1 minute at 72 ℃, with a final cycle at 72 ℃ for 5—10 minutes.

_High Fidelity PCR

For sequences to be cloned or to screen CRISPR-generated clones, High Fidelity Phusion polymerase was used in a reaction containing the DNA, 1.5 µl of 10 mM forward and reverse primers, 1 µl of 10 mM dNTPs mix (NEB), 1.5 mM $MgCl_2$, 5 µl of High Fidelity (HF) PCR buffer (NEB), 0.8 µl of Phusion polymerase and water up to 25 µl. For 50 µl reactions amounts were doubled. PCR conditions were: 1 cycle at 98 ℃ for 30 seconds, 30 cycles of 10 seconds at 98 ℃, 30 seconds at 50—65 ℃ and 30 seconds at 72 ℃, followed by a final cycle at 72 ℃ for 10 minutes.

_Long Amplification PCR

To amplify fragments bigger than 1.5 Kb a Long Amplification Taq Mix (NEB) was used in a reaction containing DNA, 1 ml of 10 mM

forward and reverse primers, 12.5 ml of Long Amp Taq Mix and water up to a final volume of 20 ml. PCR conditions were: 1 cycle at 94_ C for 30 seconds, 30—32 cycles at 94℃ for 30 seconds, 59—61℃ for 30 seconds, 65℃ for 4—5 minutes, and a final cycle at 65℃ for 10 minutes.

_Nanoplate Digital PCR

To detect normal and $(CUG)^{exp}$ *DMPK* mRNAs using the Nanoplate Digital PCR method, primers and probes were configured and purchased at Geneglobe: https://geneglobe.qiagen.com/us/product-groups/dpcr-lna-mutation-assays, using the dPCR LNA Mutation Assays (QIAGEN).

A reaction mix was prepared according to Table 2-2. Due to the hot-start, it is not necessary to keep samples on ice during reaction setup or while programming the real-time cycler. Template cDNA amount should not exceed 50 ng/reaction for 26k 24-well plates. After 40µl of reaction mix was added into each well, the nanoplate was sealed properly using the QIAcuity Nanoplate Seal provided in the QIAcuity Nanoplate Kits.

Table 2-2 dPCR setup components.

Component	Volume per reaction
4x Probe PCR Master Mix	10 µl
30x dPCR LNA Mutation Assay (FAM/HEX)	1.3 µl
Nuclease-free water	Variable
Template cDNA	Variable
Total reaction volume	40 µl

Cycling conditions were set under the dPCR parameters in the QIAcuity Software Suite or at the QIAcuity instrument: 1 cycle at 95℃ for 2 minutes, 40 cycles at 95℃ for 15 seconds, 60℃ for 30 seconds.

Data was analysed using the QIAcuity Software Suite.

2.7.2 Gel Electrophoresis

Agarose gels were used to visualise plasmids and separate DNA fragments based on their size. Standard 1% or 2% gels were prepared by dissolving 1 g or 2 g of agarose, respectively, in 100 ml of 1x TAE (40 mM Tris-HCl, 50 mM EDTA/pH 8.0 and 0.1% (w/v) glacial acetic acid). The solution was heated in the microwave and then cooled. 4 µl of 10 mg/ml ethidium bromide was added to the gel and let it solidify. Before loading the DNA samples into the wells, 1 µl of loading buffer (30% glycerol, 0.025% bromophenol blue and 0.025% xylene cyanol) was added to each sample. Marker (100 bp or 1Kb ladder) was always loaded along with the samples and gels electrophoresed at 80—110V for 0.5—1.5 hours in 1x TAE buffer. Gels were resolved using a BioRad GelDoc system where DNA was visualised under UV light.

2.7.3 Gel Extraction and Purification

DNA extraction of PCR fragments from agarose gels after electrophoresis was carried out by excising the bands with a surgical blade and using the QIAGEN Gel Extraction Kit following manufacturers' protocol. Concentration and quality of the DNA was measured using a Nanodrop 2000 spectrophotometer.

2.7.4 DNA Sequencing

6 ul of plasmid DNA (100 ng/µl) or 6—10 µl of PCR products (10 ng/µl) were sent, along with an appropriate set of primers, to a

sequencing facility in order to sequence the DNA. Sequences were then analysed using SnapGene.

2.7.5 DNA Extraction

Genomic DNA from fibroblast cells was extracted using the QIAGEN DNeasy Blood & Tissue kit. Briefly, about 2 million cells per flask were trypsinized, transferred to a 15 ml falcon tube and centrifuged for 5 minutes at 300 g. The pellet was resuspended in 200 µl of PBS and transferred to a 1.5 ml Eppendorf tube. 20 µl of 10 mg/ml proteinase K and 4 µl of 10 mg/ml RNase A were added to the cells followed by an incubation of 2 minutes at room temperature in order to get RNA-free genomic DNA. Cells were lysed by adding 200 µl of buffer AL and incubated for 10 minutes at 56℃. 200 µl of 100% ethanol was added and the mixture was transferred to a DNeasy Mini spin column. After a centrifugation of 1 minute at 8,000 g and two washes with AW1 and AW2 washing buffers, also followed by centrifugation steps, the DNA was finally eluted in 140—200 ul of sterile distilled water and stored at -20℃.

2.7.6 Total RNA Extraction

Total RNA from fibroblast cells was extracted using the Purelink RNA mini Kit. Briefly, 0.6 ml of Lysis Buffer prepared with 0.1% 2-mercaptoethanol was added to a cell pellet containing about 2 million cells using RNase-free pipette tips. Lysate was then transferred to a Homogenizer inserted in a Collection Tube and centrifuged at 12,000 g for 2 minutes. Homogenizer was removed when done. 0.6 ml of 70% ethanol was to each cell homogenate, which was then vortexed to mix thoroughly and to disperse any visible precipitate that may form after adding ethanol. 700 µl of the sample (including any remaining precipitate) was transferred to the Spin Cartridge (with the Collection Tube) and centrifuged at 12,000 × g for 15 seconds at room temperature. Flow-through was then discarded and the rest

of the sample was processed following the same steps above. Samples were then proceeded to On-column PureLink® DNase Treatment, 80 µl of PureLink® DNase mixture was prepared per sample according to Table 2-3.

Table 2-3 DNase treatment components.

Component	Volume
10X DNase I Reaction Buffer	8 µl
Resuspended DNase (~3U/µl)	10 µl
RNase Free Water	62 µl
Final Volume	80 µl

350 µl of Wash Buffer I was added to the Spin Cartridge containing the bound RNA and samples were centrifuged at 12,000 × g for 15 seconds at room temperature. Flow-through was then discarded. 80 µl of PureLink® DNase mixture was added directly onto the surface of the Spin Cartridge membrane and samples were incubated at room temperature for 15 minutes. When the DNase I treatment was done, 350 µl of Wash Buffer I was added to the Spin Cartridge and samples were centrifuged at 12,000 × g for 15 seconds at room temperature. Flow-through was then discarded. Next, 500µl of Wash Buffer II with ethanol to the Spin Cartridge and samples were

centrifuged at 12,000 x g for 15 seconds at room temperature. Flow-through was then discarded. Sample were further centrifuged at 12,000 × g for 1 minute to dry the membrane with bound RNA. Collection tubes were then discarded and a recovery tube was insert to each spin cartridge. Finally, 30 µl–100 µl of RNase–Free Water was added to the centre of the Spin Cartridge and samples were incubated at room temperature for 1 minute. Spin Cartridge and Recovery Tube were then centrifuged for 1 minute at ⩾12,000 x g at room temperature. Purified RNA was proceed to analysing RNA Yield and Quality. The concentration and quality of the purified RNA were measured using the Nanodrop 2000 and the 2100 Bioanalyzer. RNA samples were stored at -80 ℃ for future use.

2.7.7 Nuclear and Cytoplasmic RNA Extraction

Nuclear and Cytoplasmic RNA fractions from fibroblast cells were extracted using the Norgen Cytoplasmic & Nuclear RNA Purification Kit. On-column DNase treatment was applied to the nuclear RNA fractions only following the On-column PureLink® DNase Treatment Protocol to eliminate the contaminating genomic DNA (see 2.7.6). The concentration and quality of the isolated RNA fractions were measured using the Nanodrop 2000 and the 2100 Bioanalyzer. RNA samples were stored at -80℃ for future use.

2.7.8 Poly(A)$^+$ mRNA Extraction

Purification of Poly(A)$^+$ mRNA from fibroblast cells was performed using the Oligotex® mRNA Mini Kit (Qiagen). On-column DNase treatment was applied to the RNA extracts following the On-column PureLink® DNase Treatment Protocol to eliminate the contaminating genomic DNA (see 2.7.6). The concentration and quality of the isolated RNA fractions were measured using the Nanodrop 2000 and the 2100 Bioanalyzer. RNA samples were stored at -80℃ for future use.

2.7.9 5' cap-dependent RNA Extraction

Purification of 5’ capped mRNA from fibroblast cells was performed using the mRNA-ONLY™ Eukaryotic mRNA Isolation Kit (EPICENTRE Biotechnologies). On-column DNase treatment was applied to the RNA extracts following the On-column PureLink® DNase Treatment Protocol to eliminate the contaminating genomic DNA (see 2.7.6). The concentration and quality of the isolated RNA fractions were measured using the Nanodrop 2000 and the 2100 Bioanalyzer. RNA samples were stored at -80 ℃ for future use.

2.7.10 Quantification of DNA and RNA

DNA and RNA samples were quantified using a NanoDrop 2000 Spectrophotometer.

2.7.11 RNA quality assessment using the Bioanalyzer

In order to assess the concentration, quality and integrity of total RNA or nuclear/cytoplasmic RNA fractions prior to use in downstream application, samples were analysed using the Agilent 6000 Nano Kit (Agilent Technologies, USA) and a 2100 Bioanalyzer (Agilent Technologies, USA). Samples were analysed following the protocol provided by the manufacturer. The Bioanalyzer system provides a RIN (RNA Integrity Number) that objectively measures total RNA quality on a scale from 1 to 10, with 10 being the highest integrity. The 18s and 28s rRNA peaks are analysed to evaluate the RNA integrity.

2.7.12 Reverse Transcriptase PCR

RNA was transcribed into cDNA using the Thermo Scientific Verso cDNA Synthesis Kit. 1 µg of RNA was mixed with 4 µl of 5X cDNA synthesis buffer, 2 µl of dNTP Mix, 1µl of random hexamer primers, 1 µl of RT Enhancer, 1 µl of Verso Enzyme Mix and nuclease-free water up to a final volume of 20 µl. The reaction mix was incubated

in a PCR machine as follows: 50 minutes at 45℃ and 3 minutes at 95℃.

2.8 Protein Analysis

2.8.1 Solutions, Buffers and Antibodies

The solutions and buffers described in this section were used and prepared to extract proteins from cultured cells, quantify them and perform western blots. The recipes for each of them are listed:

Lysis buffer:

_ RIPA buffer

_PMSF (Protease inhibitor)

_ dH_2O

The RIPA buffer and protease inhibitor are commercially available and were acquired from Cell Signalling.

NuPAGE mix:

The volumes listed are for one sample. For more samples the volume of each solution needs to be accordingly scaled up.

_ 4x NuPAGE LDS sample buffer, 2.5 ml

_ 10x NuPAGE sample reducing agent, 1 ml

Running buffer:

_ 1x NuPAGE MOPS SDS running buffer (20x)

_ dH_2O

Transfer buffer:

_ 1x NuPAGE transfer buffer

_ 10% Methanol

_ dH_2O

All the NuPAGE solutions were obtained from Novex (Life Technologies).

TBS-Tween:

_ Tris-HCl 50 mM

_ NaCl 150 mM

_ Tween 20 0.05%

_ dH_2O

The pH of the TBS-Tween was adjusted to 7.6 using HCl.

ECL Developing solution:

The developing solution was prepared with the solutions from the commercially available Clarity Western ECL blotting substrate kit. They were mixed at a ratio 1:1.

_ Clarity western peroxide reagent

_ Clarity western luminol/enhancer reagent

Table 2-4 List of primary and secondary antibodies used in this thesis

Antibody	Source	SKU	Working concentration
Mouse monoclonal anti-MBNL1	Gift from Dr. Ian Holt)	MB1a	1:1000 for WB; 1:500 for IF
Mouse monoclonal Anti-MBNL2	Gift from Dr. Ian Holt)	MB2a	1:1000 for WB; 1:500 for IF

Mouse monoclonal anti-Lamin-B (A-11)	Santa Cruz Biotechnology	sc-377000	1:1000 for WB
Mouse monoclonal anti-α-tubulin (TU-02)	Santa Cruz Biotechnology	sc-8035	1:1000 for WB
Anti-mouse IgGκ BP-HRP	Santa Cruz Biotechnology	sc-516102	1:5000 for WB
Mouse monoclonal anti-XRN2 (H-3)	Santa Cruz Biotechnology	sc-365258	1:1000 for WB
Mouse monoclonal anti-EXOSC10 (B-8)	Santa Cruz Biotechnology	sc-374595	1:1000 for WB
Rabbit monoclonal anti-STAU1	abcam	ab137100	1:1000 for WB
Anti-rabbit IgG HRP-linked Antibody	Cell Signalling	7074	1:5000 for WB
Rabbit monoclonal anti-RENT1/hUPF1	abcam	ab133564	1:1000 for WB
Mouse monoclonal anti-GAPDH	Santa Cruz Biotechnology	sc-47724	1:2000 for WB

2.8.2 Total Protein Extraction

Confluent T75 flasks were used for the protein extraction. Cells were washed with pre-warmed PBS and detached from the surface by adding 2.5 ml of Trypsin-EDTA followed by an incubation of 5 minutes at 37℃. 10 ml of media were added to neutralise the trypsin and the cells were transferred to a 15 ml falcon tube. After a centrifugation of 10 minutes at 100 g the pellet of cells was washed with 1 ml of PBS and transferred to an Eppendorf tube. Next, cells were centrifuged again for 3 minutes at 3,000 rpm. The supernatant was decanted and the pellet was resuspended in 200—300 ml of RIPA buffer. The Eppendorf tubes were left on ice for 30 minutes followed by two freeze-thaw cycles at -80℃. Finally the protein lysate was collected by centrifuging at top speed at 4℃ for 30 minutes. The supernatant was collected and stored at -80℃ until further processing.

2.8.3 Nuclear and Cytoplasmic Protein Extraction

Nuclear and cytoplasmic protein fractions were extracted using the NE-PER Nuclear and Cytoplasmic Extraction Reagents.

2.8.4 Protein Quantification

Protein concentration was determined using the Bio-Rad DC Protein assay. Bovine serum albumin at a concentration of 1mg/ml was used to produce a standard curve. 10 ml of the protein sample were mixed with the reagents from the Bio-Rad DC assay according to manufacturer's protocol and absorbance of the protein was measured using a Bio-Rad SmartSpec 3000 Spectrophotometer set to 600 nm. Absorbance from the BSA samples was plotted against known concentrations of BSA ranging from 5 mg/ml to 50 mg/ml and the standard curve was used to estimate the concentration of the protein samples.

2.8.5 Western Blot

Western blot was performed using a commercial NuPage system (Invitrogen) according to the manufacturers' instructions.

_Sample preparation

A volume reflecting 10—30 mg of protein was mixed with 3.5 ml of NuPAGE Mix. Deionized water was added in some cases to adjust the volume to 14 ml in total. Samples were heated at 70℃ for 15 minutes to denature proteins before loading on the gel.

_SDS-PAGE

Before loading the samples on the gel the wells were carefully washed with running buffer and placed in a X-cell Surelock mini-cell system. The chamber was filled with the remaining running buffer and 14 ml of protein sample were loaded onto the wells. SeeBlue marker (acts as a visual control) and Magic Marker were always loaded on the first two wells. Samples were run on a NuPage 10% Bis-Tris gel for 1 hour at 200 V.

_Transfer of proteins to the PVDF membrane

Once the run was complete, a PVDF membrane was placed on top of the gel and in between a sandwich made with filter papers and sponges soaked in Transfer buffer. The sandwich was placed in a Xcell Blot Module in an inverted position, with the top of the gel at the bottom of the Blot Module, and the proteins from the gel were transferred to the membrane for 1.5 hours at 20 V.

_Protein detection

The membrane was rinsed twice in a tray with distilled water (10 minutes each wash) and blocked overnight at 4℃ with 5% Blotto in TBS-Tween. The next day the membrane was rinsed thrice in TBS-Tween and incubated with the primary antibodies anti-CrkRS and

anti-GAPDH, diluted in 3% BSA in TBS-Tween at a dilution 1:1000 and 1:10,000, respectively, for 2 hours at RT. The membrane was rinsed again twice in TBS-Tween and incubated with the secondary antibody anti-mouse conjugated to HRP (1 ml in 10 ml) for 30 minutes at RT. The membrane was washed thrice in TBS-Tween before being incubated for 5 minutes with the ECL Detection solution. The protein bands were then visualised using a LAS-3000 Imaging System.

2.9 Genome Engineering by CRISPR/Cas9 System

2.9.1 Target Genes

The two genes chosen for the CRISPR projects were *MBNL1* and *MBNL2*. All the sequence information for them was taken from NCBI or Ensembl.

2.9.2 gRNAs Design

Sequence from the target genes was taken from Ensembl and analysed using two gRNA design web tools: CRISPR Design from the Zhang Lab and Chopchop v.2. These web tools scan the input sequences in order to find suitable gRNA sequences and compare them with other sites in the genome, giving a final score depending on the quality of the gRNA sequence and the number of off-target effects. For this work, the gRNAs with the higher scores were selected to perform the *in vitro* assay.

2.9.3 *In vitro* Assessment of the gRNA and Cas9 Activities

A gRNA with high activity *in vitro* is more likely to work *in vivo*, so the *in vitro* activity of all the gRNAs was tested prior to the transfection of the cells.

_gRNA template generation

Oligos harbouring the T7 promoter sequence, the 20-nt sequence specific to each gRNA and part of the sequence of the gRNA scaffold were synthesized by Sigma. An additional oligo containing the gRNA scaffold sequence was also synthesised. Each of the oligos was annealed to the scaffold oligo to create a double-strand DNA fragment, to be transcribed into gRNA. For the annealing reaction, 1 ml of each oligo was mixed with 1 ml of the scaffold oligo and 8 ml of water to a final volume of 10 ml. The annealing reaction was carried out in a PCR thermocycler set up as follows: 5 minutes at 95℃, 5 cycles of 2 seconds each, starting at 95℃ and decreasing 2℃ per cycle, then 580 cycles of 1 second, starting at 83℃ and decreasing 0.1℃ per cycle, and a final incubation at 10℃ for 30 seconds. Once the reaction run was complete, the next reagents were added to the mixture: 2.5 ml of dNTPs (10mM), 2 ml of 10x NEB buffer 2.1, 0.5 ml of T4 NEB DNA polymerase and water up to a final volume of 20 ml. This new mixture was incubated at 12℃ for 20 minutes and then PCR purified using the PCR Clean-up kit (Sigma) following manufacturers' protocol. A small aliquot of the product was run on a 1% agarose gel to verify that the size of the product was the expected (120 bp).

_*In vitro* transcription of the gRNAs

The double-stranded oligos were transcribed into gRNA with the MAXIscript T7 kit (Ambion) following manufacturers' instructions. After an incubation of 4 hours at 37℃ with the T7 polymerase, 1 ml of TURBO DNase was added to the reaction and incubated for another 15 minutes at 37℃. To clean up the reaction, 2 ml of 5 M ammonium acetate and 60 ml of ice-cold 100% ethanol were added to the transcription reaction and it was incubated overnight at -20℃. The next day the reaction was centrifuged for 30 minutes at maximum speed at 4℃ and then the RNA was precipitated with 70% ethanol. After another centrifugation step, the RNA present in the

pellet was resuspended in ultra-pure water and its concentration was measured using a Nanodrop spectrophotometer.

_Cas9 *in vitro* assay

Genomic DNA was extracted from DM1 cells using the DNeasy Blood & Tissue kit (QIAGEN), according to the manufacturer's instructions. Fragments harbouring the target site for each gRNA were PCR amplified using the primers MBNL1-F': 5'-GTGTGCCCTTCTCATTGTATCA- 3' and MBNL1-R': 5'-GAAAGTATTTGCACTTTTCCCG- 3' for MBNL1; MBNL2-F': 5'-TTGCCGCTGTTTAAATT- 3' and MBNL2-R': 5'-CAAAAGCCCTAACTGCTGTTC- 3' for MBNL2. Once the PCR product was generated, the next reaction was assembled: 3 µl of 10x Cas9 reaction buffer, 300 ng of the gRNA transcribed, 3 µl of 1 mM Cas9 protein (NEB) and water up to 20 µl. This reaction was incubated for 10 minutes at 25℃ so the gRNA and the Cas9 can form a complex. Next, 180 ng of the PCR product previously amplified was added to the reaction and incubated for 1 hour at 37 ℃. Finally the samples were run on a 2% agarose gel.

2.10 shRNA lentiviral transduction

Cells were seeded on a 6-well plate (0.3 million per well) at 37 ℃ in a humidified 5% CO_2 incubator overnight. On the transduction day, the virus was thawed and mixed gently prior to use. Old media were removed from the plate and 1 ml of fresh medium was added to each well. Appropriate amount of virus (Table 2-5) and 10 µg of Polybrene were placed in each well to achieve the desired MOI. After that, the plate was swirled gently to mix and incubated at 37 ℃ in a humidified 5% CO_2 incubator overnight. Virus-containing media were remove and replaced with fresh media after 12 hours of incubation. Cells were harvested 5 days post transduction for western blot, RNA extraction or RNA FISH.

Table 2-5 List of lentiviral particles used in this thesis.

Type	Gene name	Target sequence 5'-3'	Vector	Source	MOI
Knock-in	MBNL1	cDNA: MBNL1-202 (ENST00000282488.11)	YOE-LV018	Ubigene	5
	MBNL2	cDNA: MBNL2-207 (ENST00000679496.1)	YOE-LV018	Ubigene	5
	UPF1	cDNA: UPF1-201 (ENST00000262803.10)	pLenti-P2A-tGFP	Origene	5
Knock-down	XRN2	TACATAGCTGATCGTTTAAAT	pLKO.1-CMV-neo	Sigma-Aldrich	4
	EXOSC10	sh1:AGTTGACTTGGAGCACCACTCTTACAGGA sh2:TGATACTCATCAGGCAGCACGCCTTCTTA sh3:TCAGGCAGCGAAGTTCGATCCATCAACCA sh4:CTTACGACTACAGCCAGTCAGACTTCAAG	pGFP-C-shLenti	Origene	4 (in total)

	UPF1	sh1:ATCACGGCACAGCAGATCAACAAGCTGGA sh2:CACAACCAGATCAGGAACATGGACAGCAT sh3:TGCCTGCGTGGTTTACTGTAATACCAGCA sh4:CGTGAGAGCCTCATGCAGTTCAGCAAGCC	pGFP-C-shLenti	Origene	4 (in total)
	STAU1	CCAACAGTGATCTGTATTCTT	pLKO.1-CMV-neo	Sigma-Aldrich	4
Negative control	Empty	/	YOE-LV018	Ubigene	5
	Scramble	Non-target: TATGTGCGGCAAACCAAGCG	pLKO.1-CMV-neo	Sigma-Aldrich	4
	Scramble	Non-target: GCGCGATAGCGCGAATATAC	pGFP-C-shLenti	Origene	4
	Empty	/	pLenti-P2A-tGFP	Origene	5

2.11 Microscopy

The imaging of cells for various purposes in this work was performed accordingly with different microscopes (listed below).

2.11.1 Micro Widefield Plate Reader

After performing ICC or FISH, cells were visualised on a Molecular Devices ImageXpress Micro Widefield high content plate reader, using the Nikon 40x objective. Analysis of fluorescent images was performed using MetaXpress software.

2.11.2 Zeiss Celldiscoverer 7 Microscope

After performing ICC and FISH, cells were visualised on a Zeiss Celldiscoverer 7 Microscope (Carl Zeiss, Germany), using a ZEISS Plan-Apochromat 20x/0.95 Autocorr objective. Images were captured and analysed using the ZEN (blue version) software (Carl Zeiss, Germany).

2.11.3 Confocal Microscopy Zeiss LSM880F

Confocal Microscope was used to visualise the cells after ICC or FISH when seeded on coverslips. Cell imaging was performed on a Zeiss

LSM880F confocal microscope (Carl Zeiss, Germany) using a 63x /1.4 oil objective. Images were captured using the ZEN (blue version) software (Carl Zeiss, Germany) and analysis was performed using the Fiji software (open source).

2.11.4 Zeiss Elyra Super Resolution Microscope

Super Resolution Microscope was used for single-molecule localisation in cells. Cell imaging was performed on a Zeiss Elyra Super Resolution Microscope (Carl Zeiss, Germany) using a Zeiss alpha Plan-Apochromat 100x/1.46 oil objective. Images were captured using the ZEN (black version) software (Carl Zeiss,

Germany) and analysis was performed using the Bayesian Cluster Analysis Package (generated by Dr. Dylan Owen).

_Image recording

Cells adhered to 18 mm coverslips (Zeiss cat. no.) were rinsed 1x with distilled water and mounted with 9 ul Abbe buffer (Abbelight) using double sticky spacers (grace Biloabs). Slides were immediately imaged, and not stored. Zeiss Elyra PS1 with alpha-plan apochromat 100x/1.46 NA objective was used in laser widefield mode to perform the STORM experiments. Prior to each STORM recording, the field of view was captured as a snapshot with noise reduction by averaging 16 images. 10000 or 20000 frames were recorded at 25 ms exposure, 642 and 561 lasers were used sequentially for detection of Atto/Alexa647n labelled and Atto550 labelled targets, respectively. Laser power was adjusted according to the characteristics of targets: 100% high laser power for target molecules of high density, 25% laser power for low density. A filter-set (Laser blocking filter LBF -561/642) was used, in this way no filter cube needed to be moved. Therefore the two channels were optically aligned, the x-y mismatch happened due to drifting only, which was corrected digitally in the processing workflow.

_Image processing

Raw files saved as czi (Zeiss) format were imported into the PALM module of the Zeiss Black 2012 software. Localisations were first detected by the modules Gaussian fitting, with PSF (point spread function, representing a single molecule signal in an image) detection setting was set to 9 pixels in size, and peak intensity to noise was set to 6.

Each detection will have x and y coordinates and other characteristics logged into a data table. Based on several parameters,

the total detections were filtered in order to reduce the possible false detections (coming from other emitters, =background).

2.12 Electronic Databases

Information about nucleotide, transcripts and protein sequences was obtained using National Centre for Biotechnology Information (NCBI) and European Molecular Biology Laboratory (Emsembl). Primers were designed using Primer3 Input version 4.0.0.

gRNAs were designed using ChopChop version 2, an online tool that predicts potential gRNAs based on the target sequence and off-target analysis.

2.13 Nanopore Sequencing

For targeted nanopore DNA sequencing with the enrichment of regions of interest (ROI) using CRISPR/Cas9, genomic DNA from fibroblast cells was extracted using the QIAGEN DNeasy Blood & Tissue kit according to the manufacturer's protocol. The DNA concentration and quality were assessed using a NanoDrop spectrophotometer and agarose gel electrophoresis. Guide RNAs (gRNAs) were designed using ChopChop version 2. Appropriate gRNAs were then purchased from IDT and resuspended in nuclease-free water at a concentration of 100 µM. The gRNA/Cas9 ribonucleoprotein (RNP) complex was prepared by mixing the gRNA and Cas9 protein at a ratio of 1.5:1 and incubating at 37 °C for 10 minutes. The purified genomic DNA was then mixed with the gRNA/Cas9 RNP complex and incubated at 37 °C for 1 hour. This allowed the RNP complex to bind and cleave the Region of Interest (ROI) in the genomic DNA. The DNA was purified using a PCR cleanup kit to remove any residual gRNA/Cas9 RNP complex, unbound gRNAs, and other impurities that may interfere with the downstream sequencing. The sequencing adapters were ligated to the Cas9 cut sides of the ROI using the Ligation Sequencing Kit

(SQK-LSK109, Oxford Nanopore Technologies) according to the manufacturer's protocol. The ligation reaction was incubated at 20°C for 10 minutes, followed by 65°C for 10 minutes. Excess unligated adapters and other short DNA fragments were removed using the Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies) according to the manufacturer's protocol. The purified DNA was quantified using a Qubit fluorometer. The 1D2 sequencing kit (SQK-LSK308, Oxford Nanopore Technologies) was used for construction of sequencing libraries with 500 ng input DNA. The sequencing library was prepared according to the manufacturer's protocol and sequenced on a MinION sequencing device (Oxford Nanopore Technologies) using R9.5 flow-cells for 24–48 hours. The reads were base-called using the MinKNOW software and the resulting fastq files were analysed using a variety of bioinformatics tools, such as Minimap2, Samtools, and GATK, to identify and extract the relevant information. The Nanopore sequencing service and data analysis were provided by colleagues from the University of Nottingham Next Generation Sequencing Facility, who performed quality control and helped with data interpretation.

Chapter 3 Double knock-out of *MBNL1* and *MBNL2* in DM1 fibroblast cells

3.1 Abstract

Myotonic dystrophy type 1 (DM1) is characterized by disrupted MBNL protein function, leading to aberrant alternative splicing and molecular defects. In this study, I propose the following hypotheses: 1) understanding the intricate interplay between MBNL proteins, toxic RNAs, and RNA processing is crucial for effective DM1 treatments. 2) Generating MBNL-deficient cell lines can serve as a valuable tool for further investigations, ultimately contributing to the advancement of DM1 research and improved treatment strategies. To investigate the impact of MBNL proteins on disease pathogenesis, I established the following aims:

1. Simultaneously knock out MBNL1 and MBNL2 using CRISPR/Cas9 in DM1 fibroblast cells;
2. Characterise the generated MBNL-deficient DM1 cells lines through western blotting and RNA-seq;
3. Determine the effect of MBNL double knock-out on RNA foci and mutant DMPK mRNA.

The Simultaneous knock-out of MBNL1 and MBNL2 disrupted endogenous expression and had significant consequences. The double knock-out exacerbated mis-splicing events and dysregulation of gene expression, shedding light on the underlying mechanisms. Notably, it also destabilised RNA foci and facilitated the degradation of mutant DMPK transcripts, potentially rescuing other splicing events. These findings provide valuable insights into DM1 pathogenesis and offer a promising avenue for the development of targeted therapeutic interventions aimed at restoring normal RNA processing and MBNL protein function.

3.2 MBNL gRNAs

The CRISPR system type II relies on the use of a short guide RNA (gRNA) and the Cas9 protein to specifically cleave DNA sequences of interest. However, the efficiency of gRNAs can vary due to factors such as differential RNA expression, secondary structures that hinder Cas9 target recognition, or degradation by endogenous RNA interference machinery. As a result, it is crucial to test new gRNAs *in vitro* before transfection into cells. Successful *in vitro* cleavage by the Cas9-gRNA complex, indicated by the generation of smaller PCR fragments (Figure 3-1), is a good predictor of *in vivo* efficiency.

To screen gRNAs for their efficacy in cleaving DNA sequences of interest, the *in vitro* assay represents a cost-effective and time-efficient method. This approach is particularly valuable since generating a stable human cell line can take several months. To this end, I incubated gRNAs predicted by the CRISPR Design web tool, targeting MBNL1 and MBNL2, with the Cas9 protein and a PCR fragment containing their target sites in a PCR thermocycler. Subsequently, the resulting PCR fragments were analysed on a gel to assess gRNA efficiency.

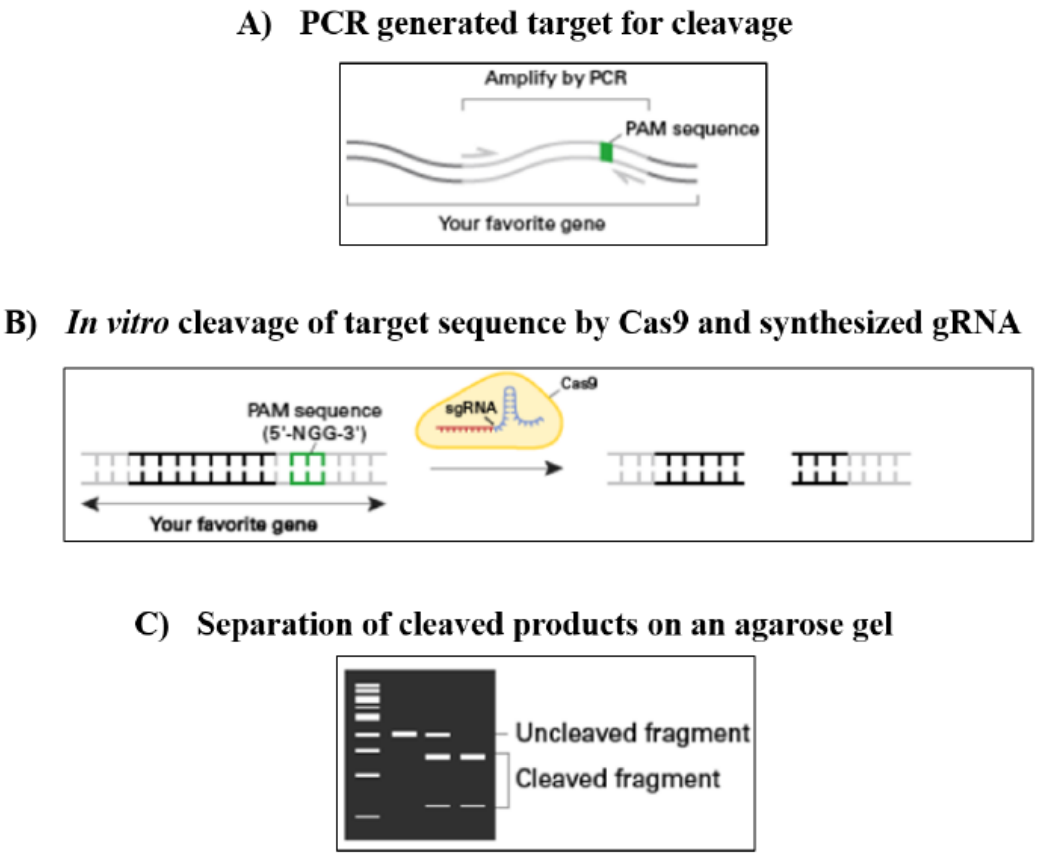
Following the *in vitro* assay, I selected two gRNAs with the highest cleaving activity (sequences shown in Table 3-1) for further investigation. Specifically, I conducted a double knock-out experiment targeting *MBNL1* and *MBNL2* in DM1 fibroblast cells (KBTeloMyoD/ KB) using these gRNAs. The CRISPR-Cas9 gRNA target sites were located in exon 4 (258 nt) of *MBNL1* and exon 6 (154 nt) of *MBNL2*, respectively.

Table 3-1 gRNA sequences targeting MBNL1 and MBNL2.

	Target sequence 5'-3'	Exon	Efficiency
MBNL1	CAACGTGGCAATTGCAACCGAGG	4	0.69
MBNL2	GCTGGGGTTAAAGACCGCGCTGG	6	0.71

The efficiency values were calculated by the CRISPR Design software: https://chopchop.cbu.uib.no/. The protospacer adjacent motif (PAM) sequences are indicated in red.

Figure 3-1 Overview of the CRISPR/Cas9 in vitro assay protocol.



(A) A PCR fragment harbouring the gRNA target site is synthesised from genomic DNA. (B) The PCR fragment is then combined with each potential gRNA and recombinant Cas9. (C) The reaction products are separated by agarose gel electrophoresis and fragments analysed. Cleavage by the Cas9 will result in two or three bands that can be distinguished in the agarose gel.

3.3 Transfection of cells and expansion of clones

The Cas9 protein and the gRNAs were expressed in cells following transfection into cells of plasmids encoding the Cas9 protein and the gRNAs (Figure 3-2). In order to select for potentially modified cells, Hygromycin linear marker, which is a 2.1 kb fragment from Clontech that confers resistance to Hygromycin-B, was used to select the engineered cells with modification to both *MBNL1* and *MBNL2*.
The design of gRNAs and transfection of cells were accomplished by project students from our lab.

3.4 Determining the efficiency of genome editing

The colonies that survived hygromycin selection were kept growing in T25 flasks, and once they were confluent they were subjected to conventional DNA extraction followed by PCR amplification to examine the efficiency of genome editing at expected target sites in the *MBNL1* and *MBNL2* gene. Two pairs of PCR primers were used to amplify fragments that containing the target site. The 2 pairs of PCR primers were positioned within the exon 4 of *MBNL1* gene and exon 6 of *MBNL2* gene, respectively. The unmodified target sites should produce a fragment of 572-bp from *MBNL1* genomic DNA and a fragment of 554-bp from *MBNL2* genomic DNA.

In order to characterise these fragments further, the PCR products were sequenced by Sanger sequencing. The mutations identified in the target sites of the various colonies are shown in Table 3-2.

Table 3-2 List of the colonies screened and the mutations present in each of them.

Type of mutation in MBNL1		Type of mutation in MBNL2	
P3A2	/	P3A2	2-bp deletion
P3A3	/	P3A3	/
P3A4	/	P3A4	2-bp deletion
P3B1	/	P3B1	2-bp deletion
P3B2	/	P3B2	/
P3B3	/	P3B3	2-bp deletion
P3B4	/	P3B4	/
P3B6	/	P3B6	2-bp deletion
P3C2	/	P3C2	/
P3C3	/	P3C3	/
P3C4	/	P3C4	/
P3C6	/	P3C6	/
P3D2	/	P3D2	/
P3D5	/	P3D5	/
P4B2	/	P4B2	1-bp deletion
P4B4	/	P4B4	/
P4B5	/	P4B5	2-bp deletion
P4B6	1-bp deletion	P4B6	2-bp deletion
P4C1	4-bp deletion	P4C1	4-bp deletion
P4C3	/	P4C3	/
P4D2	/	P4D2	/
P4D3	/	P4D3	2-bp deletion
P5B2	/	P5B2	/
P5C2	/	P5C2	/

Clones with a "/" symbol had no mutations in the target site in either of the alleles.

3.5 *MBNL1* and *MBNL2* Knock-out assessment by Western blotting

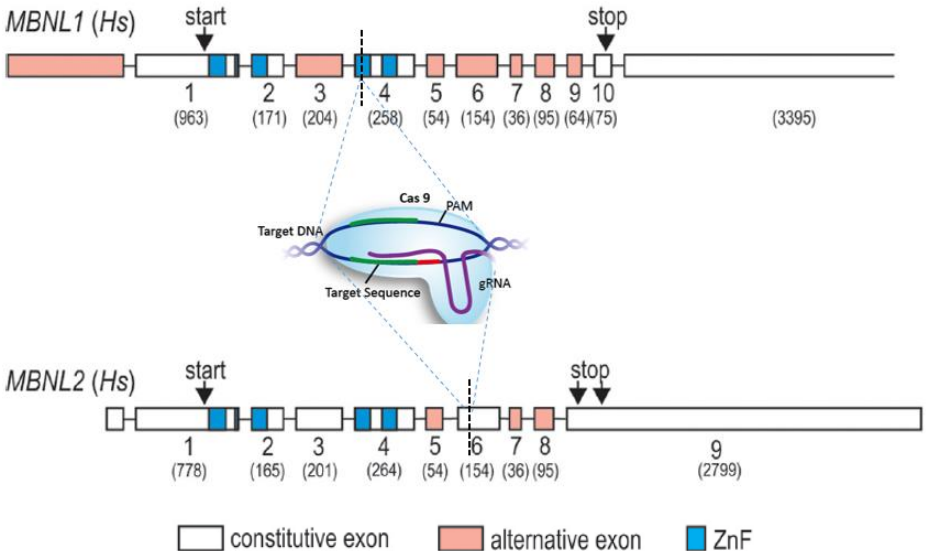
MBNL is highly expressed in wild-type DM1 fibroblast cells when two main isoforms of MBNL1 (39 kDa and 42 kDa) and one main isoform of MBNL2 (40.5 kDa) can be detected by western blots.

Following CRISPR/Cas9-mediated *MBNL* knock-out in transfected DM1 fibroblast cells (KBTeloMyoD/ KB), protein was extracted and Western blot detection of MBNL1 and MBNL2 protein performed to assess the level of the knock-out. Two *MBNL1* and *2* double knock-out cell lines, P4C1 and P4B6, were studied in detail. No endogenous MBNL1 was detected in cell line P4C1, whereas a truncated form of

MBNL1 (31.5kDa) was detected in P4B6. Additionally, truncated forms of MBNL2, measuring 31 kDa and 29.7 kDa, were detected in P4C1 and P4B6, respectively. These results suggest that CRISPR/Cas 9-mediated knock-out specifically targeting exon 4 of the MBNL1 gene and exon 6 of the MBNL2 gene led to frameshift mutations. These mutations resulted in the homozygous deletion of 1, 2, or 4 nucleotides within the both alleles of the MBNL1 or MBNL2 genes. Consequently, premature stop codons (PTCs) were introduced into their respective mRNAs. Furthermore, the frameshift mutations may have caused complete skipping of exons. Ultimately, the introduction of PTCs and/or exon skipping can disrupt the overall structure and stability of the mRNA molecules. This disruption may lead to the generation of truncated MBNL1 and MBNL2 proteins (as observed in P4B6 cells with truncated MBNL1 and truncated MBNL2, and P4C1 cells with truncated MBNL2), or degradation and destabilization of the mRNA (as observed in P4C1 cells with the complete absence of MBNL1 protein) via nonsense-mediated decay (NMD) if the PTC occurs before the last exon-exon junction of the transcript [265]. Alternatively, other faulty mRNA surveillance mechanisms may also contribute to the degradation.

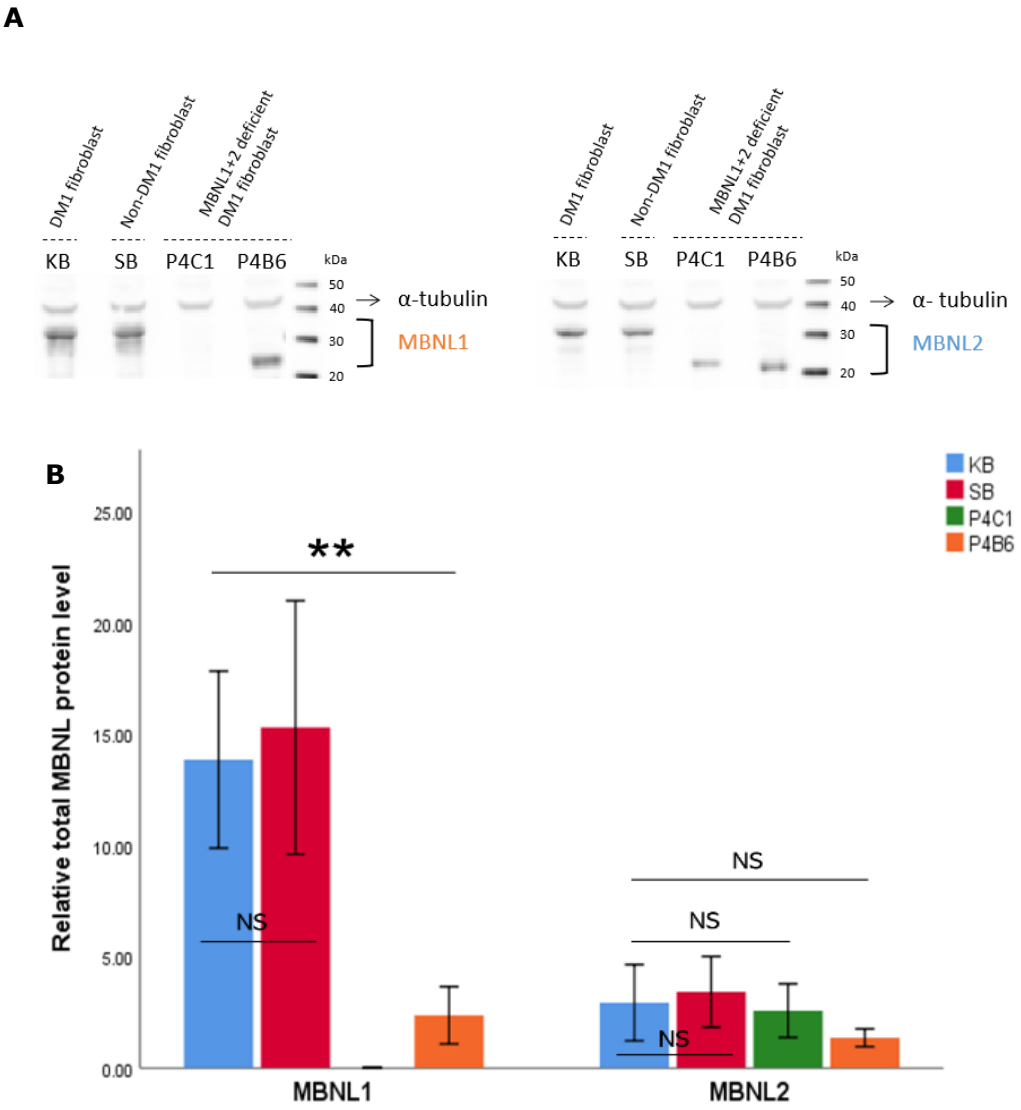
The endogenous expression of both MBNL1 and MBNL2 had therefore been effectively disrupted using the CRISPR-Cas9 system. No full length MBNL1 and MBNL2 was detectable in either P4C1 or P4B6 (Figure 3-3, Figure 0-1). Moreover, the expression of MBNL1 was completely repressed in P4C1. Therefore, these results provided me with the necessary confidence to proceed with subsequent experiments using either the P4C1 or P4B6 cell lines.

Figure 3-2 Genome Editing of the Human MBNL Locus.



The CRISPR-Cas9 target sites (shown as black dashes) are in the exon 4 (258 nt) of MBNL1 and in the exon 6 (154 nt) of MBNL2 respectively. Exon numbers refer to protein coding exons and nucleotide length of each exon is indicated in brackets. Alternatively spliced exons and ZnFs are marked red and blue respectively.

Figure 3-3 Assessment of double knock-out of MBNL1 and 2 by Western blotting

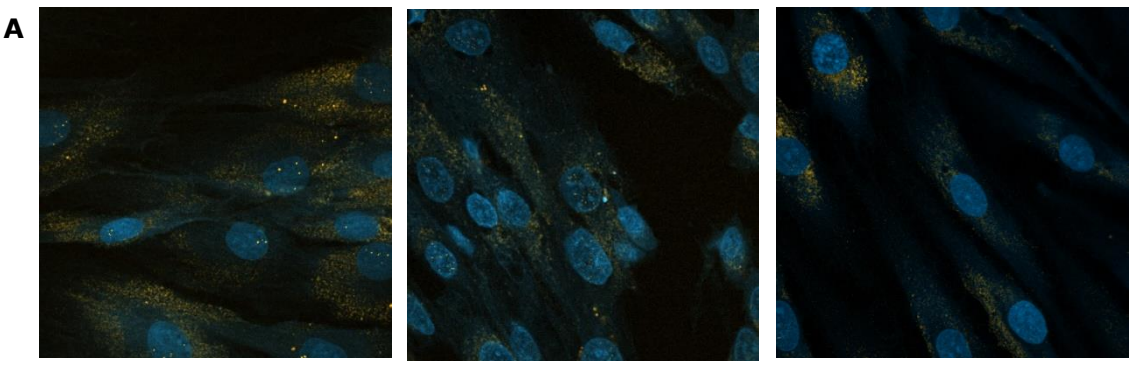


(A) Total protein extracts from KB (a DM1 patient cell line), SB (a non-DM control cell line), as well as P4C1 and P4B6 (MBNL-deficient DM1 cell lines) were analysed by Western blot to detect the expression of MBNL1 and MBNL2 proteins. α-tubulin was used as a loading control. The experiment was performed in triplicate to ensure reproducibility of the results. (B) Bar chart shows relative total MBNL protein level in these cell lines, normalised against the protein level of α-tubulin. Statistical significance was determined by two-way ANOVA (mean ± SD, n=3, ** for $P<0.01$) or indicated as not significant (NS) for $P>0.05$.

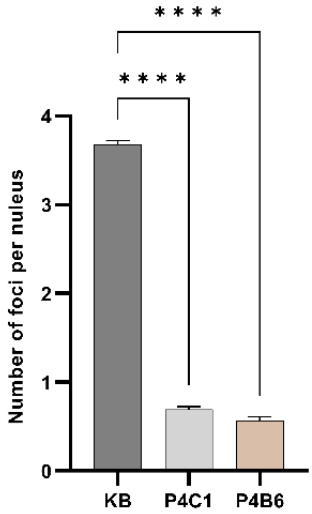
3.6 Effect of *MBNL1* and *MBNL2* double knock-out on nuclear foci

Modification of MBNL1 exon 4 and MBNL2 exon 6 appeared to significantly affect the expression level of MBNL1 protein and the integrity of MBNL2 protein in the two CRISPR-modified clones (Figure 3-3, Figure 0-1). In order to determine whether it had any effect on the production of nuclear RNA foci, the cell clones were plated on 96-well plates and RNA fluorescent in situ hybridisation (FISH) performed, using a Cy3-$(CAG)_{10}$ labelled probe. The results showed that there was a significant reduction in nuclear RNA foci in the two MBNL-deficient DM1 cells lines compared to that in the DM1 cells, the number of foci was decreased by 89.4% and 85.7% in P4C1 and P4B6, respectively (Figure 3-4). Thus, my data suggested that the absence of MBNL1 and 2 is sufficient to prevent nucleation of the predominant fraction of mutant *DMPK* RNA into foci.

Figure 3-4 Effect of MBNL1 and 2 double knock-out on RNA foci.



B



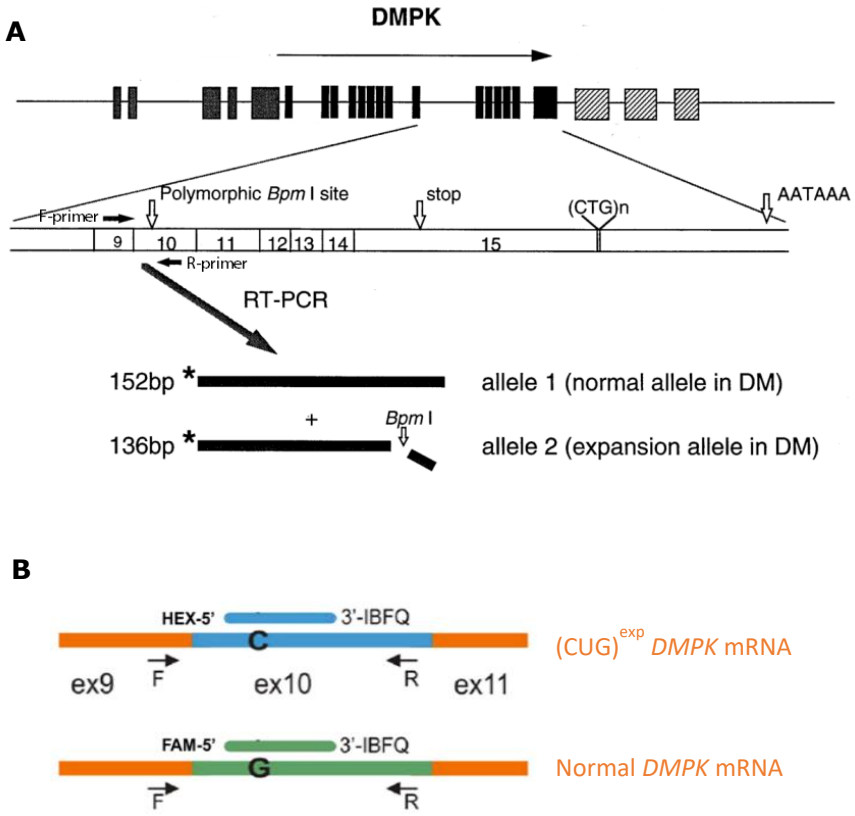
(A) Images of KB, P4C1 and P4B6 cells following RNA FISH using Cy3-$(CAG)_{10}$ probe (green) for the localisation of CUG repeats in the mutant *DMPK* mRNA and DAPI (blue) to indicate the cell nuclei. (B) Effect of *MBNL1* and *2* double knock-out on nuclear foci. Histograms show data compiled from the RNA in situ hybridisation in KB (n=7,105), P4C1 (n=4,589) and P4B6 (n=5,482) cells. n indicates the total number of cells analysed. Bars show mean ± SD. Statistical significance was determined by one-way ANOVA and **** for P<0.0001.

3.7 Effect of *MBNL* knock-out on the abundance of $(CUG)^{exp}$ *DMPK* transcripts

We have previously exploited an informative heterozygous rs527221 G>C single nucleotide polymorphism (SNP) located in the exon 10 of *DMPK.* It is important to note that the coding strand of *DMPK* is the reverse strand. In individuals who are affected by DM1 and also carry the heterozygous rs527221 SNP, the guanine (G) is commonly linked with the CTG expansion allele, whereas the cytosine (C) is associated with the normal allele (Figure 3-5 A). The prevalence of the guanine (G) allele at the rs527221 SNP is higher in the general human population, accounting for approximately 95% of individuals, while the cytosine (C) allele is less common. Assuming that the occurrence of the CTG repeat expansion is a random genetic

mutation event, the likelihood of this expansion occurring on either the guanine or cytosine allele would theoretically be equal. However, due to the higher frequency of the guanine allele in the general population, statistically speaking, more CTG expansion mutation events are expected to occur on the guanine allele. Consequently, in individuals who are affected by DM1 and also carry the heterozygous rs527221 SNP, it is frequently observed that the guanine allele is associated with the CTG expansion in the DMPK gene, while the cytosine allele is associated with the normal DMPK allele. Moreover, A BpmI recognition site is associated with this SNP (Figure 3-5 A). Here, I employed two approaches to distinguish and quantify the two types of *DMPK* transcripts: 1) conventional BmpI polymorphism assay. This assay enabled me to differentiate between the expanded and wild-type *DMPK* transcripts based on the presence or absence of a coding BpmI restriction site. Primers were designed from exon 9 and exon 10 sequences which flank the variable BpmI site (Figure 3-5 A). The DM1 cell line KB and two MBNL-deficient DM1 cell lines P4C1 and P4B6 (derived from KB) are informative for this BpmI polymorphism. 2) Nanoplate Digital PCR assay. FAM and HEX-labelled probes were designed to bind within the SNP site, they had the same nucleotide composition except for the (C or G) present in the middle of each probe (Figure 3-5 B). SBTeloMyoD (SB) was established from a donor who did not have DM but who is heterozygous for the rs527221 SNP as well. As such it provided a control for the distribution of the two types of *DMPK* transcripts.

Figure 3-5 Mechanisms of detecting DMPK transcripts.

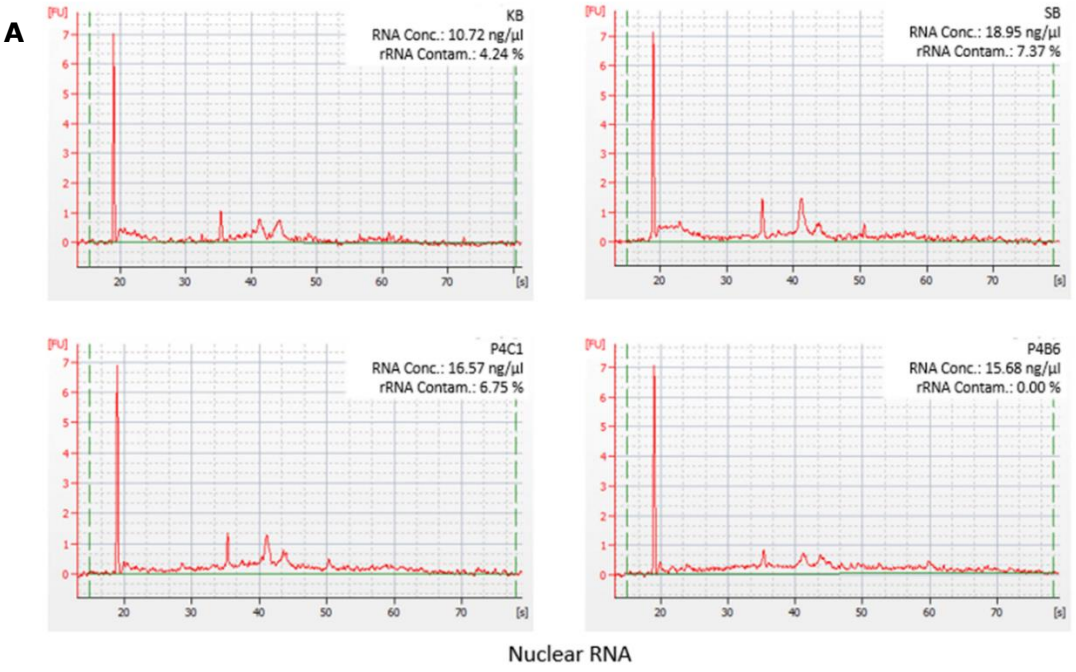


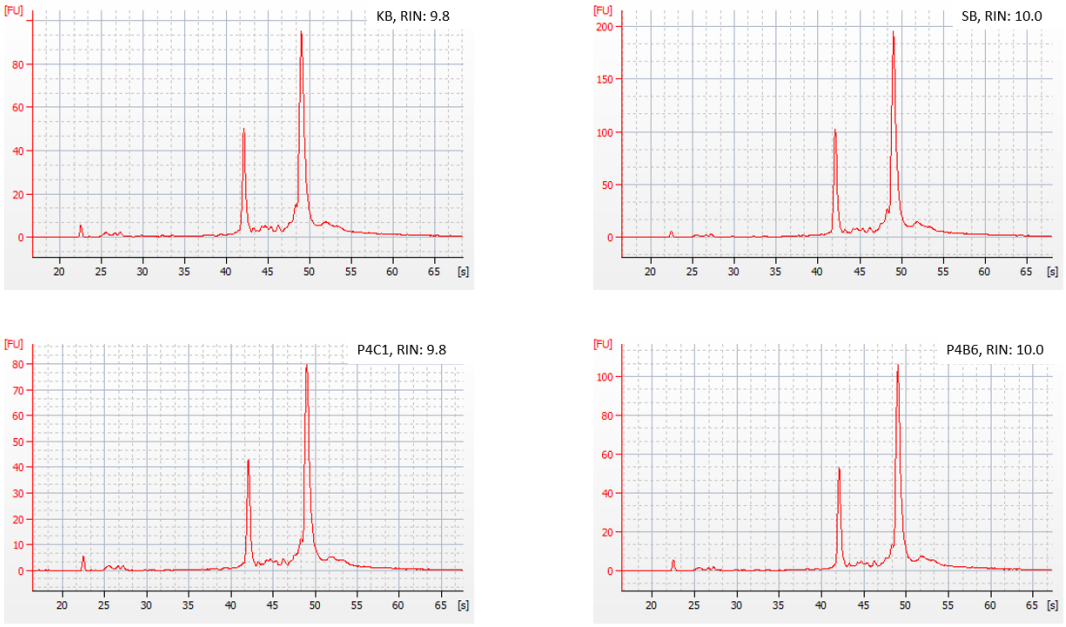
(A) BpmI polymorphism assay. The position for primers is indicated by black arrows. The $(CUG)^{exp}$ and normal *DMPK* transcripts are indicated as allele 1 and allele 2. The presence of a BpmI recognition site is only in allele 2, the expansion allele. Figure taken from [266]. (B) Scheme of probe design for digital PCR assay. FAM and HEX-labelled probes to bind within the rs527221 SNP site in the exon 10 of *DMPK*. They have the same nucleotide composition except for the single nucleotide difference (C or G) present in the middle of each probe. The probes had Iowa Black® FQ (IBFQ) as a quencher.

Nuclear and Cytoplasmic RNA fractions were isolated separately using Norgen's Cytoplasmic & Nuclear RNA Purification Kit. On-column DNase digestion was performed in nuclear fractions only to eliminate contaminating genomic DNA. The quality of the extracted RNA was examined using Agilent Bioanalyzer and 0.9% agarose DNA gel.

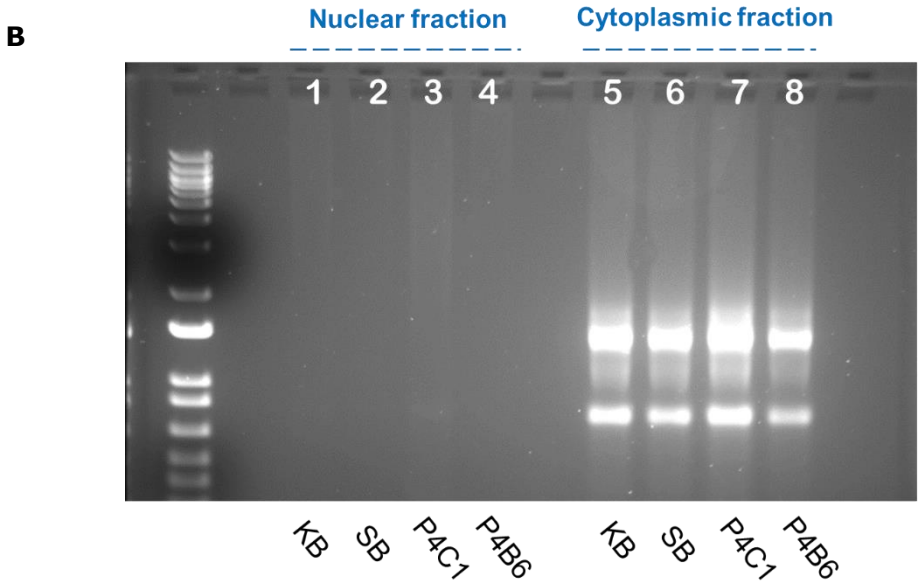
The Bioanalyzer data indicated RNA integrity scores ranging from 9.8 to 10.0 for the cytoplasmic samples (Figure 3-6 A). In the nuclear RNA samples, ribosome RNA contamination was found to be at a low level of 0.00 to 7.37% (Figure 3-6 A). Ethidium-stained agarose gel showed no trace of genomic DNA contamination in the cytoplasmic RNA fractions, and following the DNase treatment, the genomic DNA was successfully eliminated in the nuclear RNA fractions. These results provided me with the necessary confidence to proceed with subsequent experiments using these RNA samples. Following RNA quality assessment, cDNA was generated by reverse transcription.

Figure 3-6 Assessment of nuclear and cytoplasmic RNA quality.





Cytoplasmic RNA



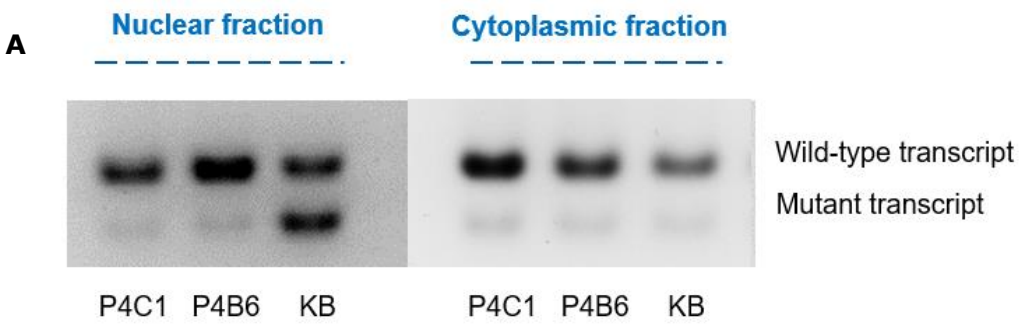
(A) Electropherogram of purified RNA from the cytoplasm and the nucleus. The x-axis shows migration time (in seconds), and the y-axis shows fluorescence intensity (in relative fluorescence units, RFU). The two prominent peaks in the electropherogram correspond to 18s and 28s ribosomal RNA (rRNA) molecules. (B) Agarose gel electrophoresis of purified RNA from cytoplasmic and nuclear fractions, stained with ethidium bromide. The gel shows clear separation of RNA bands based on size, with ribosomal RNA (rRNA) bands visible at around 28s and 18s.

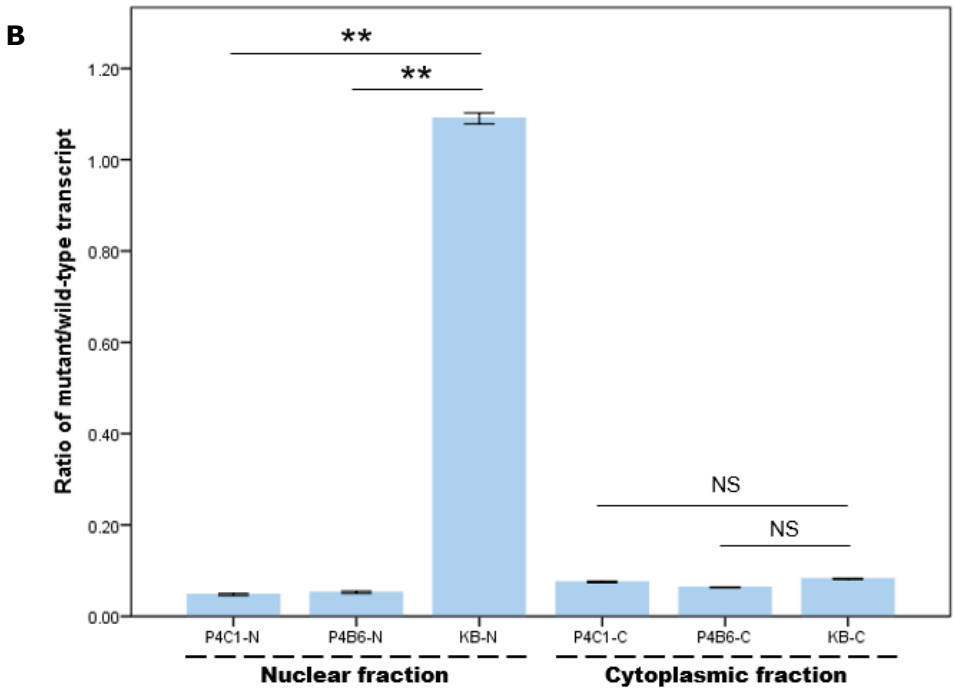
3.7.1 Method 1: BpmI restriction enzyme digestion

A 28-cycle of PCR amplification was performed using the cDNA as a template. The 154-bp amplicon generated was digested overnight with BpmI restriction enzyme resulting in two bands when visualised on a 3% ethidium-stained agarose gel, one at 152-bp and the other at 136-bp which represented the normal and mutant expanded allele respectively.

The results of BpmI digestion assays showed that in the DM1 cells (KB) the mutant *DMPK* transcript was predominantly present in the nuclear fraction with a mutant to normal *DMPK* transcript ratio of 1.15:1, while the abundance of the mutant transcript was dramatically lower in the cytoplasmic fraction, with a ratio of mutant to normal transcript of 0.05:1. Furthermore, the two MBNL-deficient DM1 cell lines (P4C1 and P4B6) showed a dramatic decrease in the abundance of nuclear mutant transcripts compared to KB, with a mutant to normal transcript ratio of 0.05:1. These cell lines also exhibited similarly low levels of cytoplasmic mutant transcripts compared to KB, with a mutant to normal transcript ratio of 0.05:1 (Figure 3-7, Figure 0-2).

Figure 3-7 Assessment of the effect of MBNL1 and 2 double knock-out on the (CUG)^{exp} DMPK transcript via BmpI digestion method.





(A) Ethidium bromide-stained gels showing RT–PCR products from nuclear and cytoplasmic RNA in KB, P4C1 and P4B6 cells. A fragment of *DMPK* mRNA was reverse-transcribed, amplified and digested with the BpmI restriction enzyme, and the resulting fragments were separated by gel electrophoresis. The experiment was performed in duplicate to ensure the reproducibility of the results. (B) Quantification of signals for the ratio of mutant to wild-type *DMPK* transcript in the nucleus and in the cytoplasm as determined by BmpI restriction enzyme digestion method. Statistical significance was determined by two-way ANOVA (mean ± SD, NS for $P>0.05$, ** for $P<0.01$) or indicated as not significant (NS) for $P>0.05$.

3.7.2 Method 2: Nanoplate Digital PCR

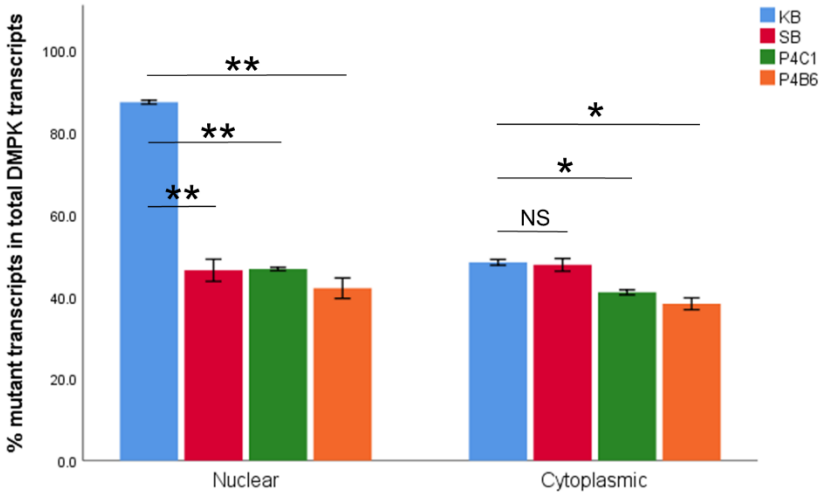
Nanoplate Digital PCR (dPCR) was performed using the cDNA as a template. To distinguish normal and mutant *DMPK* transcripts, the FAM and HEX-labelled probes were designed to bind within the SNP site. They had the same nucleotide composition except for the single nucleotide difference (C or G) present in the middle of each probe.

The dPCR set up was described in 2.7.1. Data was analysed using the QIAcuity Software Suite to determine the number of positive droplets. The counts were then used to calculate the copy numbers

of FAM-positive droplets, HEX-positive, FAM- and HEX-double positive and negative droplets in the four quadrants. PCR conditions were optimized in order to determine the best separation between negative and positive reactions.

Contrary to the results obtained from the method 1, dPCR results showed that the $(CUG)^{exp}$ *DMPK* transcript was present in both the nucleus and the cytoplasm, while a substantial portion of it was retained in the nucleus as 87.4% of the nuclear *DMPK* transcript was the $(CUG)^{exp}$ transcript (Figure 3-8). Intriguingly, the double knock-out of *MBNL1* and *2* in DM1 cells produced a dramatic reduction in the abundance of nuclear $(CUG)^{exp}$ *DMPK* transcripts with the percentage of $(CUG)^{exp}$ transcripts in total *DMPK* transcripts dropped to 42.7-47.1% and a slight but significant reduction in that in the cytoplasm, dropped from 49.5% to 38.9-40.8% (Figure 3-8).

Figure 3-8 Assessment of the effect of MBNL1 and 2 double knock-out on the $(CUG)^{exp}$ DMPK transcript via dPCR method.



Quantification of the percentage of $(CUG)^{exp}$ *DMPK* transcripts in total *DMPK* transcripts in the nucleus and in the cytoplasm as determined by dPCR with allele-specific probes. The experiment was performed in triplicate to ensure the reproducibility of the results. Statistical significance was determined by two-way ANOVA (mean ±

SD, n=3, ** for P<0.01) or indicated as not significant (NS) for P>0.05.

3.7.3 Discussion:

Results obtained from the two methods concordantly suggested that the genetic double knock-out of *MBNL1* and *2* effectively alleviate the nuclear retention of mutant *DMPK* transcripts in DM1 (Figure 3-7, Figure 3-8). With negligible amount of cytoplasmic mutant *DMPK* transcripts detected in the DM1 cells and MBNL-deficient DM1 cells via method 1, it offered no evidence from the analysis of cytoplasmic RNA to support the liberation of the mutant transcripts from nucleus to the cytoplasm following the knock-out of *MBNL1* and *2*. However, the BpmI digestion assay may underestimate the abundance of $(CUG)^{exp}$ *DMPK* transcripts from two aspects. Firstly, during the PCR annealing and extension step, there was possibility that the G-C base could mutate to a C-C mismatch or PCR products could form heteroduplexes due to cross-hybridisation. As a result of these conditions, the PCR products may be undigestible by the BpmI restriction enzyme and therefore appeared to be derived from wild-type rather than mutant transcripts. Secondly, the efficiency of BpmI digestion reactions can be limited, resulting in some target DNA fragments remaining uncleaved.

On the contrary, results obtained from method 2 suggest that in both the DM1 cells and the MBNL-deficient DM1 cells there were noticeable amount of mutant *DMPK* transcripts transported to the cytoplasm. Additionally, following the knock-out of *MBNL1* and *2*, the amount of nuclear $(CUG)^{exp}$ *DMPK* transcripts were drastically reduced, but the amount of cytoplasmic $(CUG)^{exp}$ *DMPK* transcripts were not increased or equivalently decreased, but slightly decreased, indicating that the absence of MBNL1 and 2 not only promotes the

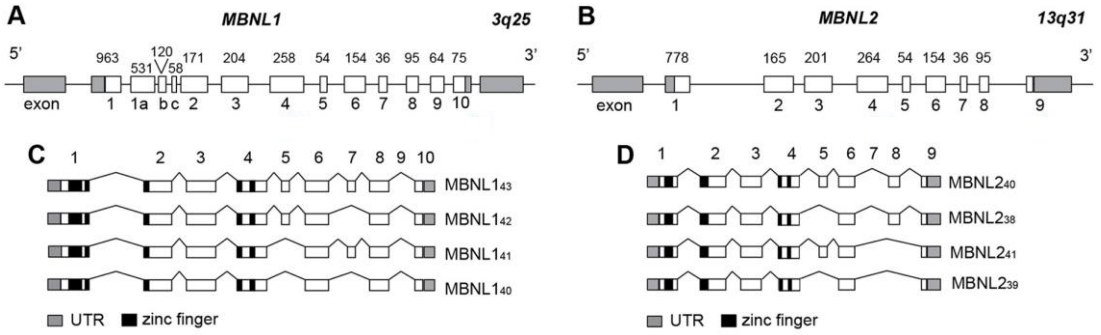
degradation of nuclear $(CUG)^{exp}$ *DMPK* transcripts but may also facilitates the export of the transcripts from the nucleus to cytoplasm.

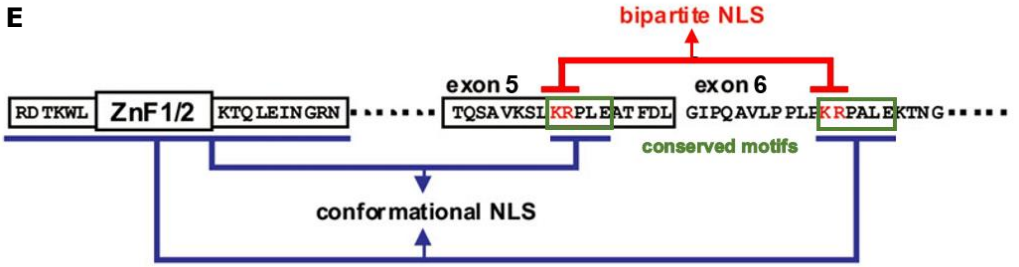
3.8 Effect of *MBNL1* and *MBNL2* double knock-out on the distribution of residual MBNL proteins in DM1 cells

CRISPR/Cas9 mediated *MBNL1* and *2* knock-out in DM1 cells produced truncated MBNL1/2. I sought to establish the composition of residual MBNLs and determine how this affected their sub-cellular distribution in DM1 cells.

MBNL1 contains two sets of nuclear localisation signal (NLS), a classical bipartite NLS (biNLS) and a novel conformational NLS (conNLS) [86]. BiNLS comprises two KR motifs embedded in conserved intervals encoded by exons 5 and 6. However, conNLS involves ZnF1-2 and the flanking regions of ZnF1-2 plus either one of the two conserved intervals encoded by exon 5 or exon 6 (Figure 3-9 E) [86]. Similarly, the mRNA that encodes MBNL2 has alternative exons corresponding to MBNL1 exon 5 and exon 6, which encodes the part of the protein associated with nuclear localisation and stronger splicing regulation (Figure 3-9 A-D).

Figure 3-9 Gene structures of MBNL1 and MBNL2.





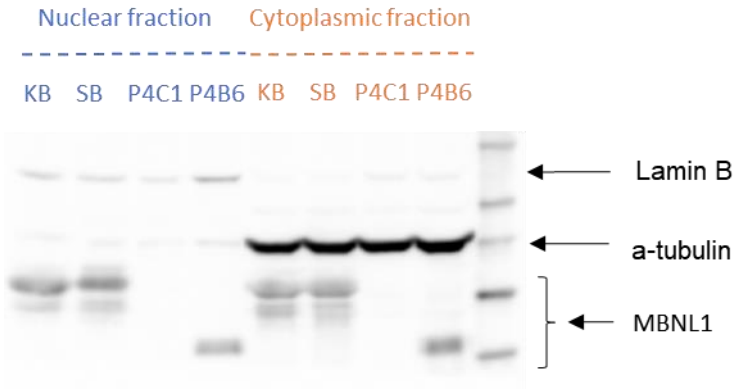
(A-D) The main alternative splice modes for MBNL1 and MBNL2 identified by previous studies and the molecular weights of the corresponding protein products. Boxes represent exons, numbered 1–10, connected by horizontal lines, the introns. Exon lengths are indicated in base pairs. (E) The mechanism of nuclear localisation of MBNL.

3.8.1 Results

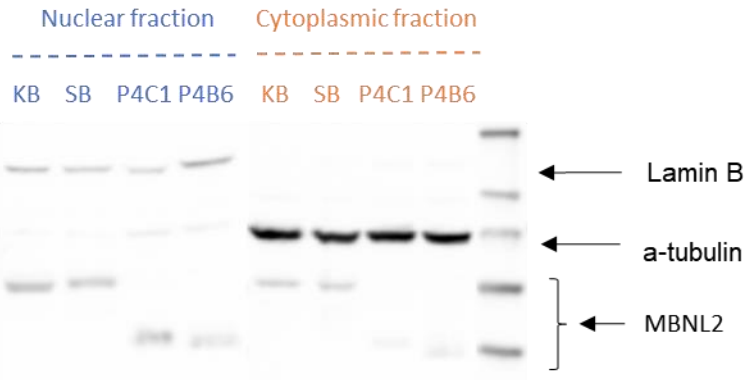
In order to examine the nuclear and cytoplasmic distribution of residual MBNL1 and MBNL2 in the CRISPR-Cas9 modified cells: P4C1 and P4B6, western blotting was performed with fractionated protein. The extent to which a protein localises to the nucleus is indicated by the nucleocytoplasmic index (NCI), ranges from 0 to 1 with higher values indicating stronger nuclear localisation in cells. In agreement with previous findings, MBNL1 was not detectable in P4C1. Interestingly, all truncated MBNL isoforms in MBNL-deficient cells displayed prominent nuclear localisation, at a level comparable to that observed in DM1 cells, except the truncated MBNL2 in P4C1 exhibited slightly but significantly stronger nuclear localisation compared to that observed in DM1 cells (Figure 3-10 A-B, Figure 0-3). Taking the mechanism of nuclear localisation of MBNL1 and MBNL2 into consideration, these results indicate that the presence of ZnF1, in addition to either one of the two conserved intervals encoded by exon 5 or exon 6, is sufficient to induce nuclear localisation of MBNL proteins (Table 3-3, Figure 3-9 E).

Figure 3-10 Sub-cellular localisation of MBNL1 and 2 proteins.

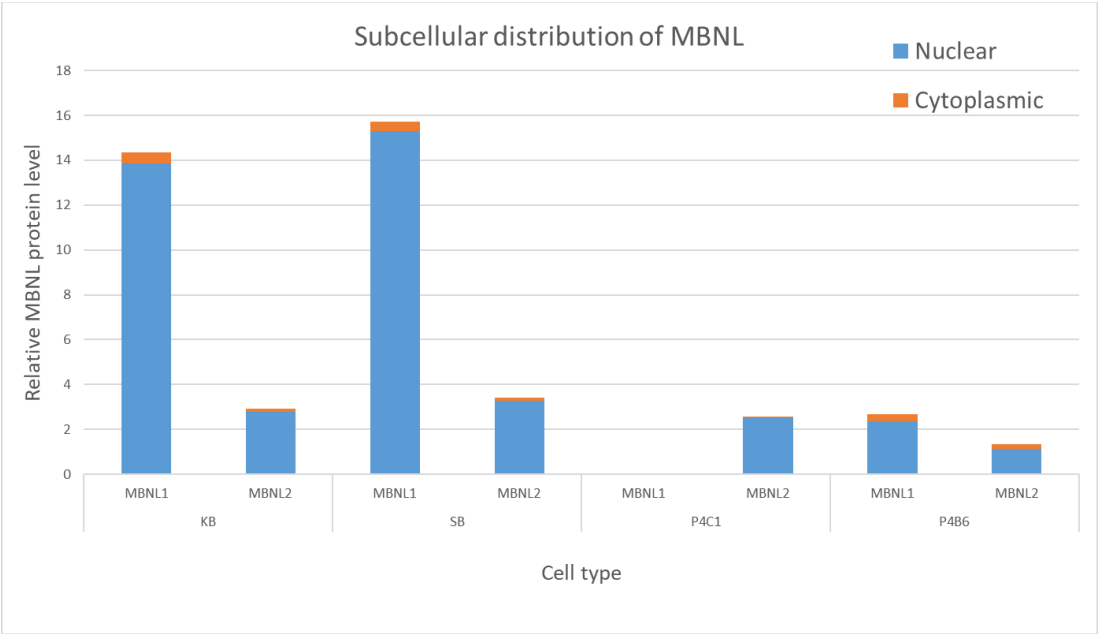
A

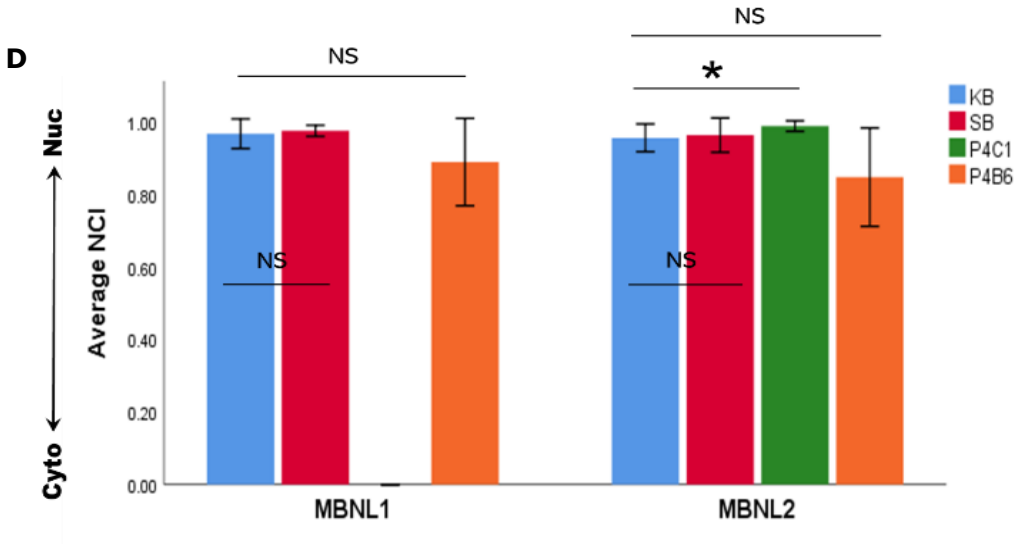


B



C





(A-B) Western blot analysis was performed to investigate the subcellular distribution of MBNL1 and MBNL2 proteins in different cell lines. Nuclear and cytoplasmic protein fractions were isolated from KB, SB, P4C1, and P4B6 cells, and protein expression levels were probed using specific antibodies. Lamin-B and α-tubulin were used as loading controls for the nuclear and cytoplasmic fractions, respectively. The experiment was performed in triplicate to ensure reproducibility of the results. (C) Quantification of signals in (A,B) for the relative MBNL protein level in nucleus and cytoplasm, normalised against the ratio of Lamin-B/α-tubulin in total protein (Figure 0-1). (D) Bar chart shows average NCI values of wild-type MBNL in KB and SB cells and residual truncated MBNL in P4C1 and P4B6 cells. The nucleocytoplasmic index (NCI) ranges from 0 to 1 with higher values indicating stronger nuclear localisation in cells. Statistical significance was determined by two-way ANOVA (mean ± SD, n=3, * for $P<0.05$) or indicated as not significant (NS) for $P>0.05$.

3.8.2 Discussion

MBNL is an RNA-binding protein that regulates its own alternative splicing through a self-regulatory mechanism. The specific targeting of exon 4 in the MBNL1 gene by CRISPR/Cas9 resulted in either complete silencing or significant reduction of MBNL1 protein expression in P4C1 and P4B6 cells. This finding aligns with previous studies by Hale, M.A., et al. (2018), which demonstrated that the

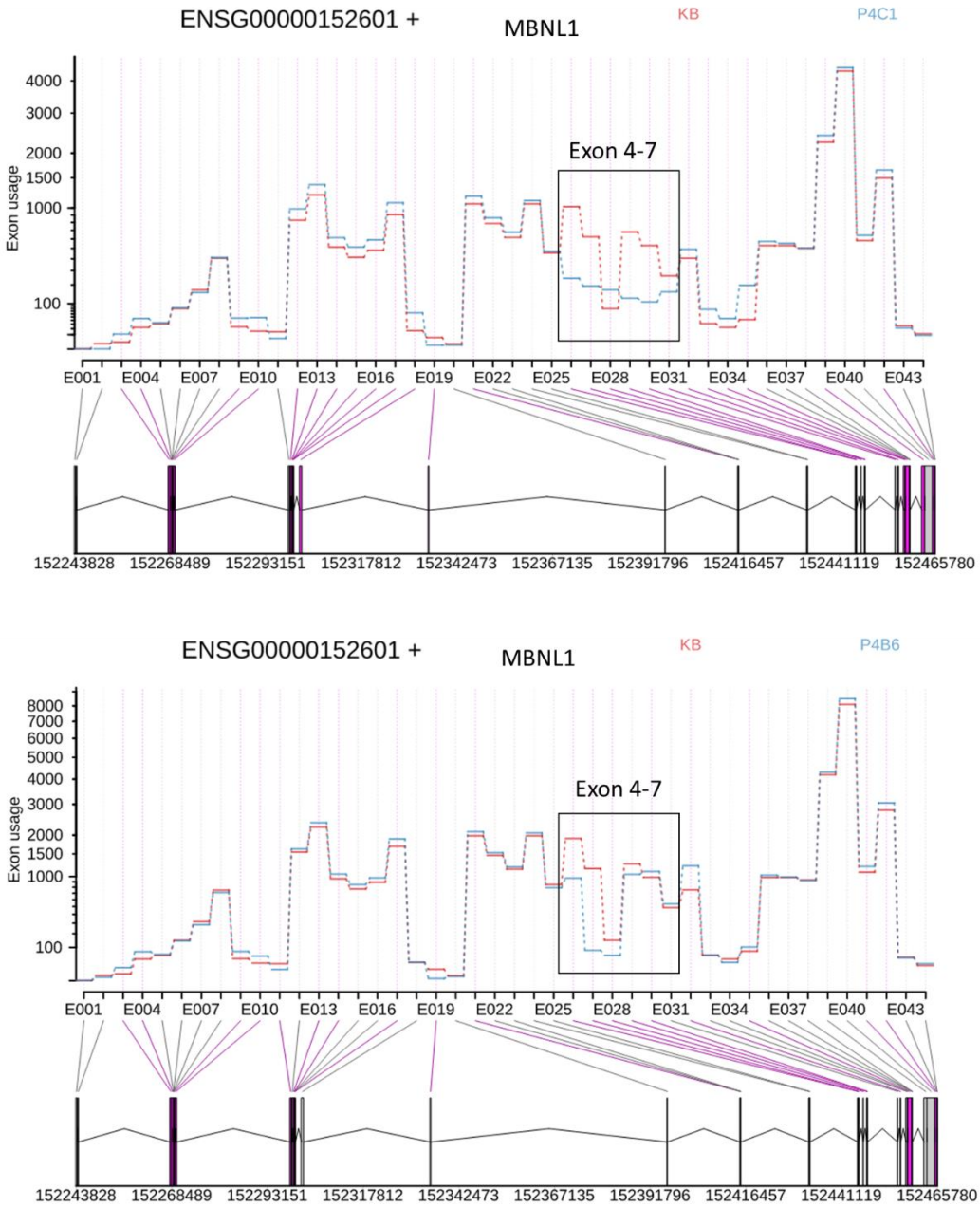
ZnF3-4 domain encoded by exon 4 enhances the stability of MBNL1 [89]. Furthermore, the ZnF3-4 domain is responsible for binding to target mRNAs and enabling MBNL1 to regulate their alternative splicing.

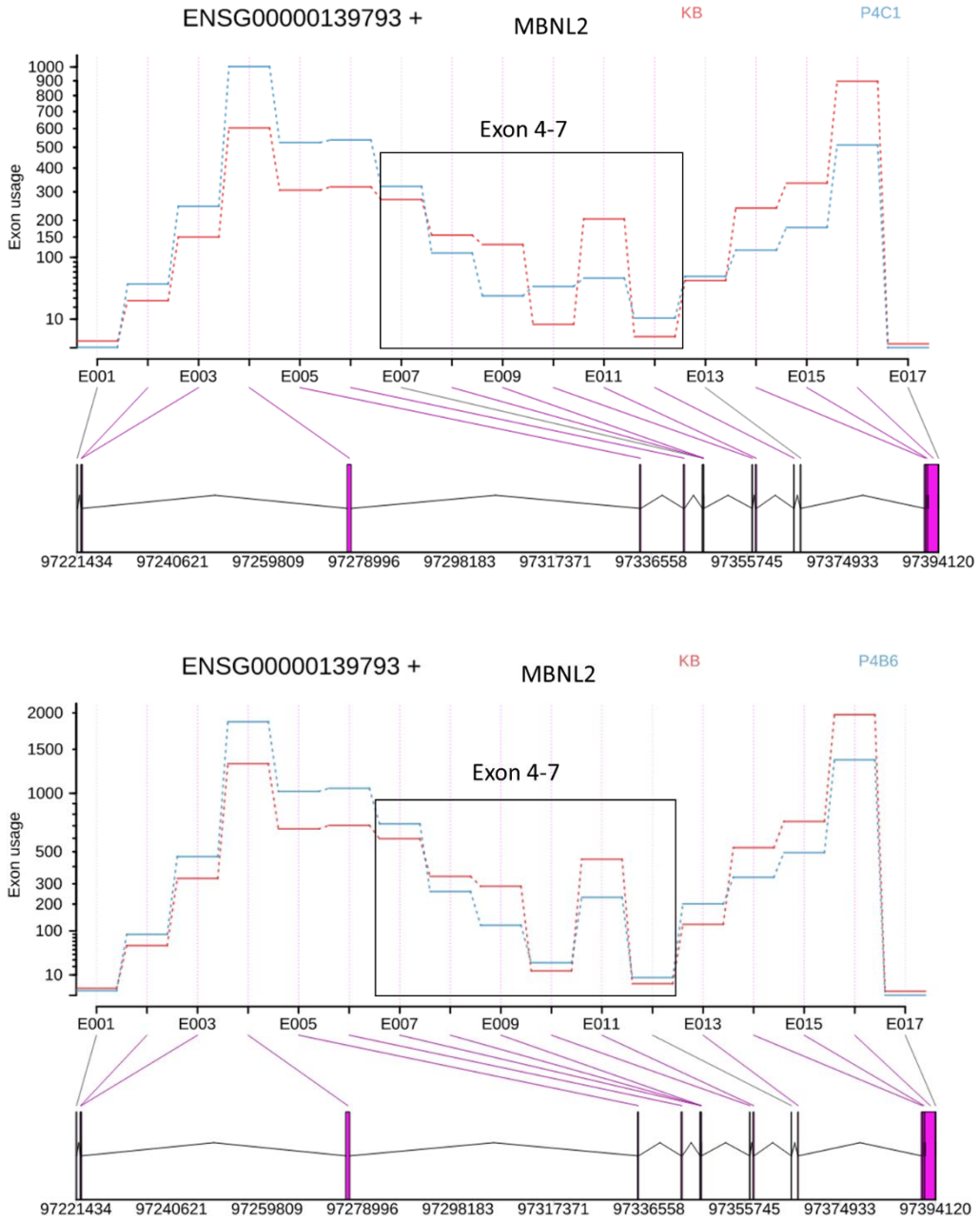
Through my RNA-Seq analysis, it was revealed that CRISPR/Cas9-mediated knock-out of *MBNL1* caused complete skipping of exon 4, potentially disrupting the normal self-regulation mechanism of MBNL1. Without the ZnF3-4 domain encoded by exon 4, MBNL1 may lose its ability to bind and regulate its own mRNA effectively. Consequently, dysregulation in the alternative splicing of *MBNL1* itself can lead to unintended mis-splicing events. In fact, I observed mis-splicing of *MBNL1* mRNA in P4C1 and P4B6 cells, where exons 5, 6, and 7 exhibited down-regulation.

These downstream exons are reliant on the proper functioning of MBNL1 and its regulatory interactions to be accurately included in the final mRNA transcript. Thus, the absence of functional MBNL1 due to mis-splicing can result in unintended skipping or exclusion of downstream exons, including exons 5, 6, and 7.

On the other hand, the expression level of MBNL2 protein with C-terminal truncation remained unaffected in P4C1 and P4B6 cells compared to wild-type MBNL2 in DM1 cells, despite specific targeting of exon 6 in the *MBNL2* gene by CRISPR/Cas9. Unlike targeting exon 4 in the *MBNL1* gene, which caused complete skipping of exon 4 and depletion of downstream exons, targeting exon 6 in the *MBNL2* gene solely led to the skipping of exon 6. However, since exon 6 is a constitutive exon which is essential for maintaining the proper structure and function of the protein, its depletion can result in altered protein structure, loss of functional domains, impaired RNA binding, and disrupted protein-protein interactions.

Figure 3-11 Effect of MBNL1 and 2 double knock-out on the MBNL1 and 2 mRNA compositions





The figures depict the results of RNA-Seq data analysis using DEXSeq, revealing significant differential exon usage of *MBNL1* exon 4-7 (A-B) and *MBNL2* exon 4-7 (C-D) between the DM1 cell line (KB, shown in red) and the MBNL-deficient DM1 cell lines (P4C1 and P4B6, shown in blue). The exon numbering and corresponding counting bins used for quantification are as follows: *MBNL1* exon 4 (counting bin E026-027), exon 5 (counting bin E028), exon 6 (counting bin E029-030), and exon 7 (counting bin E031); *MBNL2* exon 4 (counting bin E007-009), exon 5 (counting bin E010), exon 6 (counting bin E011), and exon 7 (counting bin E012). The top panel of the figure presents the fitted values of exon usage based on a linear model, after accounting for overall changes in gene expression.

The bottom panel displays the flattened gene model. The purple highlighted counting bin exons indicate regions of significant differential exon usage. To ensure reproducibility, the experiment was performed in triplicate.

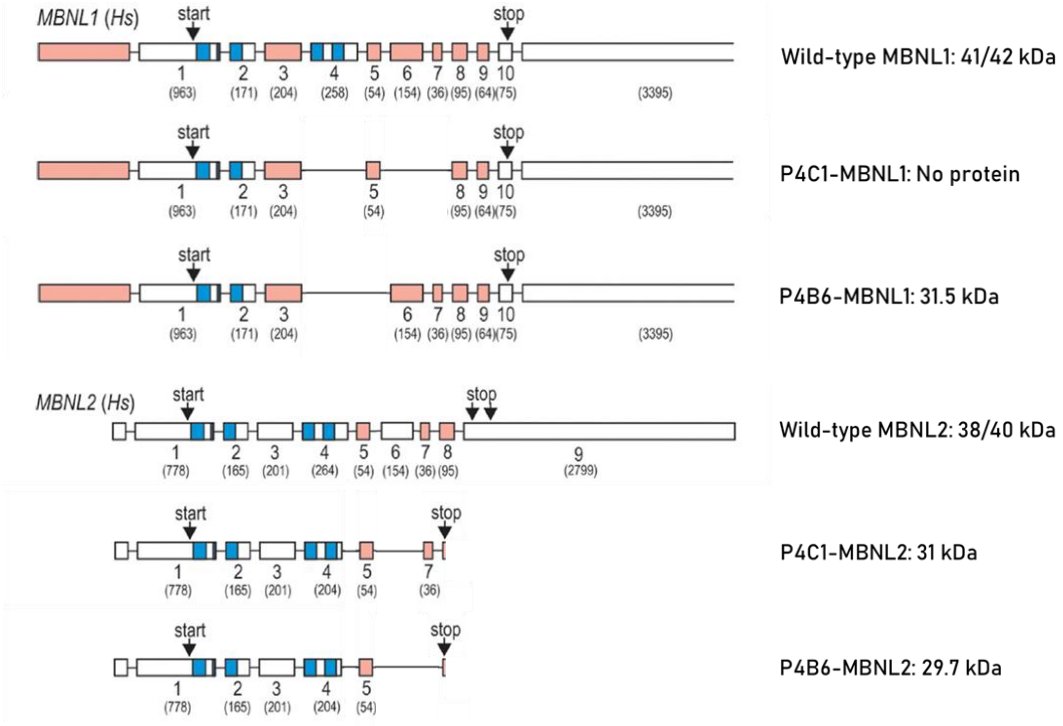
Table 3-3 Aberrantly spliced exons of MBNL1 and 2 identified in P4C1 and P4B6.

MBNL	Exon	Expression	Log_2foldchange
P4C1-MBNL1	Exon 4 (encoding ZnF3-4)	↓	-2.06 (padj=7.61E-222)
	Exon 5	↑	1.10 (padj=1.73E-14)
	Exon 6	↓	-2.44 (padj=3.07E-121)
	Exon 7	↓	-0.71 (padj=1.94E-3)
P4B6-MBNL1	Exon 4 (encoding ZnF3-4)	↓	-0.97 (padj=1.09E-185)
	Exon 5	↓	-1.34 (padj=1.07E-14)
	Exon 6	—	
	Exon 7	—	
P4C1-MBNL2	Exon 4 (encoding ZnF3-4)	—	
	Exon 5	↑	2.80 (padj=1.24E-18)
	Exon 6	↓	-1.79 (padj=2.41E-34)
	Exon 7	↑	2.87 (padj=4.52E-4)
	Exon 4 (encoding ZnF 3-4)	—	

P4B6-MBNL2	Exon 5	↑	0.81 (padj=0.047)
	Exon 6	↓	-0.95 (padj=4.49E-24)
	Exon 7	—	

“↓” indicates down-regulation, “↑” indicates up-regulation, and “—” indicates no significant change. N=3, adjusted P values (padj) of <0.05 were considered to be statistically significant.

Figure 3-12 Speculation of the MBNL mRNA compositions.



The diagrams show the speculated *MBNL1* and *2* mRNA compositions and molecular weights of the corresponding protein products in the two MBNL-deficient cell lines (P4C1 and P4B6). Alternatively spliced exons and ZnFs are marked pink and blue, respectively.

Additionally, truncated MBNL1 and 2 remained predominantly nuclear when the inclusion level of exon 5 or exon 6 was highly raised or maintain unchanged in the *MBNL* transcripts, which were consistent with previous studies that these two exons encode domains which are fundamental for MBNL nuclear localisation [86].

Furthermore, systematic analysis of the intracellular localisation of MBNL1 and MBNL2 also suggests that complete depletion of MBNL1 protein in nucleus caused a compensatory effect through an increase in nuclear MBNL2, indicating cross-regulation among MBNL proteins (Figure 3-10, P4C1).

It has been revealed that the enhanced nuclear CUG repeat RNA localisation is dependent on the composition of the zinc finger motifs (ZnFs), linker region and nuclear localisation (NLS) of MBNL [86]. The affinity of MBNL proteins in $(CUG)^{exp}$ foci is very high for all MBNL paralogs [107]. However, I observed that truncated MBNL2 with the loss of domain encoded by constitutive exon 6, was not sufficient to induce RNA foci formation suggesting that constitutive exon 6 encodes domain which may be essential for RNA binding (Figure 3-4 A, P4C1). It is noteworthy that although part of *MBNL2* exon 4 (Figure 3 11 C-D, counting bin E008-009) exhibited down-regulation, the exon usage level of the region in exon 4 encoding ZnF 3-4 remained unaffected in P4C1 or P4B6 cells. Additionally, MBNL sequestration by $(CUG)^{exp}$ *DMPK* transcripts did not stimulate MBNL import into nucleus (Figure 3-10 A-B, KB versus SB).

Overall, these findings highlight the critical role of MBNL in self-regulation and alternative splicing, emphasizing the importance of specific exons in maintaining protein function and localization. Further research is needed to fully understand the molecular mechanisms and functional consequences of these observations.

3.9 Effect of *MBNL1* and *MBNL2* double knock-out on Alternative splicing and gene expression

Total RNA was extracted from the cultured cells using the Purelink RNA Mini Kit (Life Technologies), following strict quality control measures to ensure high RNA integrity. Subsequently, poly(A)

enrichment was employed to selectively capture mRNA molecules from the total RNA pool. The isolated mRNA samples were then subjected to library preparation, which involved the conversion of RNA into complementary DNA (cDNA) fragments using reverse transcription and subsequent addition of sequencing adapters. These prepared libraries were then sequenced using the Illumina high-throughput sequencing platform. The sequencing generated millions of short reads representing fragments of the transcriptome. Raw sequencing reads obtained from the sequencer underwent quality trimming to remove adapters and low-quality bases. Next, the trimmed reads were mapped to the reference genome using the HISAT2 alignment tool. Subsequently, read counts were quantified for each annotated feature, including genes or exons, using the featureCounts program. Principal component analysis (PCA) was applied to assess the global patterns and identify potential sample outliers. To identify differentially expressed (DE) genes, the DESeq package was employed, which implements rigorous statistical methods. Moreover, to examine isoform abundance and detect differential exon usage, de novo transcriptome assembly was conducted using StringTie.

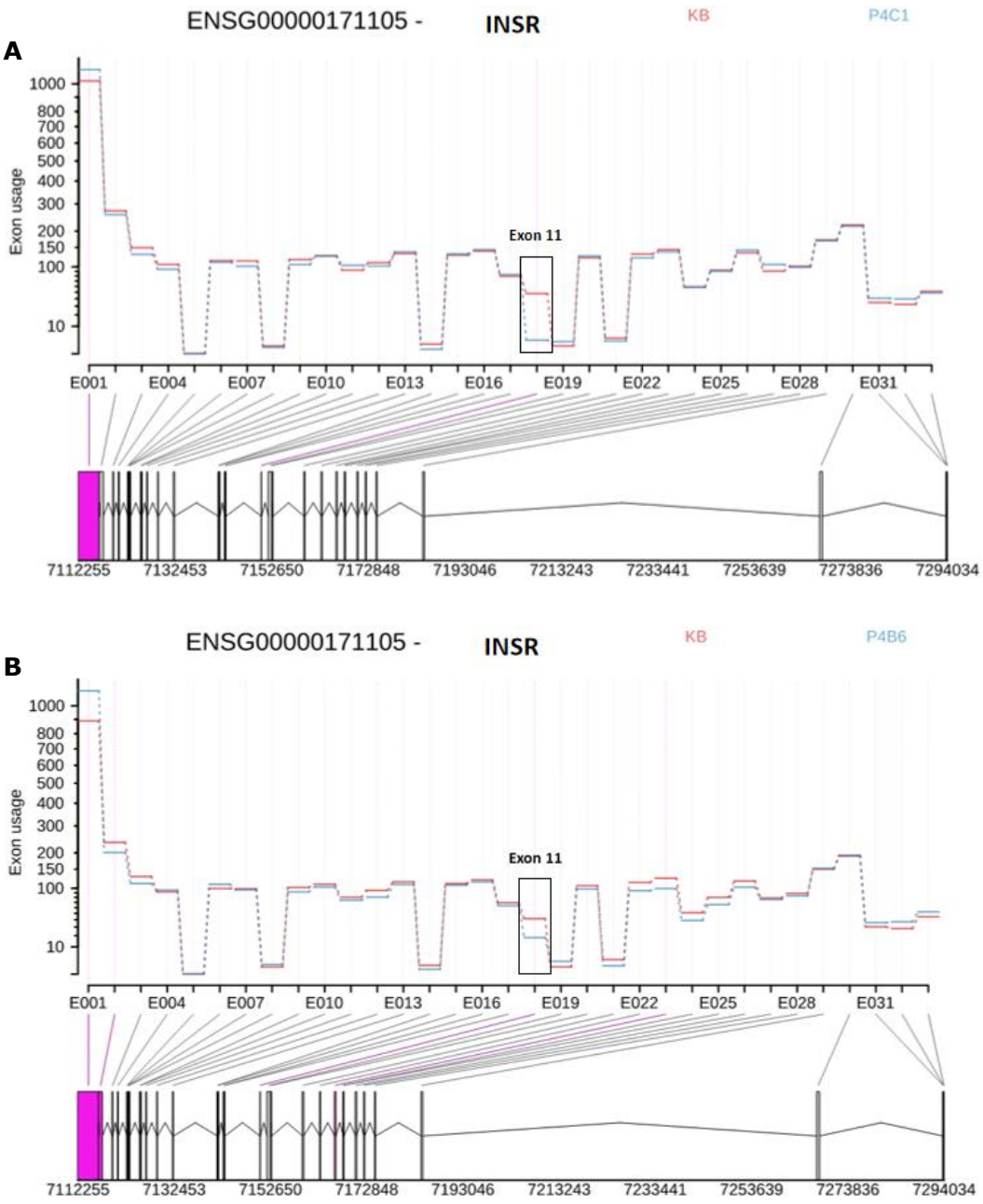
It has been shown that MBNL levels are reduced due to sequestration by expanded CUG repeats in DM cells. Loss of MBNL function has been implicated in mis-regulated splicing of pre-mRNAs like insulin receptor (INSR/IR) and nuclear factor IX (NFIX) which leads to disease symptoms like insulin resistance and regeneration deficiencies in patients with DM.

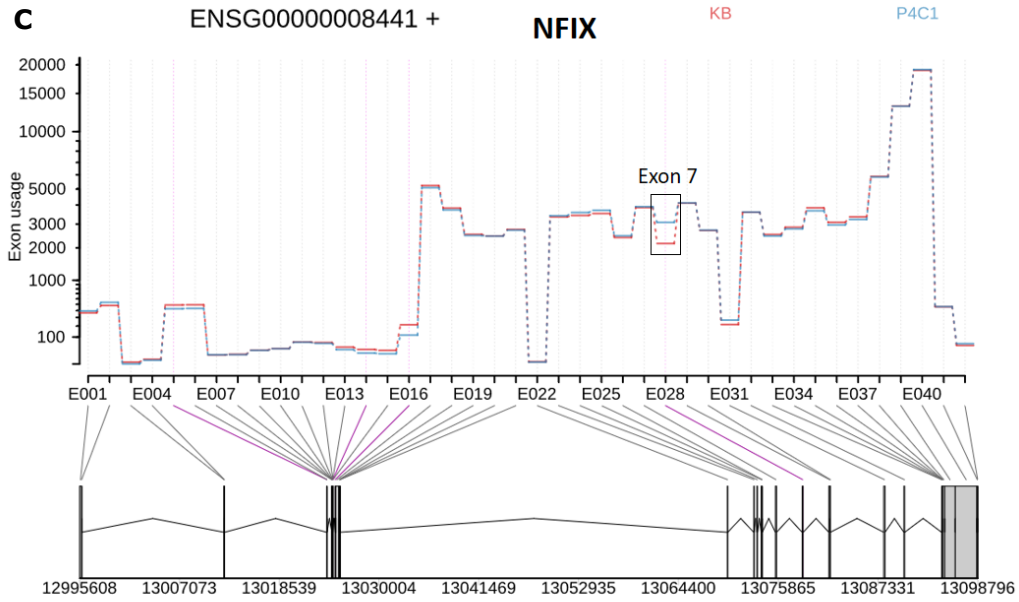
Mis-splicing of insulin receptor (INSR) gene leads to the generation of an isoform with reduced affinity for insulin due to the exclusion of exon 11, resulting in increased blood glucose levels in DM patients [63]. The splicing event is regulated by the RNA-binding proteins MBNL and CELF. Previous studies have demonstrated that MBNL

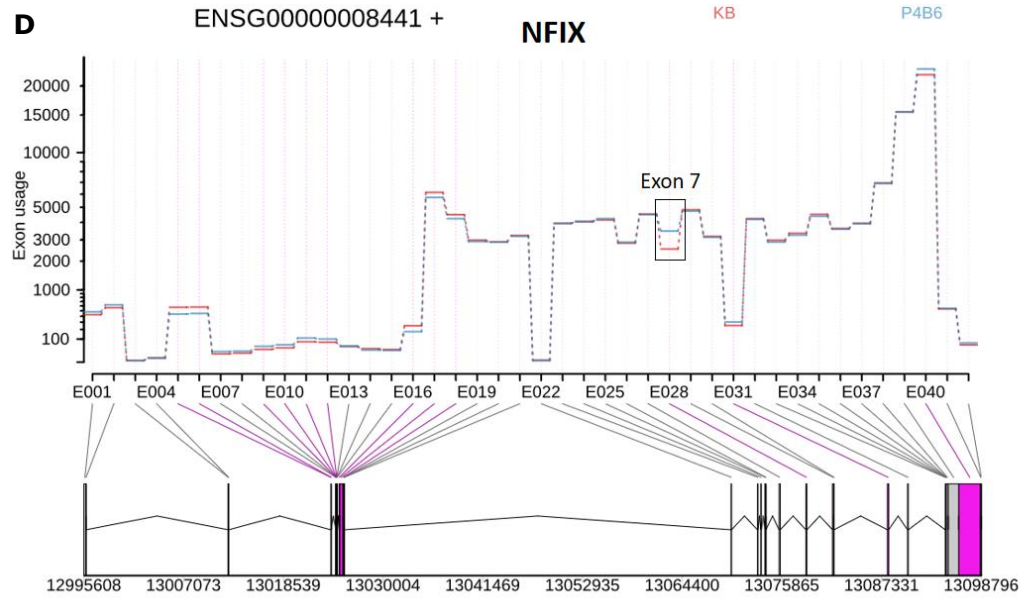
promotes the inclusion of exon 11 by binding to a conserved intronic enhancer downstream of the exon. In addition, Nuclear factor IX (NFIX), a transcription factor expressed in various tissues including skeletal muscle, is involved in the embryonic to fetal myogenic gene expression program [267]. Mis-splicing of *NFIX* exon 7 in DM patients, HSA^{LR} mice, and Mbnl1 knockout mice, leads to increased exon 7 inclusion and disrupts the association of *NFIX* with other transcriptional coactivators, resulting in the disruption of prenatal myogenic gene expression [72, 78].

To investigate the effect of MBNL1 and MBNL2 double knock-out on alternative splicing, I employed RNA-Seq assays to compare the transcripts of *INSR* and *NFIX* between DM1 cells (KB) and MBNL-deficient DM1 cells (P4C1 and P4B6). I observed a significant decrease in the inclusion of exon 11 in the *INSR* transcripts in both P4C1 (94.9% decrease) and P4B6 (56.7% decrease) (Figure 3-13 A-B), while the inclusion of exon 7 in the *NFIX* transcripts was significantly increased in both P4C1 (42.5% increase) and P4B6 (36.8% increase) compared to KB (Figure 3-13 C-D). These findings suggest that MBNL proteins play an important role in regulating the splicing of the *INSR* gene and *NFIX* gene.

Figure 3-13 Effect of MBNL1 and 2 double knock-out on the INSR and NFIX alternative splicing.







RNA-Seq data analysis using DEXSeq shows significant differential exon usage of *INSR* exon 11 (A-B, counting bin E018) and *NFIX* exon 7 (C-D, counting bin E028) between the DM1 cell line (KB, red) and the MBNL-deficient DM1 cell lines (P4C1 and P4B6, blue). The top panel shows the fitted values of exon usage according to the linear model but after subtraction of overall changes in gene expression, and the bottom panel shows the flattened gene model. The counting bin exons highlighted in purple represent the significant differential

exon usage. The experiment was performed in triplicate to ensure the reproducibility of the results.

Bridging integrator 1 (BIN1) is expressed in many tissues and functions as a regulator of actin and membrane dynamics [268]. In particular, BIN1 is essential for the formation of T-tubules in skeletal muscle, which are invaginations of the plasma membrane that help transmit electrical impulses throughout the muscle fiber, allowing for coordinated muscle contraction. The alternative splicing of *BIN1* is regulated by the MBNL family of proteins, mis-splicing of *BIN1* can disrupt the formation of the T-tubule network, leading to muscle weakness and wasting. [119, 269]. *BIN1* exon 11 (referred to as BIN-201, exon 10 in the Table 3-4), which encodes the protein's phosphoinositide-binding domain, is abnormally excluded in CDM patient-derived cells and DM1 skeletal muscle [177]. The phosphoinositide-binding domain of BIN1 is critical for its localisation to membranes and for its role in membrane remodelling [177]. To gain further insight into the role of MBNL proteins in *BIN1* splicing, I conducted RNA-Seq data analysis to compare the *BIN1* transcripts between DM1 cells (KB) and MBNL-deficient DM1 cells (P4C1 and P4B6). My analysis revealed a significant reduction in the inclusion of exon 11 in the two MBNL-deficient DM1 cell lines (Figure 3-14 B-C), suggesting that MBNL proteins play an important role in regulating the alternative splicing of *BIN1*.

BIN1 has also been implicated in the regulation of tau protein [270, 271], and down-regulation of the neuronal isoform of *BIN1* has been shown to promote tau pathology in the brain [271]. Tau pathology refers to the accumulation of abnormal tau protein, which is a microtubule-associated protein that plays a crucial role in stabilising microtubules and supporting material transport in neurons [272, 273]. In healthy cells, tau protein is highly regulated and maintained

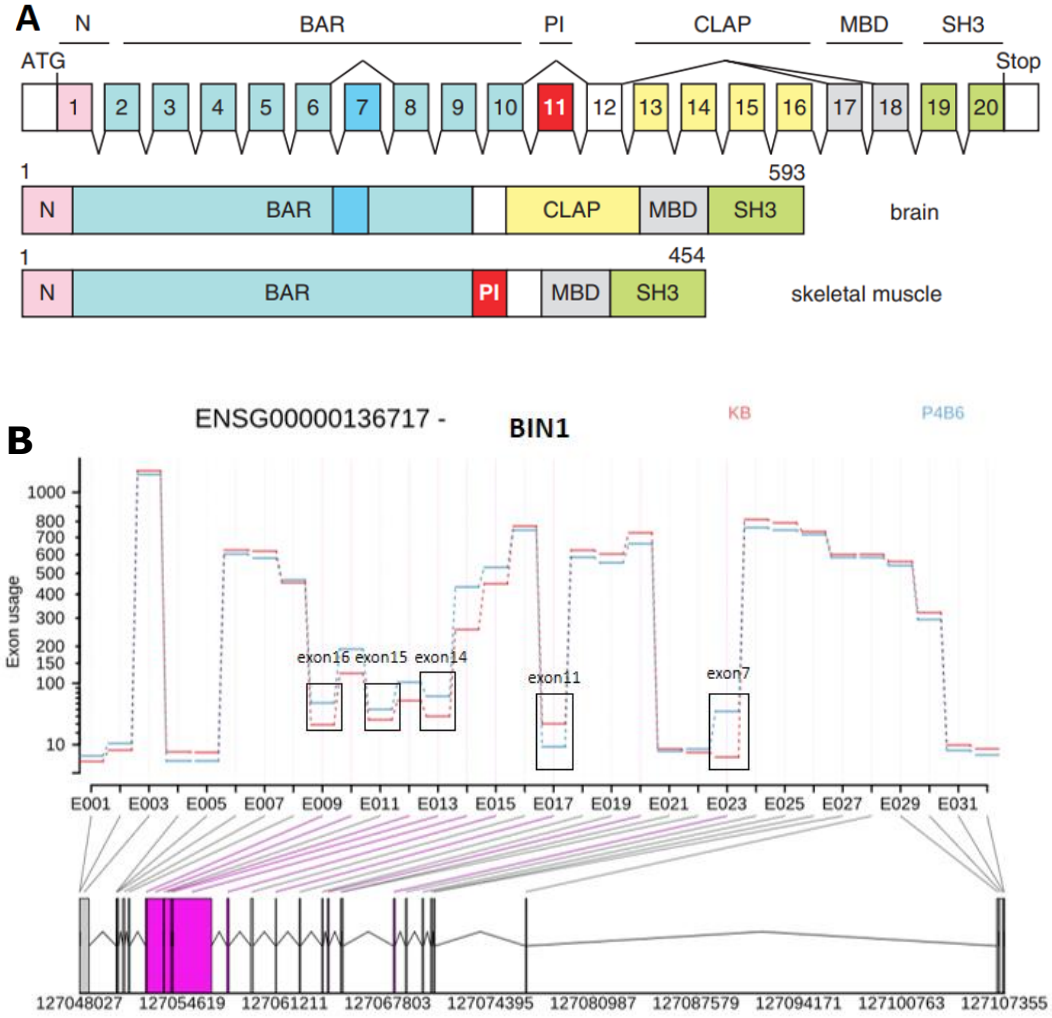
in a soluble state [274, 275]. However, in neurodegenerative diseases like Alzheimer's, Parkinson's, and frontotemporal dementia, tau protein becomes abnormally phosphorylated and aggregates into insoluble fibrils, leading to the formation of neurofibrillary tangles (NFTs) [272, 276, 277]. In DM, tau pathology is observed in several brain regions, including the hippocampus, cortex, and basal ganglia, and NFT accumulation is thought to contribute to cognitive and behavioural changes in DM patients [278, 279]. Although the *BIN1* mis-splicing has not been established in the DM brain, my analysis revealed splicing alterations associated with the expression of brain isoform of *BIN1* in DM1 fibroblast cells. Specifically, exons 7 and 14-16 (referred to as BIN-202, exons 7 and 13 to 15 in the Table 3-4) (Figure 3-14 A) which encode brain specific domains, were included at significantly higher levels in P4C1 and P4B6 fibroblast cells compared to KB cells (Figure 3-14 B-C), but there was no significant difference between KB and SB cells (Figure 3-14 D). Increased inclusion of these brain-specific exons in the *BIN1* transcript may result in elevated expression of *BIN1* with brain-specific domains, which may have a beneficial effect on DM1 brain by regulating tau pathology. However, this hypothesis is currently speculative, and further research is necessary to elucidate the precise impact of *BIN1* splicing alterations on tau pathology and cognitive function in DM1 patients.

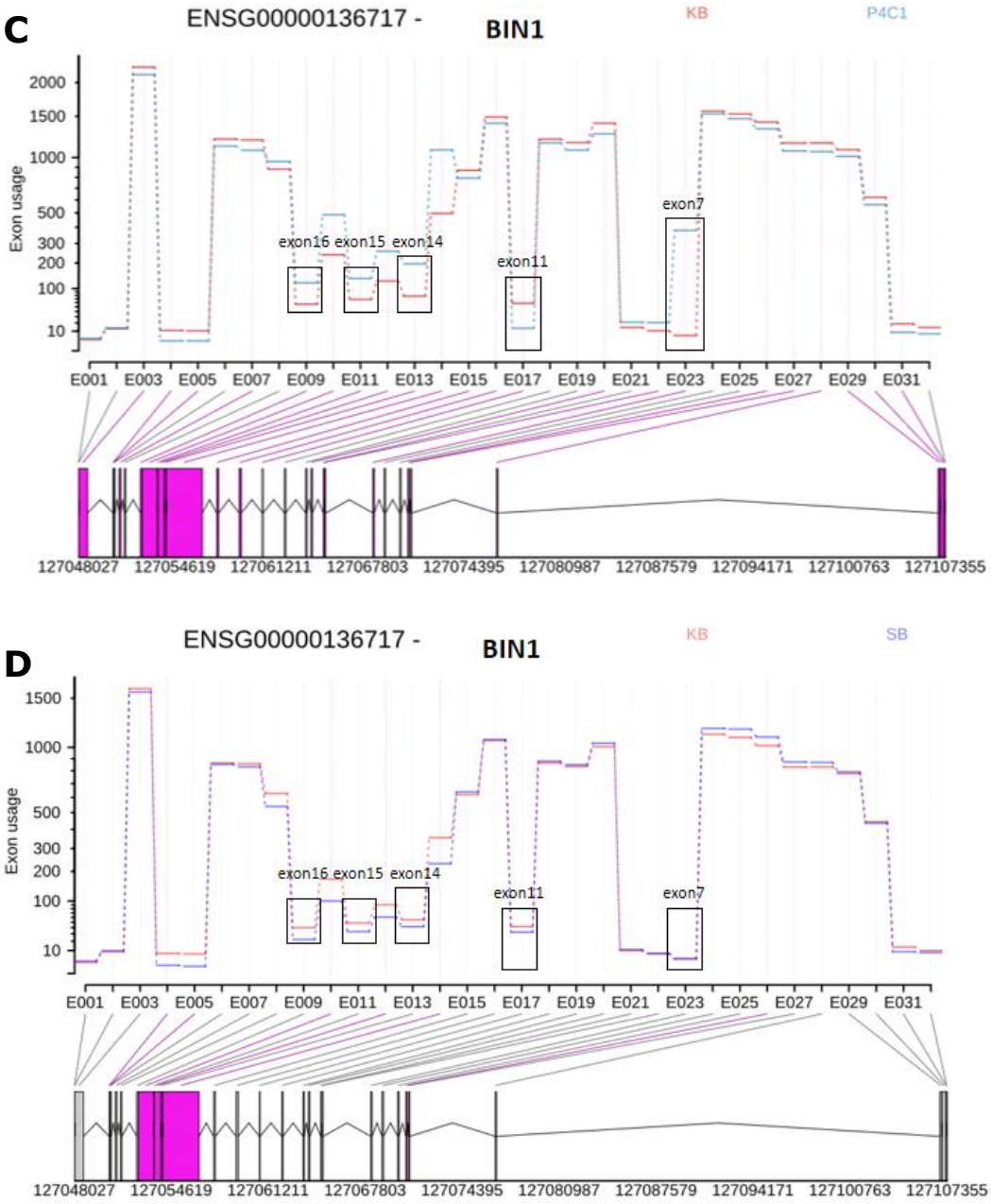
Thus, my results suggest that MBNL proteins play a crucial role in regulating alternative splicing of the *BIN1* gene, maintaining tissue-specific splicing patterns, and may have important implications for understanding the molecular mechanisms underlying various neuromuscular disorders involved in myotonic dystrophy.

The findings presented in this study align with previous research indicating that MBNL proteins are integral in regulating the alternative splicing of *INSR*, *NFIX*, and *BIN1* as sequestration of

MBNLs leads to mis-splicing of these genes in DM1 [63, 78, 119, 177]. Furthermore, the results indicate that the depletion of both MBNL1 and MBNL2 can exacerbate the mis-splicing of these genes in DM1.

Figure 3-14 Effect of MBNL1 and 2 double knock-out on the BIN1 alternative splicing.





(A) exon structure and protein domains of BIN1. BIN1 contains several functional domains, including a N-terminal amphipathic helix (N), a BIN1-amphiphysin-Rvs167 domain (BAR), a phosphoinositide-binding domain (PI), a clathrin-associated protein-binding domain (CLAP), a Myc-binding domain (MBD), and a Src homology 3 domain (SH3). Plots B, C and D show the fitted values of exon usage but after subtraction of overall changes in gene expression. The counting bin exons highlighted in purple represent the significant differential exon usage. The experiment was performed in triplicate to ensure the reproducibility of the results.

The RNA-Seq data further suggested some novel aberrant alternative splicing events in MBNL-deficient cell lines. In some cases, the mis-regulated transcripts and their corresponding proteins have been previously implicated in myotonic dystrophy type 1, but no studies have suggested a role for them in MBNLs-induced spliceopathy.

For example, knock-out of *XPA*, a DNA damage recognition and repair factor was shown to repress somatic CAG expansion in a mouse model of spinocerebellar ataxia type1 (SCA1) which is characterized by an expanded CAG repeat tract at the *SCA1* locus [280]. However, this effect is tissue-specific; loss of XPA only reduces CAG trinucleotide repeat instability in neuronal tissues among those analysed [280]. There are two main isoforms of *XPA* transcripts which are distinguished by the inclusion/exclusion of exon 6 (231nt). The XPA^{ex6+} transcript is subject to nonsense mediated decay (NMD). Interestingly, I noticed a significant increase in the inclusion level of *XPA* exon 6 in the MBNL-deficient DM1 cells (P4C1 and P4B6), compared to the DM1 cells (KB) (Table 3-4), indicating that the knock-out of *MBNL* could cause mis-splicing of *XPA* transcript and lead to XPA deficiency by triggering NMD. Subsequently, XPA deficiency may contribute to the stability of CTG trinucleotide expansion in *DMPK.*

SFPQ (splicing factor proline and glutamine rich) is one of the core paraspeckle proteins that are essential for the formation and structural integrity of paraspeckles. Paraspeckles have been implicated in modulating post-transcriptional processes in cells by sequestrating RNA-binding proteins (RBPs), mRNAs and microRNAs, and retaining them in the nucleus as well as localising them within specific subnuclear compartments [281, 282]. Furthermore, studies have demonstrated that *SFPQ* knockdown in fibroblasts cells containing the (G_4C_2) repeat expansion mutation in the *C9ORF72*

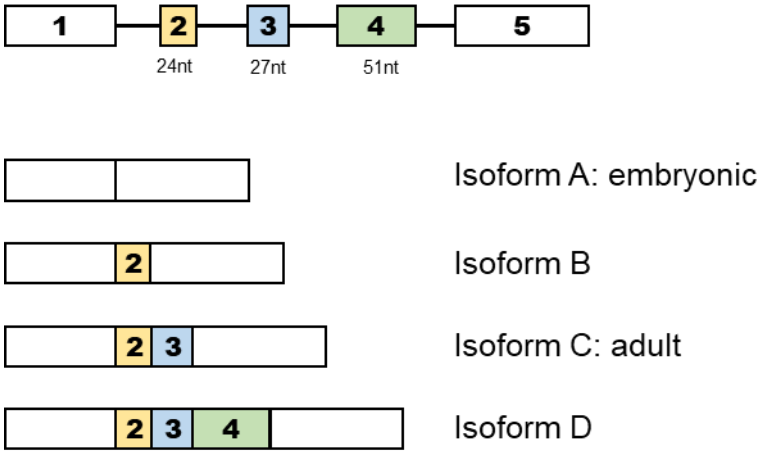
gene reduced (G_4C_2)n RNA foci [283]. This finding indicates that the essentiality of SFPQ for paraspeckle formation and the association of (G_4C_2)n repeat RNA with paraspeckles. My RNA-Seq data showed that SFPQ expression was down-regulated in MBNL-deficient DM1 cells, which might be due to the boosted expression of non-protein coding mRNA isoforms of *SFPQ* that are subject to NMD (Table 3-4). These results suggest that MBNL may play a role in the alternative splicing of *SFPQ* pre-mRNA and the regulation of SFPQ expression. It's also possible that the SFPQ down-regulation contributes to the dissipation of RNA foci in MBNL-deficient DM1 cells.

ATP11C (ATPase phospholipid transporting 11C) is a P-type ATPase that is involved in the maintenance of plasma membrane lipid asymmetry. Interestingly, I found that the regulation of *ATP11C* exon 29 alternative splicing is partially rescued in the MBNL-deficient DM1 cells (P4C1). I observed 1.15-fold higher exon 29 inclusion level in SB (non-DM, control) and a 0.78-fold higher level in P4C1, compared to KB (DM1). These findings suggest that in spite of the absence of MBNL, some aberrant alternative splicing events can be corrected due to the dissipation of RNA foci. Similarly, it has been revealed that *ATP11C* transcript exhibits exon 29 skipping in high-metastatic breast cancer cells, compared to low-metastatic breast cells [284]. Some studies have reported alterations in ATP11C expression or activity in DM1 cells or animal models, suggesting that ATP11C may be involved in DM1 pathology. However, the exact role of ATP11C in DM1 remains unclear, and further research is needed to fully elucidate its association with the disease.

Myotubularin-related 1 (*MTMR1*) alternative splicing plays a critical role in muscle differentiation, as evidenced by the regulated expression levels of specific mRNA isoforms during development. Specifically, isoform C, a muscle-specific *MTMR1* mRNA isoform, is increasingly expressed during development, while fetal *MTMR1*

mRNA isoforms A and B decrease in expression levels [285, 286] (Figure 3-15). *MTMR1* mRNA isoform C is predominant in adult skeletal muscle [286] (Figure 3-15). Previous studies reported that muscle-specific alternative splicing of *MTMR1* was impaired in both DM1 and DM2 muscle tissues as a severe reduction in the level of adult C isoform was found in muscle biopsies from juvenile/adult onset DM1 and DM2 [285, 286]. Remarkably, My RNA-Seq data showed that there was a further reduction in the level of *MTMR1* mRNA isoform C following the double knock-out of *MBNL1* and *2* in DM1 due to the inclusion level of *MTMR1* exon 2 and 3 was significantly decreased (Table 3-4) (Figure 3-15). Similarly, an aberrant splicing pattern: an adult-to-fetal shift was also seen in Myocyte enhancer binding factor (MEF) family members which play a pivotal role in muscle development and maintenance [287]. Exon 5a inclusion isoforms of *MFF2A* and *MEF2C* mRNA which are adult isoforms are predominantly expressed in normal tissues, whereas they are strikingly down-regulated in DM and instead, fetal isoforms of *MEF2A* and *MEF2C* mRNA are predominant in DM [288]. My RNA-Seq data showed that following the double knock-out of *MBNL1* and *2*, the exon 5a inclusion level in *MEF2A* and *MEF2C* mRNA was further reduced in DM1 (Table 3-4). These results suggest that MBNL proteins are involved in the alternative splicing of *MTMR1* and MEF family transcripts. Furthermore, the absence of MBNLs in DM1 exacerbated the adult-to-fetal shift in these alternative splicing events that have been previously reported in DM1.

Figure 3-15 Schematic representation of MTMR1 alternative splicing in skeletal muscle.



White boxes represent conservative exons, coloured boxes represent alternative exons.

Table 3-4 Aberrantly spliced exons and introns identified in KB, P4B6 and P4C1 via RNA-Seq data analysis.

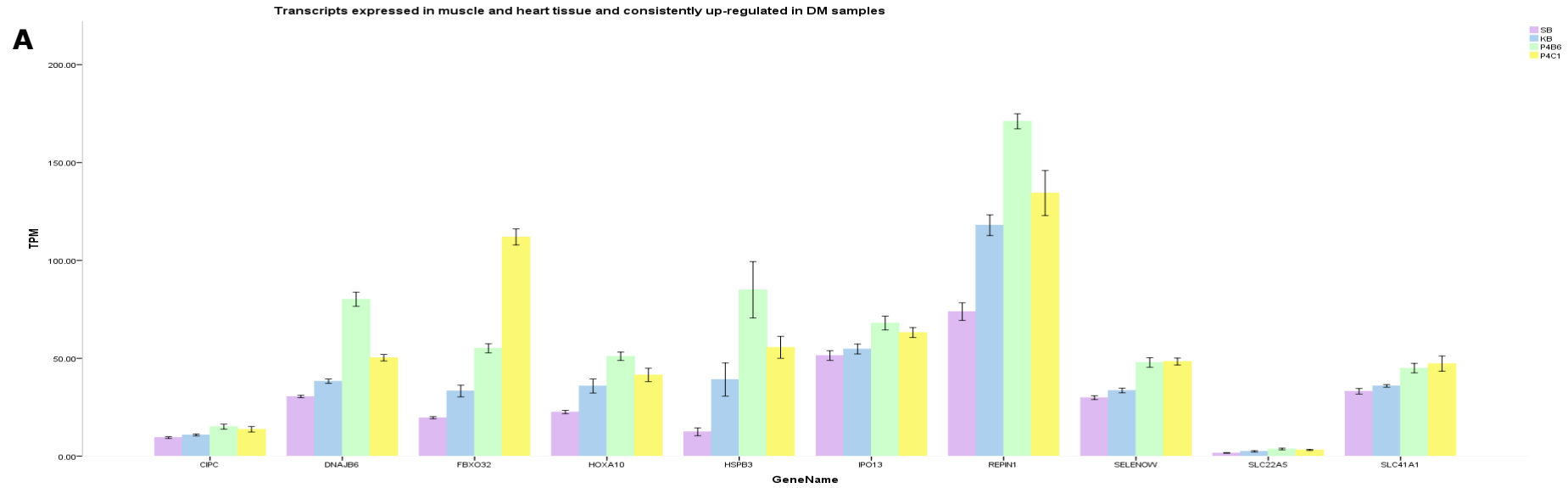
geneName	Exon/Intron	SB vs KB		P4B6 vs KB		P4C1 vs KB		GeneInfo
		log2FoldChange	Adjusted P-Value	log2FoldChange	Adjusted P-Value	log2FoldChange	Adjusted P-Value	
ANK1	ANK1-201, Exon39, 559nt	/	/	1.74	1.07E-10	1.34	2.67E-06	Ankyrin 1
IRF1-AS1	IRF1-AS1-201, lncRNA, exon4, 4548nt	/	/	1.92	1.06E-12	1.22	1.00E-03	Interferon Regulatory Factor 1-antisense RNA1
SLC30A5	SLC30A5-201, exon4, 779nt	0.42	3.60E-03	0.70	5.42E-10	0.72	1.18E-05	Solute carrier family 30 member 5
NPC1	NPC1-212, exon4, 459nt	/	/	1.41	3.85E-07	0.80	3.96E-02	NPC intracellular cholesterol transporter 1
NPC1	NPC1-209, exon2, 592nt	/	/	1.59	2.25E-12	0.73	2.62E-05	
XPA	XPA-202, NMD,exon6, 231nt	/	/	1.81	3.84E-18	1.47	1.10E-10	DNA damage recognition and repair factor
HOXC10	HOXC10-201, Exon1, 865nt	2.17	1.82E-10	5.16	4.64E-205	3.95	1.11E-94	Homeobox C4
HOXC10	HOXC10-201, Exon2, 1111nt	1.38	1.98E-10	4.47	0.00E+00	3.41	8.10E-144	
ATP11C	Exon29, 105nt	1.15	1.12E-17	0.78	2.55E-04	0.64	2.39E-03	ATPase phospholipid transporting 11C
PCNA	Exon1, UTR, 127nt	/	/	1.06	7.40E-07	0.91	5.36E-04	Proliferating cell nuclear antigen
HNRNPH1	Intron 12-13	/	/	1.37	2.74E-17	0.99	3.92E-07	Heterogeneous nuclear ribonucleoprotein H1
HNRNPH1	Intron 9-10	/	/	2.43	3.95E-32	1.51	2.19E-13	
MAP2	MAP2-202, exon7	/	/	-1.31	2.01E-09	-1.05	8.43E-05	Microtubule associated protein 2
DDX5	Intron retention across the entire transcript	1.67	1.77E-09	3.00	5.92E-40	3.57	2.49E-59	DEAD-box helicase 5
MEF2A	MEF2A-207, exon7, 24nt	/	/	1.27	3.01E-08	1.81	1.06E-18	Myocyte enhancer factor 2A
RBM3	RBM3-201, exon1, 521nt	-1.06	2.60E-20	2.09	9.50E-73	0.84	2.45E-13	RNA binding motif protein 3
SYNE1	SYNE1-207 exon14, 165nt	/	/	-1.66	2.12E-06	-1.26	4.64E-03	
SYNE1	SYNE1-207 exon13, 138nt	/	/	-1.68	8.35E-04	-1.04	4.99E-02	Spectrin repeat containing nuclear envelope protein 1
SYNE1	SYNE1-207 exon12, 108nt	/	/	-1.34	2.57E-03	-1.21	5.35E-03	
FADS2	FADS2-201, exon1, 171nt	0.87	4.13E-08	1.20	5.24E-07	1.12	1.19E-07	Flap structure-specific endonuclease 1
VEGFA	VEGFA-224, exon6, 123nt	/	/	-1.36	1.00E-02	-1.77	0.00E+00	Vascular endothelial growth factor A
VEGFA	VEGFA-224, exon7, 132nt	0.29	1.37E-02	1.04	2.43E-12	1.12	6.83E-39	
AC015813.1	AC015813.1, lncRNA	-1.26	6.46E-08	0.99	5.34E-14	0.81	1.03E-09	Novel transcript
AC060780.1	AC060780.1-202, lncRNA	-0.69	2.37E-10	1.97	1.63E-78	1.22	9.70E-48	Novel transcript
FN1	FN1-203, exon33, 270nt	/	/	0.90	0.00E+00	1.08	7.17E-284	Fibronectin 1
FN1	FN1-203, exon25, 273nt	-0.23	1.21E-48	1.23	0.00E+00	1.94	0.00E+00	
CAPZB	CAPZB-203, exon9, 113nt	0.71	3.67E-03	0.95	6.72E-04	0.94	1.25E-02	Capping actin protein of muscle Z-line subunit beta
CAPZB	CAPZB-205, exon1, 475nt	/	/	1.96	4.57E-16	1.04	1.29E-03	
LIMK2	LIMK2-201, exon2, 100nt	/	/	-1.08	2.71E-06	-1.42	4.71E-11	LIM domain kinase 2
LIMK2	LIMK2-202, exon1, 368nt	/	/	0.69	5.55E-04	1.04	8.92E-11	
ZNF37A	ZNF37A-203, proceed transcript, exon3, 499nt	-1.12	1.60E-02	-1.43	3.65E-07	-1.37	1.24E-07	Zinc finger protein 37A
PVR	PVR-205, exon6, 159nt	/	/	0.85	1.27E-02	1.52	2.84E-17	PVR cell adhesion molecule
CEACAM19	CEACAM19-207, exon2, 364nt	-1.57	5.07E-08	-1.05	3.07E-04	-1.99	4.15E-13	CEA cell adhesion molecule 19
CEACAM19	CEACAM19-207, exon3, 133nt	-1.40	3.77E-07	-1.18	7.95E-05	-2.47	3.66E-17	
CIC	CIC-203, exon2, 2804nt	0.22	2.56E-17	0.70	5.56E-221	0.64	2.99E-103	Capicua transcriptional repressor
CIC	CIC-201, exon1, 752nt	/	/	-1.23	2.02E-29	-1.15	2.77E-22	

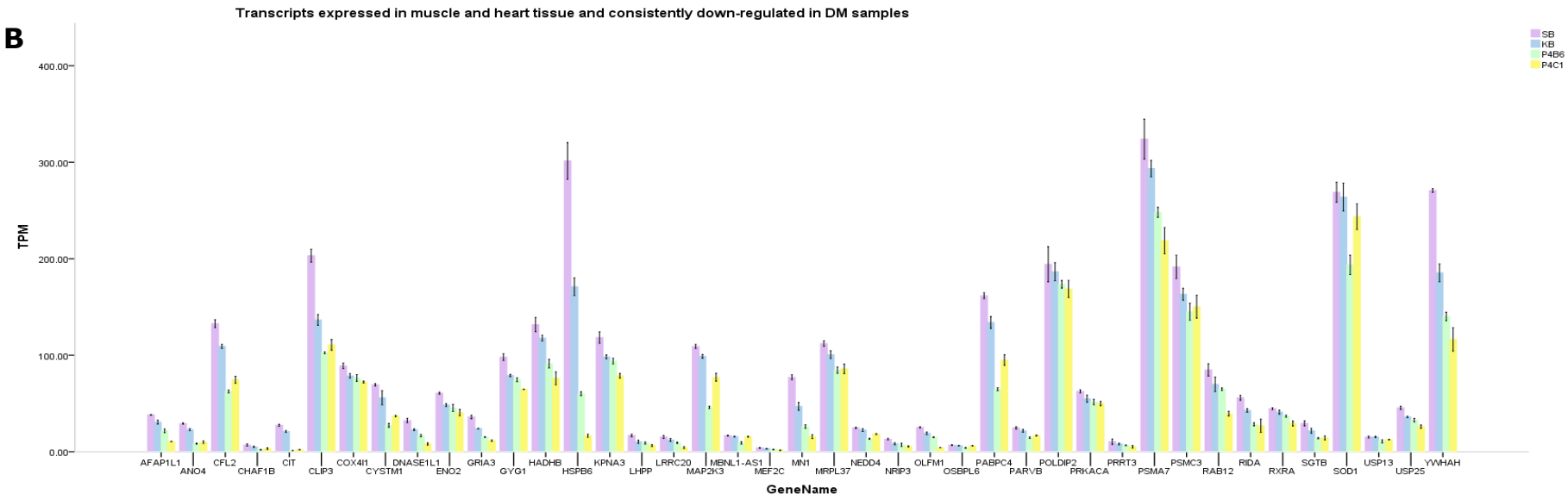
TCF4	TCF4-277, exon1, 482nt	1.06	5.00E-07	0.82	3.16E-04	1.19	4.10E-09	Transcription factor 4
DDX3X	DDX3X-259, exon4, 179nt	/	/	0.75	6.74E-03	0.73	1.87E-02	DEAD-box helicase 3 X-linked
TPM2	TPM2-201, exon6, 76nt	-1.70	3.57E-57	1.37	2.38E-68	1.74	2.96E-110	Tropomyosin 2
MYO18A	MYO18A-202, exon11, 12nt	/	/	1.94	2.88E-19	0.84	1.10E-02	Myosin XVIIIA
MYO18A	MYO18A-202, exon2, 1080nt	0.41	3.55E-11	0.97	1.98E-65	0.91	1.36E-58	
MYO18A	MYO18A-202, exon1, 86nt	/	/	1.11	4.76E-06	1.18	5.90E-08	
MYO18A	MYO18A-207, exon2, 69nt	/	/	-2.33	0.00E+00	-2.57	0.00E+00	
SF1	SF-203, exon1, 483nt	0.58	6.10E-03	0.92	3.54E-14	0.70	1.62E-05	Splicing factor 1
CTSB	CTSB-259, exon5, 96nt	/	/	-1.04	1.47E-06	-1.24	3.63E-08	Cathepsin B
CTSB	CTSB-248, UTR, exon2, 97nt and exon3, 88nt	1.91	6.01E-200	2.34	7.22E-143	2.02	1.92E-58	
RCAN1	RCAN1-212, exon1, 162nt	0.98	7.36E-13	1.80	3.03E-84	1.54	8.94E-42	Regulator of calcineurin 1
RCAN1	RCAN1-201, exon1, 354nt	-0.66	2.85E-09	-3.27	6.88E-206	-1.93	5.81E-54	
TPM1	TPM1-203, exon1, 219nt	-1.56	8.19E-91	-2.40	1.55E-138	-5.76	0.00E+00	Tropomyosin 1
TPM1	TPM1-201, exon6, 76nt	/	/	-1.18	2.45E-258	-2.39	0.00E+00	
TPM1	TPM1-202, exon9, 127nt and exon10, 198nt	/	/	2.09	4.36E-61	2.31	1.09E-194	
TPM1	TMP-204, exon8	/	/	1.17	6.73E-17	0.69	6.18E-06	
ACTN1	ACTN1-203, exon20, 66nt	0.14	2.20E-02	-1.22	2.95E-128	-3.08	0.00E+00	Actinin alpha 1
DAB1	DAB1-203, exon1-6 and exon8-15	2.17	3.29E-29	0.91	1.98E-03	1.71	3.21E-17	DAB adaptor protein 1
SECTM1	SECTM1-203, exon1	/	/	1.13	3.17E-02	1.58	2.40E-04	Secreted and transmembrane 1
SFPQ	SFPQ-206, NMD, exon7, 48nt	-0.48	1.17E-06	0.91	1.95E-30	0.62	4.47E-08	Splicing factor proline and glutamine rich
AL137782.1	AL137782.1-201, lncRNA, exon1-8	/	/	-1.33	9.09E-100	-1.43	3.33E-63	Novel transcript
AL137782.1	AL137782.1-201, lncRNA, exon15	-1.26	9.04E-03	-1.31	3.87E-03	-1.77	2.35E-05	
NCOR2	NCOR2-204, exon45, 225nt	-0.43	6.56E-05	2.22	8.21E-194	2.38	7.62E-149	Nuclear receptor corepressor 2
NCOR2	NCOR2-203, exon45, 87nt	/	/	0.86	1.11E-28	0.92	1.74E-48	
LY6E	LY6E-215, exon4, 164nt	/	/	1.57	4.82E-06	1.62	8.65E-04	Lymphocyte antigen 6 family member E
CAMK2G	CAMK2G-203, exon17, 114nt	/	/	1.17	4.12E-35	0.84	3.28E-14	Calcium/calmodulin dependent protein kinase II gamma
CAMK2G	CAMK2G-202, exon13, 66nt	/	/	-1.03	2.61E-09	-1.95	2.24E-40	
RTN4	RTN4-205, exon1, 363nt	-0.58	3.30E-02	1.38	2.46E-23	1.92	6.50E-57	Reticulon 4
INSR	INSR-201, exon11, 36nt	/	/	-1.20	9.03E-06	-4.30	1.45E-30	Insulin receptor
NFIX	NFIX-203, exon7, 123nt	/	/	0.45	3.03E-81	0.51	5.32E-39	Nuclear factor I X
BIN1	BIN1-202, exon13, 108nt, exon14, 24nt and exon15, 108nt	/	/	1.07	7.78E-07	1.34	9.64E-15	Bridging integrator 1
BIN1	BIN1-201, exon10 , 45nt	/	/	-1.81	1.42E-06	-2.16	1.87E-12	
BIN1	BIN1-202, exon7, 93nt	/	/	3.91	1.04E-15	5.96	8.48E-228	
SOS1	SOS1-202, exon21, 45nt	/	/	-0.85	3.46E-20	-1.69	3.44E-38	SOS Ras/Rac guanine nucleotide exchange factor 1
DTNA	DTNA-202, exon19, 169nt	/	/	-2.34	3.89E-03	-1.83	1.03E-02	Dystrobrevin alpha
MAP4K4	MAP4K4-219, exon17, 231nt	-0.76	9.04E-03	2.91	1.06E-156	3.29	1.87E-168	Mitogen-activated protein kinase kinase kinase kinase 4
LRRFIP2	LRRFIP2-201, exon15, 90nt	/	/	/	/	-1.19	1.44E-02	LRR binding FLII interacting protein 2
LRRFIP2	LRRFIP2-201, exon10, 51nt	/	/	/	/	-2.01	8.12E-03	
LRRFIP2	LRRFIP2-201, exon9, 75nt	/	/	/	/	-1.99	1.92E-05	

KIF13A	KIF13A-201, exon38, 105nt	/	/	1.56	4.63E-32	2.51	1.23E-112	Kinesin family member 13A
KIF13A	KIF13A-201, exon26, 39nt	/	/	-0.72	1.44E-08	-3.60	4.00E-187	
GFPT1	GFPT1-201, exon9, 54nt	/	/	/	/	-2.63	2.25E-02	Glutamine--fructose-6-phosphate transaminase 1
ABLIM3	ABLIM3-201, exon14, 99nt	/	/	1.93	1.63E-11	2.82	3.33E-27	Actin binding LIM protein family member 3
ABLIM3	ABLIM3-201, exon15, 48nt	/	/	-0.79	7.10E-12	-1.41	3.69E-33	
CAPN3	CAPN3-206, exon6, 144nt	/	/	/	/	-0.88	4.37E-02	CAPN3
CAPN3	CAPN3-206, exon15, 18nt	/	/	/	/	-2.11	3.46E-13	
CAPN3	CAPN3-206, exon16, 114nt	/	/	/	/	-1.87	3.48E-26	
MTMR1	MTMR1-208, exon3, 24nt	/	/	/	/	-18.33	8.90E-09	Myotubularin related protein 1
PKM	PKM-214, exon10, 167nt	0.06	4.19E-02	/	/	-1.06	7.28E-183	Pyruvate kinase M1/2
IMPDH2	IMPDH2-232, exon9, 72nt	-0.72	3.11E-04	-1.34	1.21E-09	-1.97	1.50E-21	Inosine monophosphate dehydrogenase 2

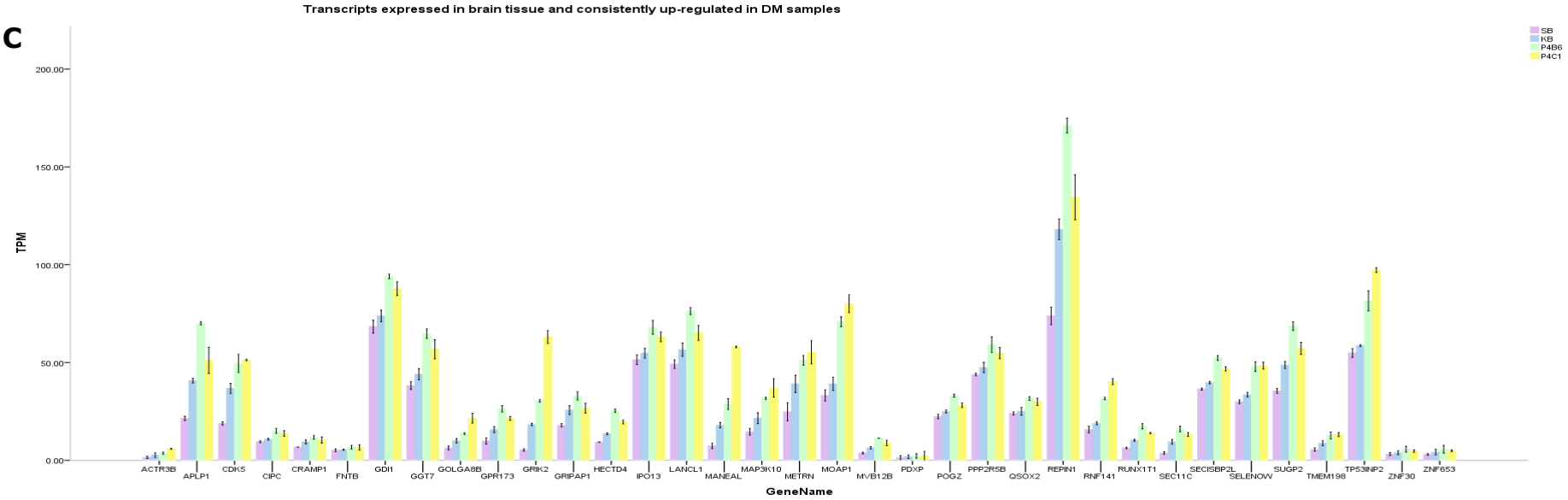
SB which is a non-patient cell line was used as a control. Log2FoldChange marked in blue or red indicated up-regulated or down-regulated alternative splicing, respectively compared with KB. Adjusted P values of <0.05 were considered to be statistically significant. "/" indicated no significant change. The experiment was performed in triplicate to ensure the reproducibility of the results.

Figure 3-16 Aberrant gene expression identified in KB, P4B6 and P4C1 via RNA-Seq data analysis.

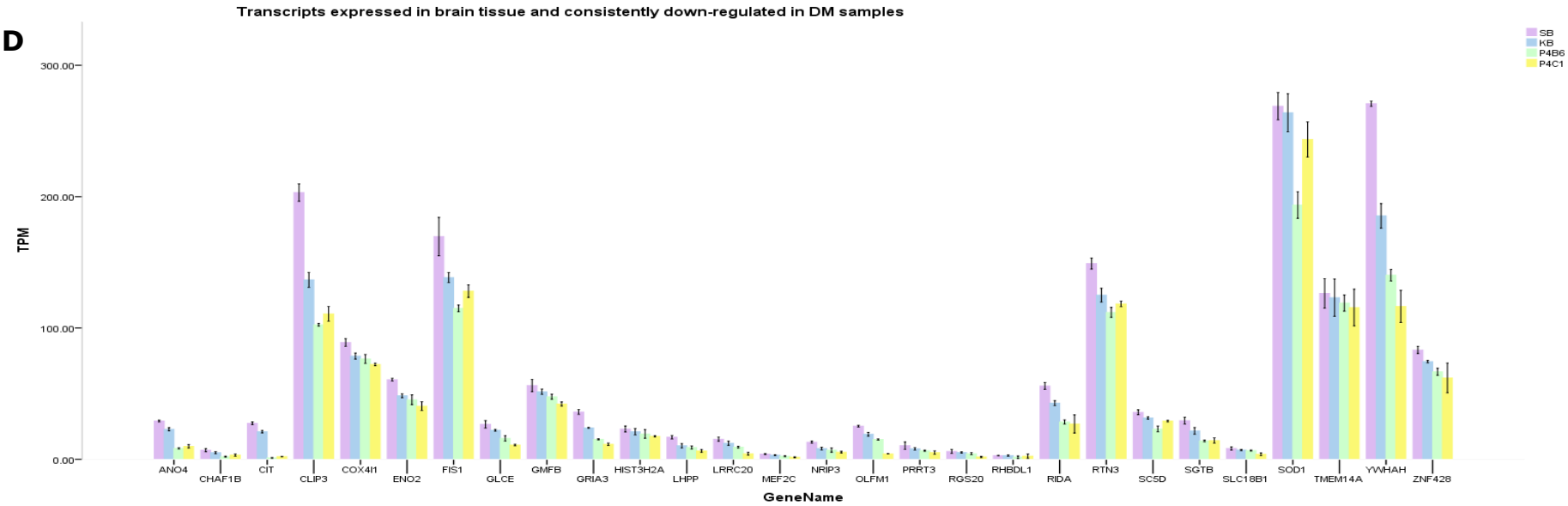


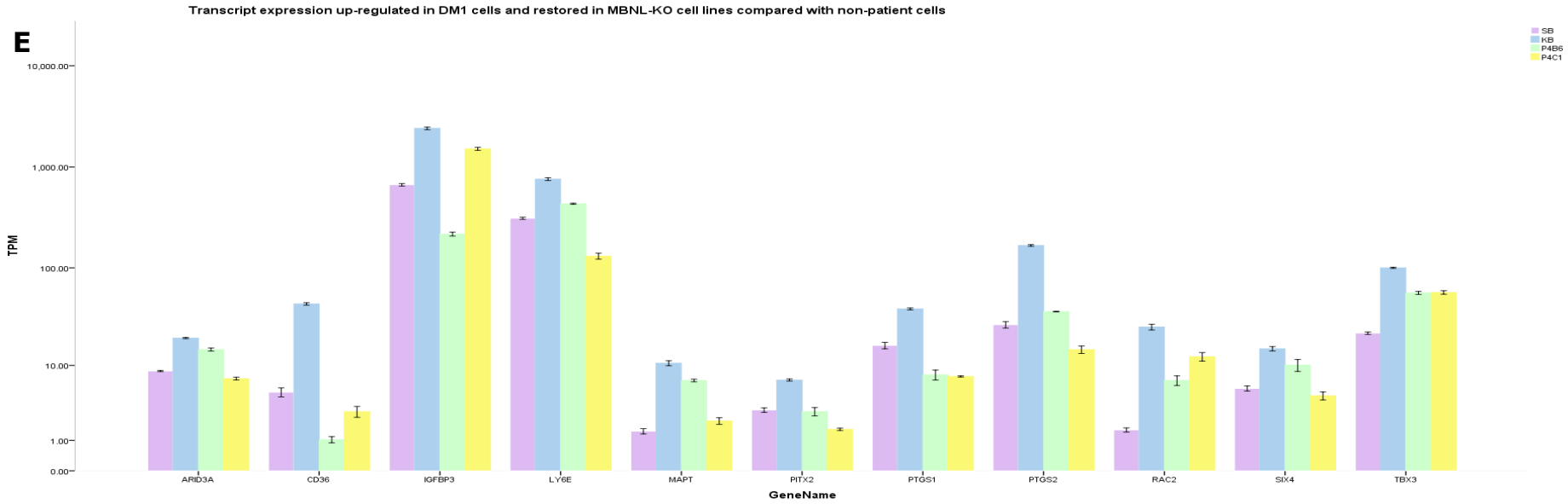

B
Transcripts expressed in muscle and heart tissue and consistently down-regulated in DM samples
SB
KB
P4B6
P4C1
TPM
400.00
300.00
200.00
100.00
0.00
AFAP1L1
ANO4
CFL2
CHAF1B
CIT
CLIP3
COX4I1
CYSTM1
DNASE1L1
ENO2
GRIA3
GYG1
HADHB
HSPB6
KPNA3
LHPP
LRRC20
MAP2K3
MBNL1-AS1
MEF2C
MN1
MRPL37
NEDD4
NRIP3
OLFM1
OSBPL6
PABPC4
PARVB
POLDIP2
PRKACA
PRRT3
PSMA7
PSMC3
RAB12
RIDA
RXRA
SGTB
SOD1
USP13
USP25
YWHAH
GeneName

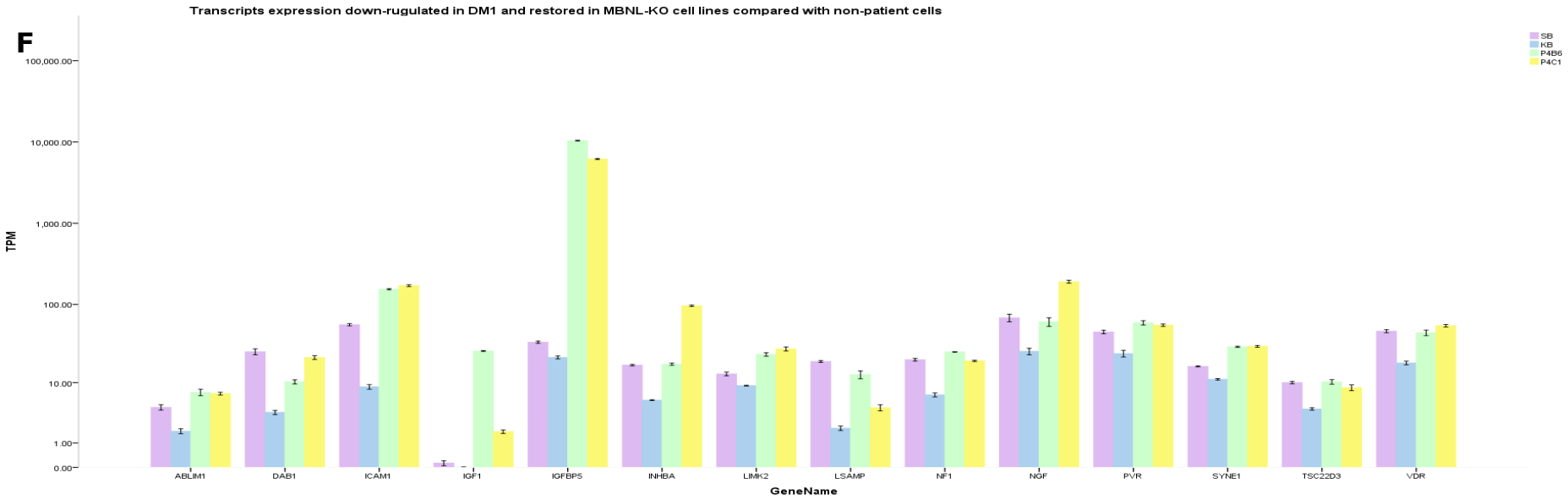
C


Transcripts expressed in brain tissue and consistently up-regulated in DM samples
SB
KB
P4B6
P4C1
TPM
200.00
150.00
100.00
50.00
0.00
ACTR3B
APLP1
CDK5
CIPC
CRAMP1
FNTB
GDI1
GGT7
GOLGA8B
GPR173
GRIK2
GRIPAP1
HECTD4
IPO13
LANCL1
MANEAL
MAP3K10
METRN
MOAP1
MVB12B
PDXP
POGZ
PPP2R5B
QSOX2
REPIN1
RNF141
RUNX1T1
SEC11C
SECISBP2L
SELENOW
SUGP2
TMEM19B
TP53INP2
ZNF30
ZNF653
GeneName

D


Transcripts expressed in brain tissue and consistently down-regulated in DM samples
SB
KB
P4B6
P4C1
TPM
300.00
200.00
100.00
0.00
ANO4
CHAF1B
CIT
CLIP3
COX4I1
ENO2
FIS1
GLCE
GMFB
GRIA3
HIST3H2A
LHPP
LRRC20
MEF2C
NRIP3
OLFM1
PRRT3
RGS20
RHBDL1
RIDA
RTN3
SC5D
SGTB
SLC18B1
SOD1
TMEM14A
YWHAH
ZNF428
GeneName


E
Transcript expression up-regulated in DM1 cells and restored in MBNL-KO cell lines compared with non-patient cells
SB
KB
P4B6
P4C1
TPM
10,000.00
1,000.00
100.00
10.00
1.00
0.00
ARID3A
CD36
IGFBP3
LY6E
MAPT
PITX2
PTGS1
PTGS2
RAC2
SIX4
TBX3
GeneName


F
Transcripts expression down-rugulated in DM1 and restored in MBNL-KO cell lines compared with non-patient cells
SB
KB
P4B6
P4C1
TPM
100,000.00
10,000.00
1,000.00
100.00
10.00
1.00
0.00
ABLIM1
DAB1
ICAM1
IGF1
IGFBP5
INHBA
LIMK2
LSAMP
NF1
NGF
PVR
SYNE1
TSC22D3
VDR
GeneName

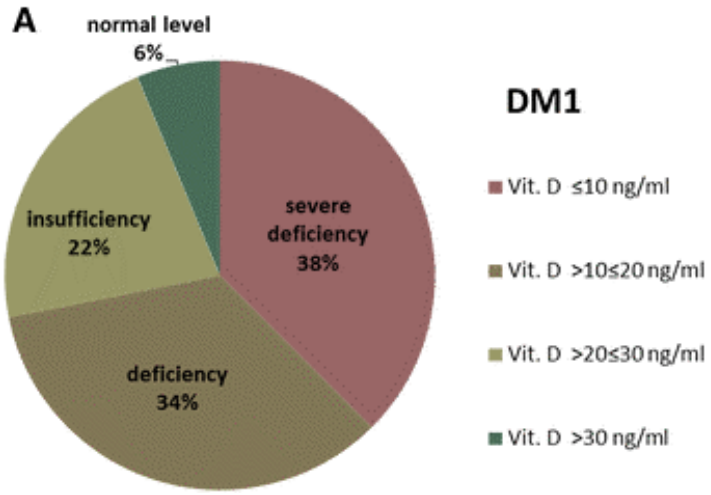
SB which is a non-patient cell line was used as a control. Sample size n=3, adjusted P values of <0.05 were considered to be statistically significant. (A-B) Transcripts expressed in muscle and heart tissue and consistently up/down-regulated in the DM1 cells and MBNL-deficient DM1 cells. (C-D) Transcripts expressed in brain tissue and consistently up/down-regulated in the DM1 cells and MBNL-Deficient DM1 cells. (E-F), Transcript expression up/down-regulated in the DM1 cells and restored in the MBNL-deficient DM1 cells. The experiment was performed in triplicate to ensure the reproducibility of the results.

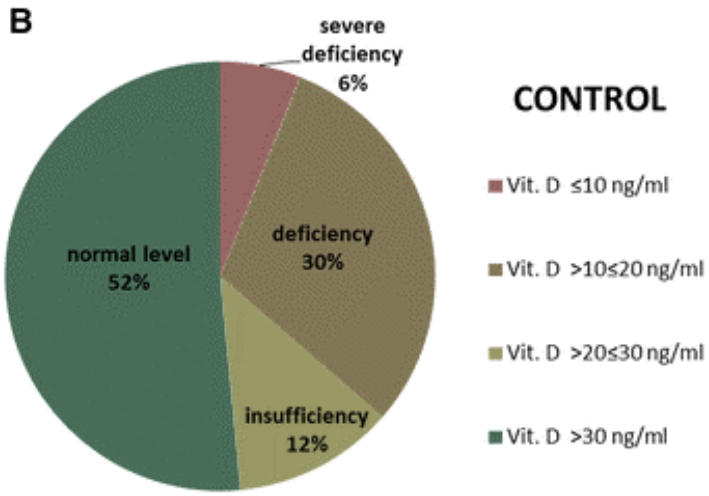
In addition to the observed alterations in alternative splicing, my study also identified numerous changes in gene expression that have been previously reported in association with DM1. To gain a better understanding of the molecular mechanisms underlying the disease, I focused my investigation on gene expression changes in brain, muscle, and heart tissues, as they are known to be particularly affected in DM1.

One gene of interest is NF1, which encodes for neurofibromin 1, a tumor-suppressor protein. Mutations in the NF1 gene have been implicated in the development of neurofibromatosis type 1, a genetic disorder characterized by uncontrolled or poorly controlled cell growth. Interestingly, neurofibromatosis and myotonic dystrophy have been observed to co-occur in a non-consanguineous family with a high degree of concordance [289]. In my study, global gene expression analysis using RNA-Seq data revealed a decrease in NF1 gene expression levels in DM1 cells compared to control cells (non-patient, SB). Intriguingly, following double knock-out of MBNL1 and MBNL2 in DM1 cells, NF1 expression was recovered to the same level as control cells (Figure 3-16 F). However, the exact role of MBNL proteins in the regulation of NF1 gene expression and its potential implications for the development of neurofibromatosis remain unclear and require further investigation.

Another gene of interest is vitamin D receptor (VDR). Studies have shown that vitamin D deficiency affects a significant proportion of DM1 patients, with up to 72% of individuals reporting various levels of deficiency [290] (Figure 3-17). The vitamin D receptor (VDR) plays a crucial role in mediating the biological effects of vitamin D in various tissues and cell types. Importantly, it exerts both genomic and non-genomic effects, all of which contribute to muscle differentiation and proliferation. One of its key genomic actions is to stimulate insulin-like growth factor-1 (IGF1). My research revealed that VDR expression was down-regulated in wild-type DM1, but was restored in MBNL-deficient DM1 cells (Figure 3-16 F). Consistently, I found that IGF1 transcription levels were reduced to almost zero in wild-type DM1, but were significantly increased in MBNL-deficient DM1 cells (Figure 3-16 F). Despite these findings, the precise mechanisms underlying the involvement of MBNL in vitamin D deficiency in DM1 remain unclear and require further investigation.

Figure 3-17 Vitamin D deficiency in DM1.





Prevalence of DM1 (A) and normal subjects (B) with different degrees of vitamin D deficit. Figure adapted from [290].

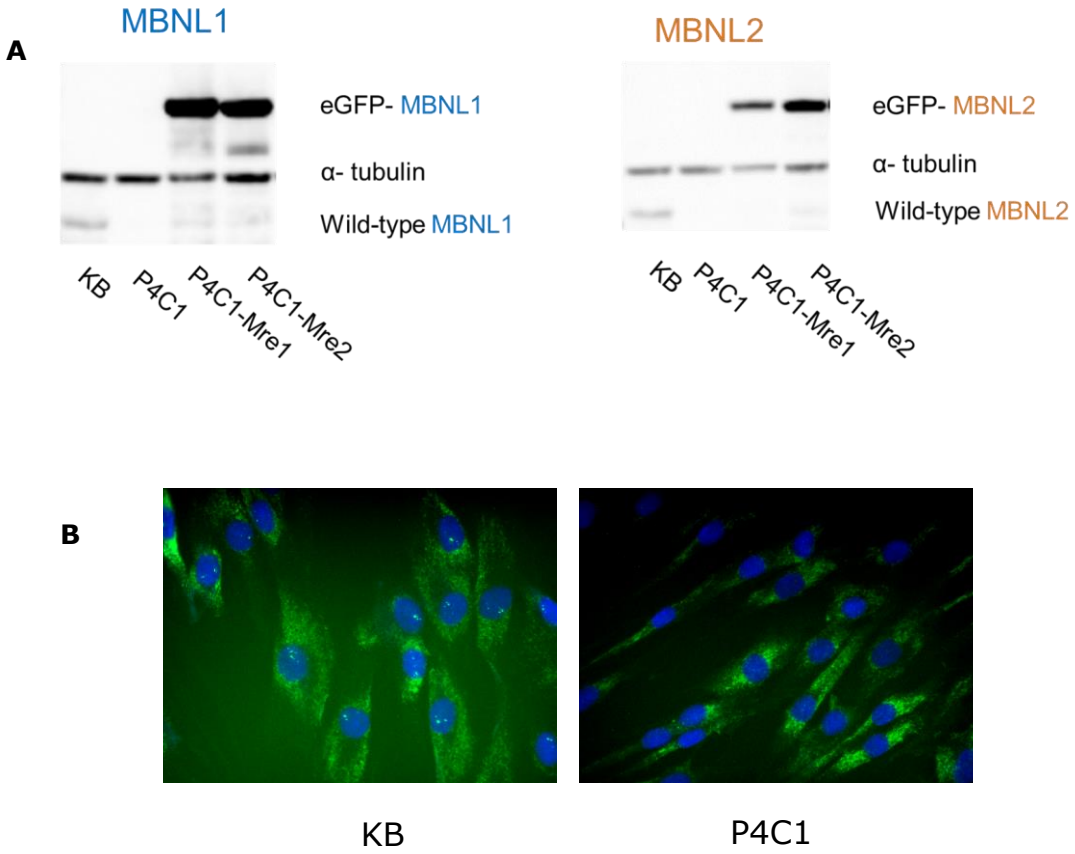
The results of this study suggest that in addition to alterations in alternative splicing, changes in gene expression may also contribute to the pathogenesis of DM1. The identification of NF1 and VDR as genes of interest may provide important insights into the underlying molecular mechanisms of the disease and could potentially lead to the development of new therapeutic targets. However, further investigation is needed to fully understand the role of MBNL proteins in regulating these genes and the potential implications for the development of Neurofibromatosis and vitamin D deficiency in DM1.

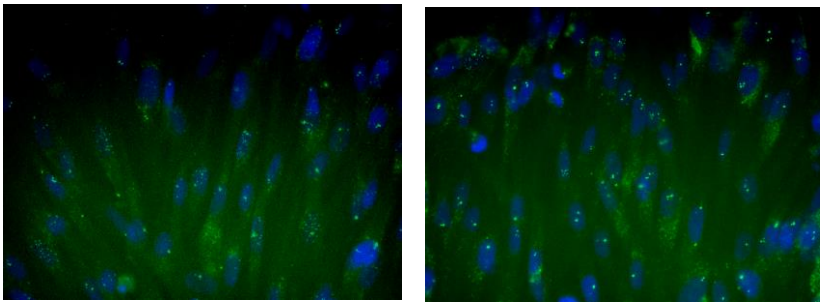
3.10 Re-expression of MBNL1 and 2

I conducted MBNL re-expression assays via lentiviral MBNL gene transfer to introduce eGFP-tagged MBNL1 and 2 cDNAs to P4C1 cells. I selected two MBNL re-expressing clones (P4C1-Mre1 and P4C1-Mre2) where both eGFP-tagged MBNL1 and 2 were successfully introduced and expressed (Figure 3-18 A). Next, I performed RNA in situ hybridisation assays and the results showed that the number of foci rose to a comparable level in the MBNL re-

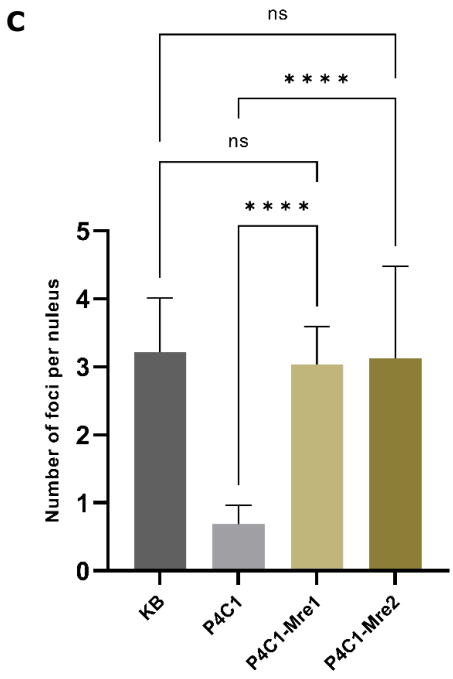
expressing cells to that observed in the wild-type DM1 (Figure 3-18 B and C). It further demonstrated that MBNLs were the determining factors of foci formation, the MBNL-$(CUG)^{exp}$ foci are dynamic structures and their formation is reversible. However, although the expression level of added-back MBNL1 and 2 was greater than the previous level of endogenous MBNLs, the number of foci didn't significantly increase. One explanation for this would be that MBNL sequestration had reached the RNA dependant level in our DM1 cells, that is the expansion RNA level was rate limiting. Thus the amount of MBNLs was more than that could be sequestered. Another explanation would be that the eGFP tagging to the MBNL proteins interferes with the MBNL and CUG repeats interaction.

Figure 3-18 Re-expression of MBNL1 and 2.





P4C1-Mre1 P4C1-Mre2



Re-expression of MBNL1 and MBNL2 in P4C1 (MBNL-deficient DM1 cells). (A) MBNL1 and 2 re-expression assessment by Western blots. P4C1-Mre1 and P4C1-Mre2 are MBNL re-expressing DM1 cell lines, derived from P4C1. (B) Images of KB, P4C1 P4C1-Mre1 and P4C1-Mre2 cells following RNA FISH using Cy3-$(CAG)_{10}$ probe (green) for the localisation of CUG repeats in the mutant *DMPK* mRNA and DAPI (blue) indicates the cell nuclei. (C) Quantification of Nuclear foci per nucleus. Histograms show data compiled from the RNA in situ hybridisation in KB (n=4355), P4C1 (n=5621) and P4C1-Mre1 (n=5694) and P4C1-Mre2 (n=6123) cells. n indicates the total number of cells analysed. Statistical significance was determined by one-way ANOVA (mean ± SD, **** for $P<0.0001$) or indicated as not significant (NS) for $P>0.05$.

3.11 Selective Nanopore sequencing of human *DMPK* by Cas9-assisted targeting of chromosome segments (CATCH)

The length of CTG repeat expansions in the *DMPK* gene is positively correlated with the severity of myotonic dystrophy and increases with successive generations, contributing to anticipation. Southern blot analysis often reveals a broad smear of genomic DNA restriction fragments containing the expanded CTG repeats, indicating somatic mosaicism [43, 291]. Somatic mosaicism can occur due to the instability of the CTG repeat expansion in the DMPK gene during cell division. This means that some cells in an affected individual may have a different number of CTG repeats than other cells, leading to variability in symptoms and disease severity and difficulty in measuring allele size.

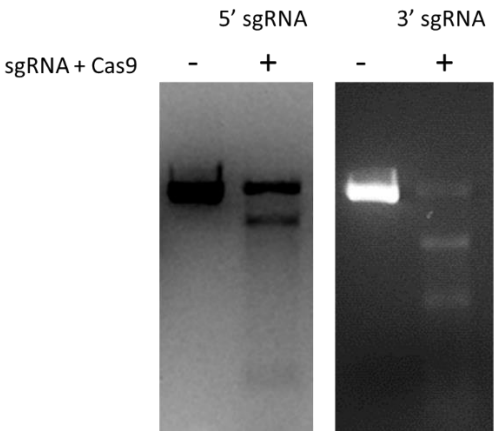
Although CTG repeat length correlates with disease severity and age of onset, the correlations are imprecise, and measuring allele size alone cannot accurately predict prognosis or anticipate disease progression. Technical difficulties in defining average allele length and measuring small shifts in repeat length over time contribute to this variation [291].

To further our understanding of the correlation between CTG repeat length and anticipation in myotonic dystrophy, more comprehensive and accurate studies are required.

I utilised the Cas9-assisted targeting of chromosome segments (CATCH) technique in conjunction with Nanopore DNA sequencing to accurately measure the CTG repeat lengths within the human *DMPK* gene. This approach enables precise targeting of specific chromosome regions and allows for high-throughput sequencing with long read lengths, enabling more accurate and comprehensive analysis of CTG repeat length variability. The

following target sequences by CRISPR/Cas9 were used for isolation of *DMPK* locus: 5' to the CTG repeat region, 5'-GGGCGTGTATAGACACCTGG-3' and 3' to the CTG repeat region, 5'-TGCGAACCAACGATAGGTGG-3'. The sgRNA activity was tested *in vitro* prior to Nanopore sequencing (Figure 3-19).

Figure 3-19 Assessment of the sgRNA activity by Cas9 in vitro assays.



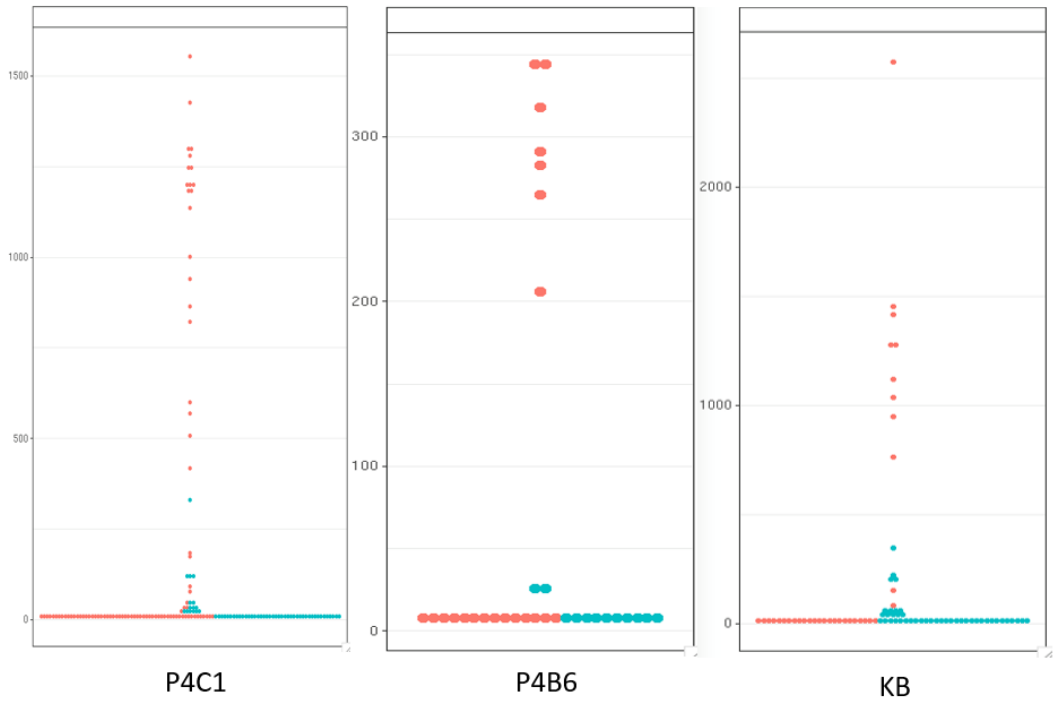
Ethidium bromide-stained gel showing the sgRNA activity following the performance of Cas9 *in vitro* assays. 5' sgRNA and 3' sgRNA recognise the sequences 5' and 3' to the CTG region in the *DMPK* gene, respectively.

The *DMPK* gene was cleaved at two sites flanking the CTG repeat region, according to the complementary gRNAs, and the fragment was isolated by pulsed-field gel electrophoresis (PFGE). 20—100 μg of target DNA was extracted from KB, P4C1 and P4B6 cells allowing sequencing library preparation from native DNA without pre-amplification. Nanopore sequencing, coupled with sequencing data analysis tools (see 2.13 for detail introduction), was employed to analyse the isolated fragments and accurately measure the CTG repeat length. Here, Nanopore sequencing has

preliminarily proven to be an effective approach to estimate precise tandem repeat length in each cell with its ability to produce long, single-molecule reads.

Nanopore sequencing generated a total of 213 raw reads in P4C1, out of which 162 were mapped to chromosome 19. Among the mapped reads, 143 contained CTG repeats within the DMPK gene. Similarly, P4B6 produced 81 raw reads, with 49 reads successfully mapped to chromosome 19, and 33 of these mapped reads contained CTG repeats. For KB, a total of 138 raw reads were obtained, with 98 mapped to chromosome 19, and 78 of these mapped reads contained CTG repeats.

Figure 3-20 Distribution of the length of CTG repeats in DM1 cell lines.



Y-axes indicate the number of CAG·CTG repeat units. Sequence reads from the forward and reverse strand of DNA are shown in orange and blue, respectively.

The analysis of the reads obtained from these three cell lines revealed important findings. Firstly, a substantial amount of reads (points in the bottom rows in the Figure 3-20) contained CTG repeats ranging from 6 to 8 units in these cell lines. This indicates that these reads represent the non-expanded allele of the DMPK gene, which appears to be somatically completely stable in these cell lines. Furthermore, it is evident that all three cell lines exhibited expanded CTG repeats in the DMPK gene (Figure 3-20). P4C1 displayed CTG repeat lengths ranging from 6 to 1553 units, with a median of 940 units for the expanded repeat lengths (>50 units). P4B6 exhibited CTG repeat lengths ranging from 6 to 346 units, with a median of 291 units for the expanded repeat lengths. Additionally, KB demonstrated CTG repeat lengths ranging from 6 to 2574 units, with a median of 284 units for the expanded repeat lengths. Notably, the sizes of CTG repeats varied among the cell lines, indicating somatic mosaicism [43, 291] in DM1 cells. The median values provide valuable information regarding the central tendency of the expanded repeat lengths in each cell line. These results suggest that Nanopore sequencing is an effective method for accurately estimating precise tandem repeat lengths with its capability of generating long, single-molecule reads. However, there was observed bias towards the forward strand in the nanopore sequencing data. One possible explanation for this bias would be that the library preparation protocol for Nanopore sequencing involves fragmenting the DNA and attaching adapters to the ends of the fragments. If there is bias during this step, such as preferential ligation of the adapters to one end of the DNA fragment, this could result in more reads being generated from one strand than the other. Another possibility could be base-calling errors. The base-calling software used to generate sequence data from the raw nanopore signal can introduce errors.

If the base-caller is more accurate for one strand than the other, this could result in more reads being generated from the more accurately base-called strand.

Despite the strand bias, Nanopore sequencing provides several benefits compared to other methods such as PCR-based methods or Southern blot analysis: 1) Long-read sequencing enables the detection of complex repeat structures, including interruptions and expansions in non-canonical regions, that may be missed by other methods. 2) The direct sequencing of the CTG repeat region eliminates the need for amplification, reducing the risk of PCR artifacts and bias. 3) Nanopore sequencing is a powerful technique for individually measuring the CTG repeat length in a single strand of DNA. Due to somatic mosaicism, the size of CTG repeat expansions can vary significantly between cells or tissues. In contrast, PCR-based methods often display CTG repeat fragments as a broad smear band, making it challenging to accurately measure the length. By providing long-read sequencing data, Nanopore sequencing can overcome this limitation, enabling the detection of complex repeat structures and identifying potential somatic mosaicism. 4) The rapid turnaround time of Nanopore sequencing allows for the timely diagnosis of myotonic dystrophy and facilitates research studies.

Overall, Nanopore sequencing offers a promising solution for precise CTG repeat length determination in myotonic dystrophy patients, providing new insights into the genetic mechanisms and variability underlying the disorder.

3.12 Discussion and conclusions

MBNL proteins are a class of RNA-binding proteins that play a crucial role in regulating alternative splicing events in cells [61].

DM1 is caused by a CTG repeat expansion in the 3'-UTR of the *DMPK* gene [40]. Once transcribed, expanded CUG repeats fold into highly stable and long RNA hairpin structures that sequester MBNLs and other nuclear factors to form ribonuclear foci [73, 292]. The sequestration of MBNLs causes their loss-of-function, which directly impairs the alternative splicing of several key regulatory target pre-mRNAs in muscles and neural cells. Particularly, mis-splicing of muscle chloride channel (*CLCN1*) [62], calcium channel voltage-dependent L type alpha 1S subunit (*CACNA1S*) [64], bridging integrator 1 (*BIN1*) [177] and insulin receptor (*INSR*) [63], results in myotonia, muscle weakness and insulin resistance, respectively, constituting key hallmarks of DM1. However, the effects of *MBNL1* and *2* double knock-out on DM1 are complex and multi-faceted. While the sequestration of MBNLs causes the amount of free MBNLs to be insufficient in DM1 cells, further knock-out of *MBNL1* and 2 in DM1 cells exacerbates many mis-splicing events and the mis-regulation of expression of many genes implicated in the disease. Additionally, it can lead to the emergence of novel aberrant splicing events and dysregulated gene expressions. These findings suggest that there is an excess of MBNL in KB nuclei. Furthermore, the double-inactivation of MBNL1 and 2 can also cause the instability of RNA foci formed by mutant *DMPK* transcripts, facilitating the degradation or nuclear export of these transcripts. Subsequently, the dissipation of RNA foci could lead to the release of other RNA binding proteins that are sequestered by the toxic RNAs, potentially rescuing other aberrant alternative splicing events and dysregulated gene expressions.

Therefore, it is important to consider both the release of MBNLs and the promotion of mutant *DMPK* mRNA decay when developing therapeutic strategies for DM1. A more nuanced understanding of

the complex interplay between MBNL proteins, toxic RNAs, and RNA processing is needed for researchers to develop more effective treatments for this debilitating disease.

Chapter 4 Detection of *DMPK* mRNA using Super-Resolution Single Molecule Localisation Microscopy (SMLM)

4.1 Abstract

The pathophysiology of myotonic dystrophy is primarily driven by the sequestration of MBNL proteins and other nuclear factors by repeat expansion DMPK transcripts, resulting in the formation of RNA foci. To gain a deeper understanding of these RNA foci and their dynamics, I propose the following hypotheses: 1) By utilizing super-resolution single molecule localization microscopy (SMLM), I can unravel the characteristics of $(CUG)^{exp}$ RNA foci in both DM1 cells and MBNL-deficient DM1 cells. 2) The application of STORM technology, as part of SMLM, can provide a valuable tool for studying RNA foci in other diseases associated with short tandem repeat mutations.

To achieve my research goals, I established the following aims:

1. Utilise RNA-FISH and stochastic optical reconstruction microscopy (STORM) to overcome optical limitations and gain unprecedented insights into the organization and distribution of RNA foci.
2. Investigate the role of MBNL proteins in the dynamics of RNA foci.

My findings revealed the presence of both large and micro RNA foci, with quantification of fluorophore signals distinguishing between them. Micro foci were identified as breakdown products of large RNA foci, while large foci were confined to the nucleus, and micro foci had the ability to migrate to the cytoplasm, potentially contributing to RAN translation in DM1. Additionally, my study expanded upon previous research, showing that the absence of MBNL1 modulates the breakdown of mutant DMPK

mRNA. Specifically, in the absence of MBNLs, mutant DMPK mRNA formed micro RNA foci instead of large, stable foci, promoting their degradation or transportation to the cytoplasm. These findings deepen our understanding of RNA foci dynamics and the role of MBNL proteins in DM1 pathogenesis. Furthermore, this research enhances my knowledge of the underlying mechanisms contributing to DM1 and opens new avenues for investigating similar pathological processes in related disorders.

4.2 Principle of STORM

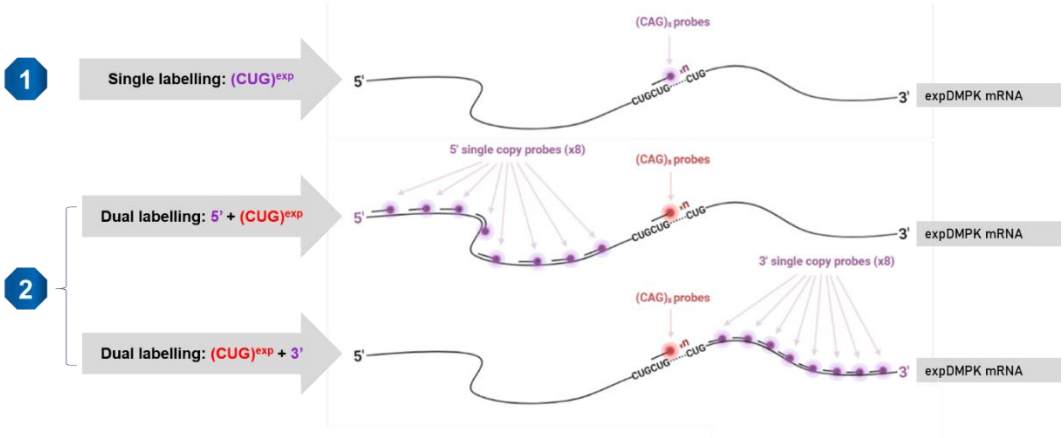
I applied RNA-FISH and STORM [252, 258] to examine potential micro foci or break-down products of expansion *DMPK* mRNA transcripts, the visualisation of which is limited when using conventional microscopes. For example, under the high-sensitivity widefield microscope, large $(CUG)^{exp}$ foci were clearly visible in DM1, and some tiny faint foci were also observed in MBNL-deficient DM1 cells, P4B6 and P4C1 (Figure 4-2 A, E and G). However, due to the resolution limit in conventional microscopy and overly concentrated probe-labelling in a single RNA focus, the size or the intensity of RNA foci cannot be accurately estimated. By taking advantage of STORM, we have bypassed these issues. STORM is a type of super-resolution single molecule localisation microscopy (SMLM), which captures fluorophore signals by recording their coordinates and other corresponding parameters over 10-20 thousand frames [253, 258]. Only a sparse subset of fluorophores are stochastically activated in each frame, their associated point spread functions (PSFs) do not overlap on the detector, allowing the precise localisation of each emitter as the centre of their PSF. To reconstruct the full fluorophore population distribution, fluorophores are 'turned on and off' and acquired in successive frames over time, with each emitter's localised coordinates recorded in each frame. The reconstructed image data

therefore consists of a list of localisation coordinates from all data frames, which are shown as crosses in the images, as well as the associated localisation uncertainties. Then by running the Bayesian cluster analysis with the obtained fluorophore coordinates [263], I were able to quantify how many localisations there were in each cluster.

4.3 CUG repeat single-labelling

In the first instance, I attempted to detect the CUG repeat expansion in the *DMPK* mRNA only by using an Atto647n-$(CAG)_8$ probe (Figure 4-1). The $(CAG)_8$ probe is not expected to detect the wild-type DMPK transcript which typically contains 6-8 CUG units in KB and SB cells. However, it is possible that the $(CAG)_8$ probe could potentially bind to the wild-type DMPK transcripts, albeit with signals that are too weak to be reliably detected or distinguished from background signals.

Figure 4-1 Scheme of the probe design for the stochastic optical reconstruction microscopy (STORM).

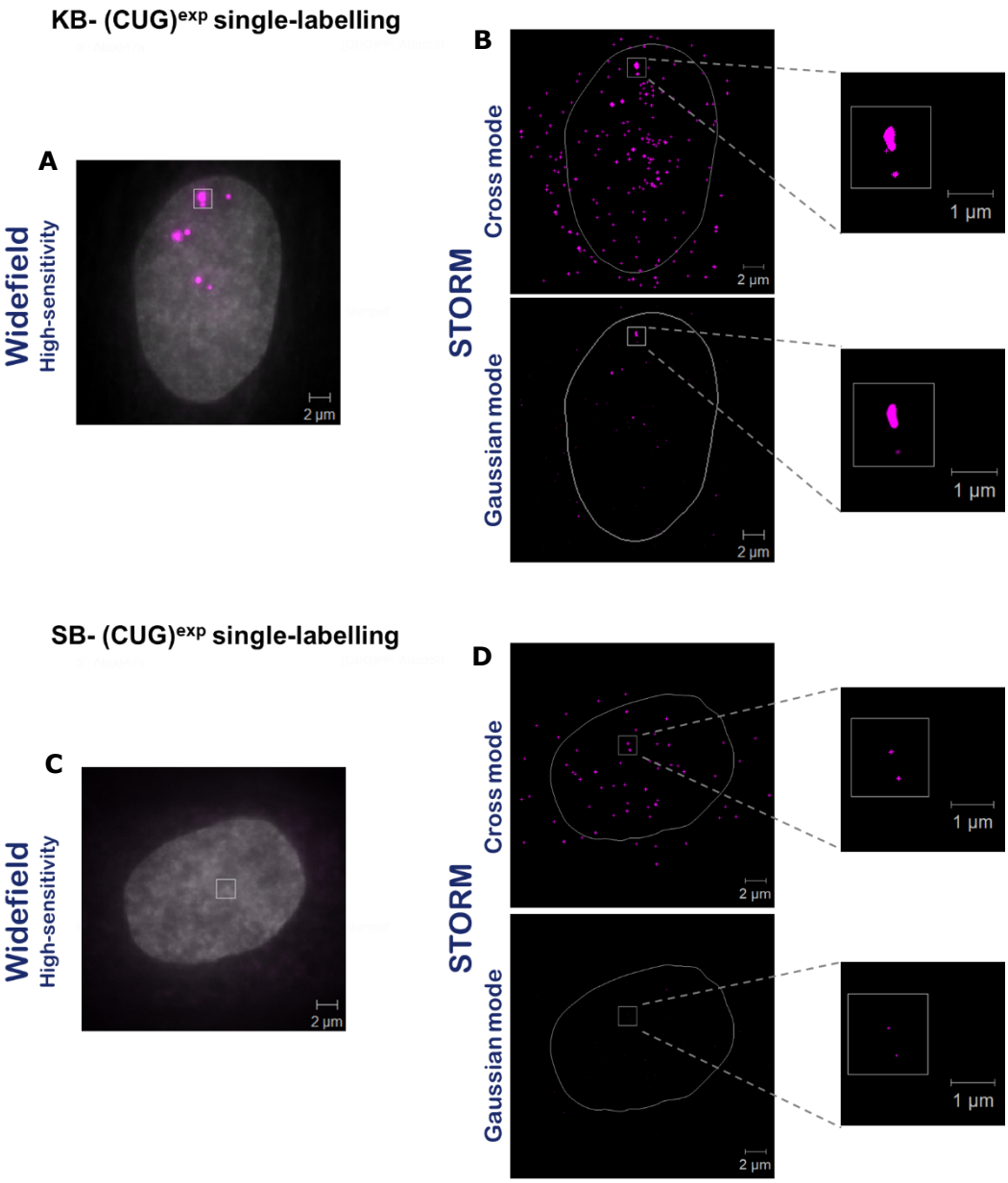


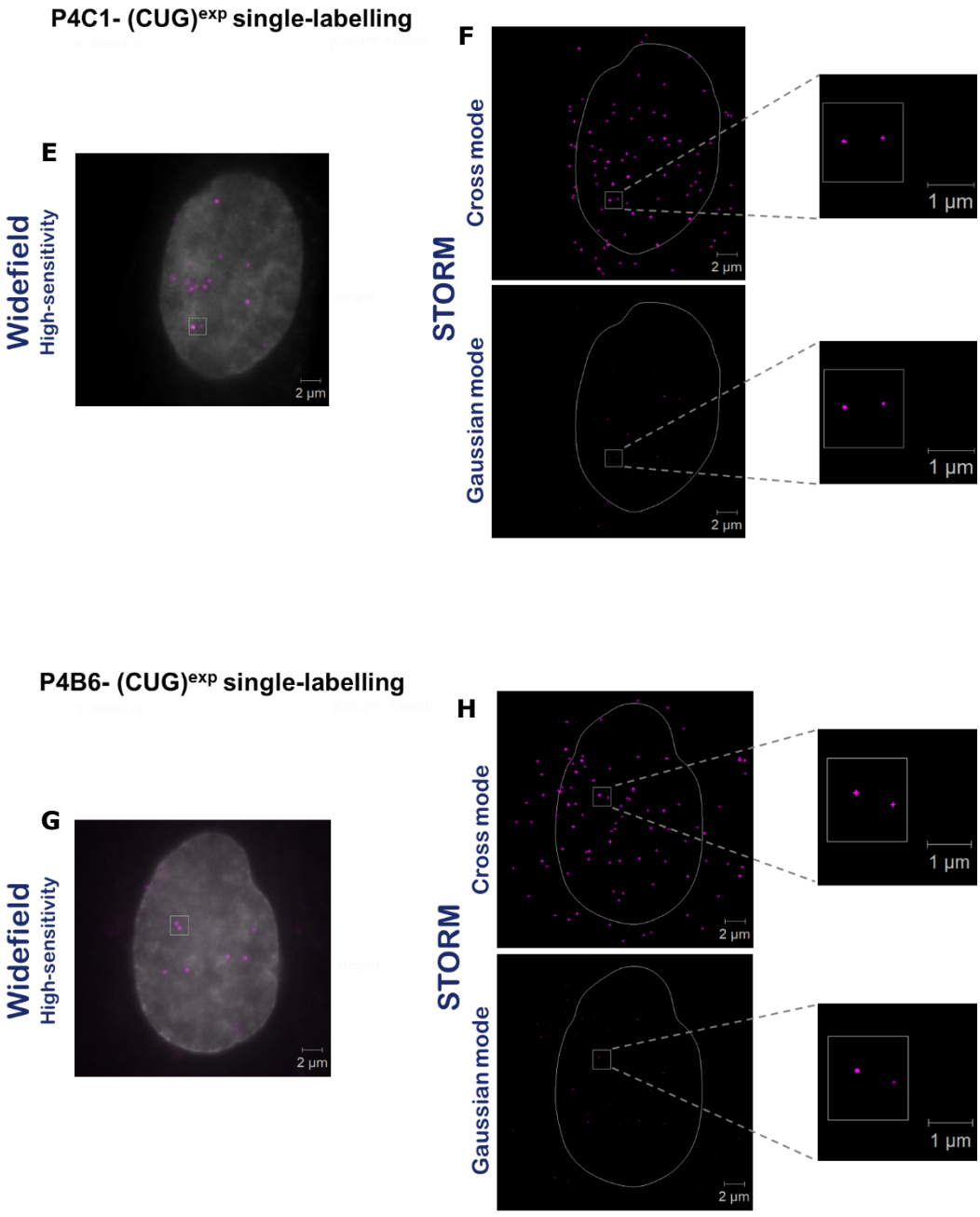
In the single labelling experiment, Atto647n-$(CAG)_8$ probes were designed to bind the CUG repeat region in the *DMPK* transcript. In the dual labelling experiment, Atto550-$(CAG)_8$ probes were designed to bind to the CUG repeat region and 8 Atto647n-labelled 5' or 3' single copy probes were designed to bind to the 5' or 3'

flanking region of the CUG repeats in the *DMPK* transcript, respectively.

Figure 4-3 shows the compiled data on the quantification of the number of localisations in $(CUG)^{exp}$ clusters, obtained through Bayesian cluster analysis with the corresponding fluorophore coordinates. To ensure consistency with my observations using conventional microscopy, I set two thresholds. Signals from SB were utilised to identify clusters derived from non-$(CUG)^{exp}$ or off-target signals. Specifically, I found that 95% of clusters identified with the $(CAG)_8$ probe in SB contained $\leqslant$ 10 localisations. Based on these criteria, I categorised the clusters into three distinct groups: Large $(CUG)^{exp}$ foci (containing >250 localisations), micro $(CUG)^{exp}$ foci (containing 11-250 localisations) and non-$(CUG)^{exp}$ signals (containing $\leqslant$ 10 localisations). Notably, I observed large foci with 1,000 or even 2,000 localisations only in the DM1 nucleus. In contrast, micro foci were detected in the nuclei of both DM1 and MBNL deficient DM1 cells. These results suggest that micro $(CUG)^{exp}$ foci may represent break-down products of the expansion CUG repeat region in the *DMPK* mRNA.

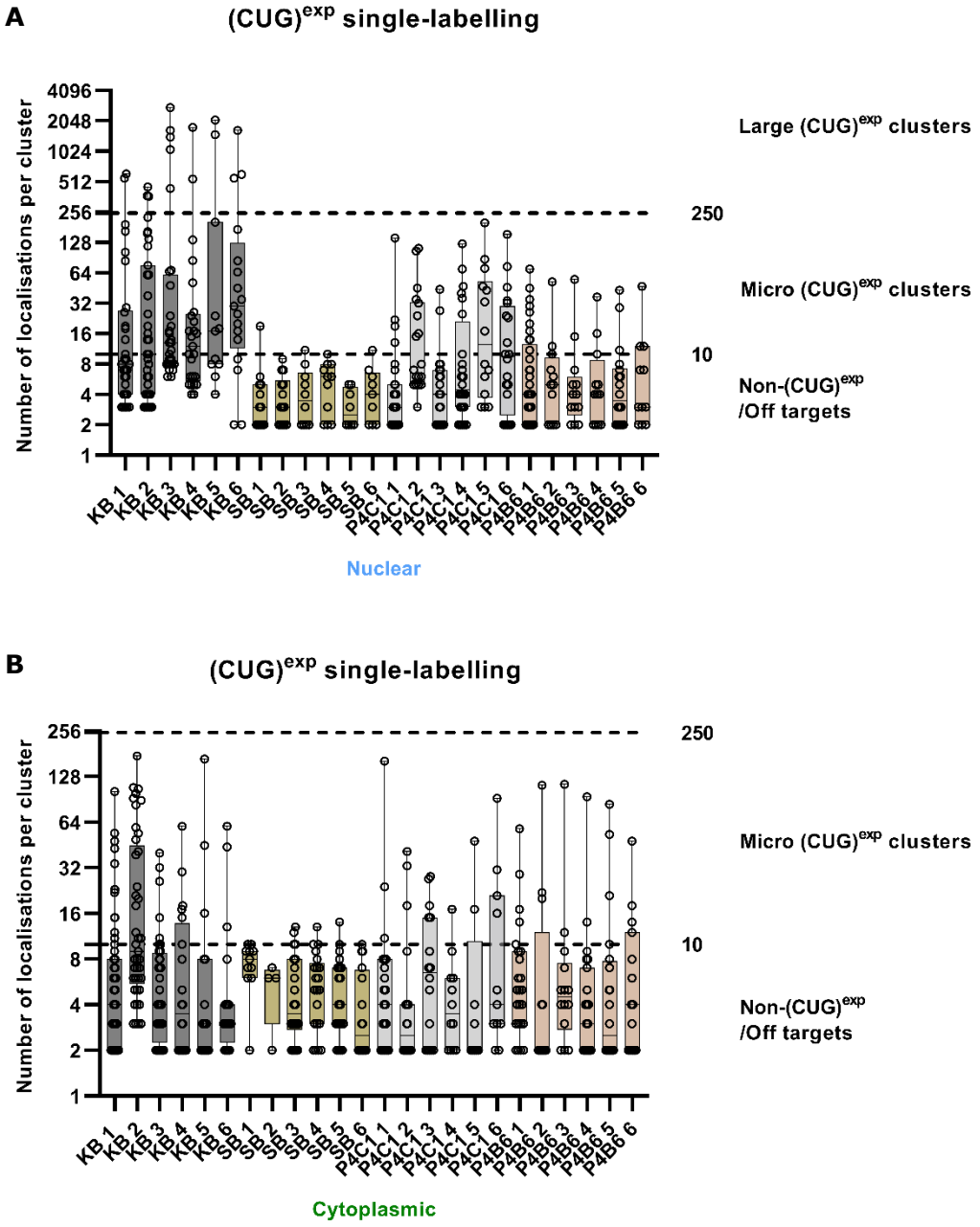
Figure 4-2 Detecting the CUG repeat expansion in the DMPK transcript with STROM.





A, C, E and G show the conventional high-sensitivity widefield images of DM1 (KB), Non-DM (SB) and 2 MBNL-deficient DM1 cell lines (P4C1 and P4B6) following RNA FISH using Atto647N-$(CAG)_8$ probes (magenta) for the localisation of the CUG repeats and Hoechst (grey) to indicate the cell nuclei; B, D, F and H show the STORM images of the above cell lines. Individual fluorophore localisation coordinates were recorded by STORM. Under cross mode, each fluorophore signal (blinking) is shown as a cross; under gaussian mode, fluorophore signals are reconstructed to be shown in the way by mimicking the visual effect of human eyes.

Figure 4-3 Quantification of the number of localisations in (CUG)^{exp} clusters.

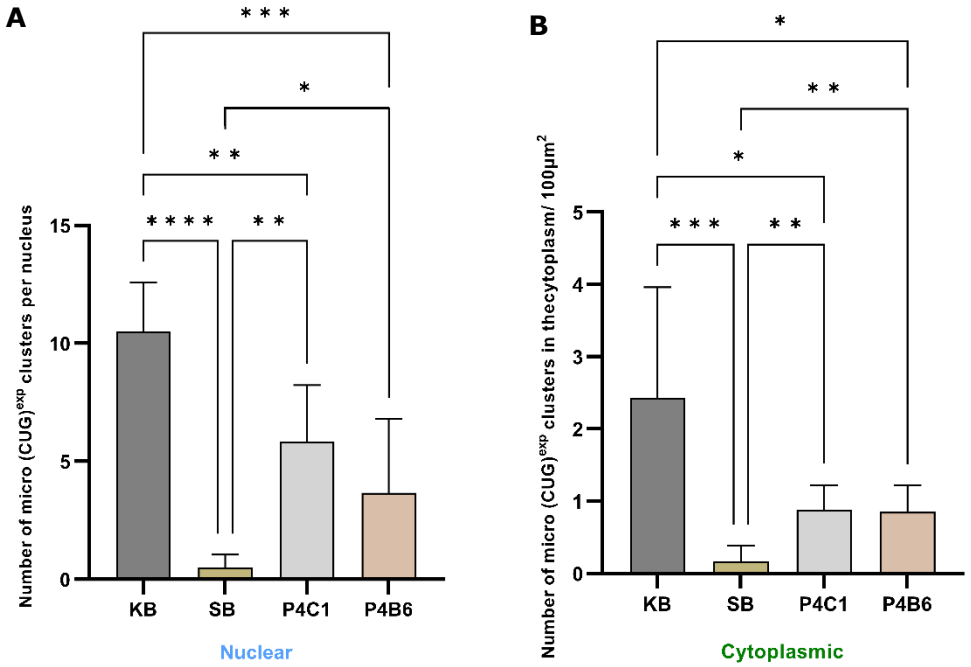


The diagrams show the number of fluorophore localisations in each RNA cluster identified with the $(CAG)_8$ probe in the nucleus (A) and in the cytoplasm (B). The boxes represent the interquartile range (IQR) of the data, encompassing the middle 50% of observations. The whiskers, depicted as lines extending from the boxes, indicate the minimum and maximum values among all data points. Clusters containing >250 localisations are considered as large clusters/foci; clusters containing 11-250 localisations are considered as micro clusters/foci; clusters

containing $\leqslant$ 10 localisations are considered as non-$(CUG)^{exp}$ or off-targets.

To accurately quantify the number of clusters/foci in the cytoplasm, I computed the average number of clusters/foci per 100 μm^2 cytoplasmic area. My results were consistent with previous studies, as no large $(CUG)^{exp}$ foci were detected in the cytoplasm of DM1 cells, KB nor the MBNL-depleted derivatives P4B6 and P4C1 (Figure 4-3 B). But intriguingly, I observed the presence of micro $(CUG)^{exp}$ foci in the cytoplasm of all three cell lines (Figure 4-3 B). However, the number of micro foci in the nucleus and cytoplasm of P4B6 and P4C1 were significantly decreased compared to those observed in KB cells (Figure 4-4). Thus it would appear that the absence of MBNL proteins in P4B6 and P4C1 impacts on the formation of large $(CUG)^{exp}$ foci and also on the micro foci. One possible explanation for my findings is that the micro foci represent the breakdown products of the large nuclear foci and their reduced structure allows the micro foci to be exported to the cytoplasm. It may also suggest that the MBNL proteins strengthen the nuclear retention of mutant *DMPK* transcripts and in the absence of MBNL, the mutant *DMPK* transcripts are degraded more rapidly.

Figure 4-4 Quantification of the number of micro (CUG)^{exp} foci in KB, SB, P4C1 and P4B6 (n=6).



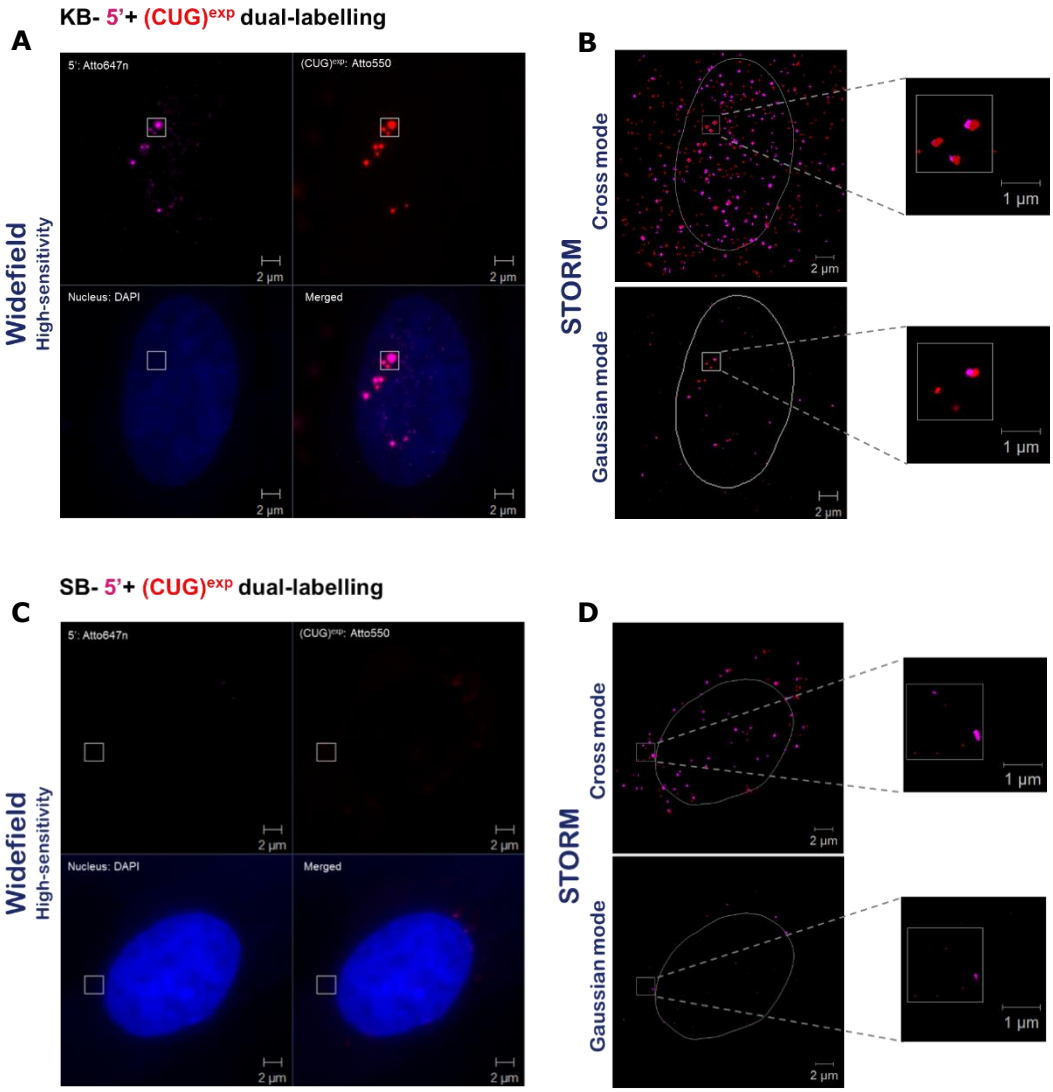
(A) Histograms show the mean number of micro (CUG)^{exp} foci per nucleus. (B) Histograms show the mean number of micro (CUG)^{exp} foci per 100µm^2 cytoplasmic area. n indicates the total number of cells analysed. Statistical significance was determined by one-way ANOVA (mean ± SD, * for P<0.05, ** for P<0.01, *** P<0.001, **** for P<0.0001).

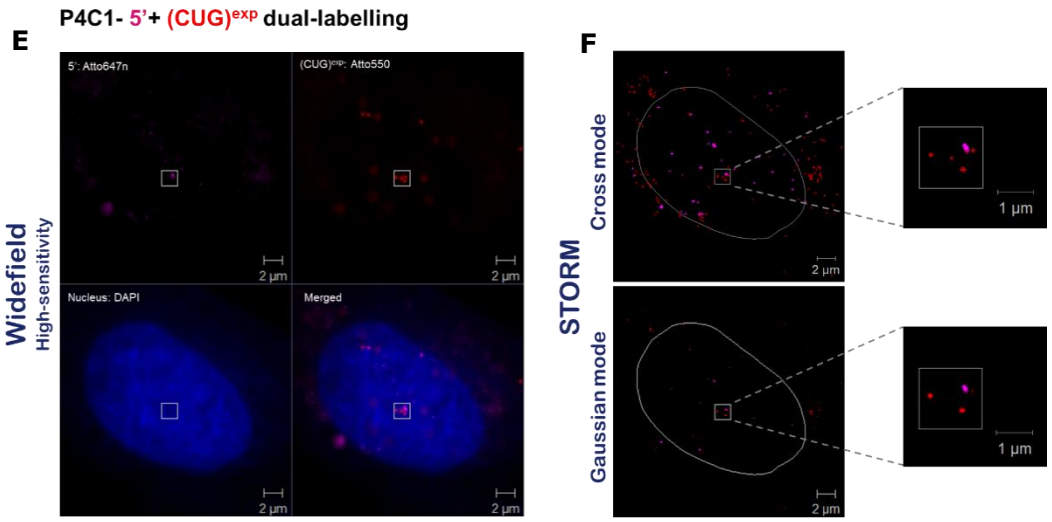
4.4 Dual labelling the CUG repeats and single copy 5′ or 3′ sequences

The above findings led me to question the extent to which (CUG)^{exp} transcripts are flanked by 5′ or 3′ sequence, which may reflect different degradation states of the mutant *DMPK* mRNA. In order to investigate this, I designed probes from three intervals of the *DMPK* transcript which allowed me to determine signals from regions 5′ and 3′ to the repeat expansion and from the repeat itself. For the 5′ and 3′ labelling experiments, I designed and tested a collection of 8 non-overlapping Atto647n-labelled single copy probes for each side of the repeat. I used an Atto550-(CAG)$_8$

probe to bind to the CUG repeat region in the *DMPK* transcript. Detailed probe binding positions in the DMPK cDNA are illustrated in Figure 4-1 and Figure 0-9. I performed two sets of dual-labelling STORM assays independently (Figure 4-5) (Figure 4-6). In one I used the 5′ probe set with the CUG repeat region in a dual-labelling experiment and in the other I used the 3′ probe set with the CUG repeat region dual-labelling.

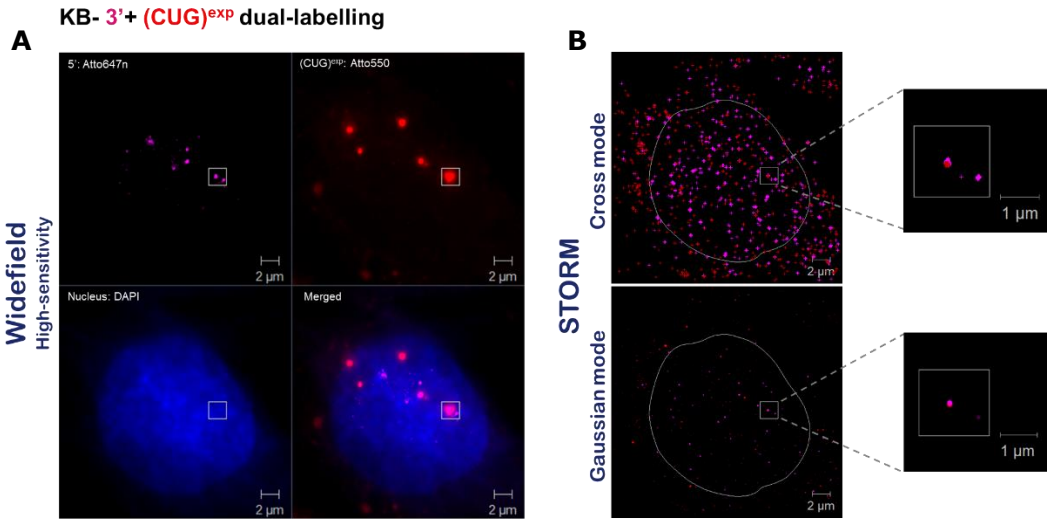
Figure 4-5 Detecting the CUG repeat region and its 5′ end flanking regions in the DMPK transcript with STORM.

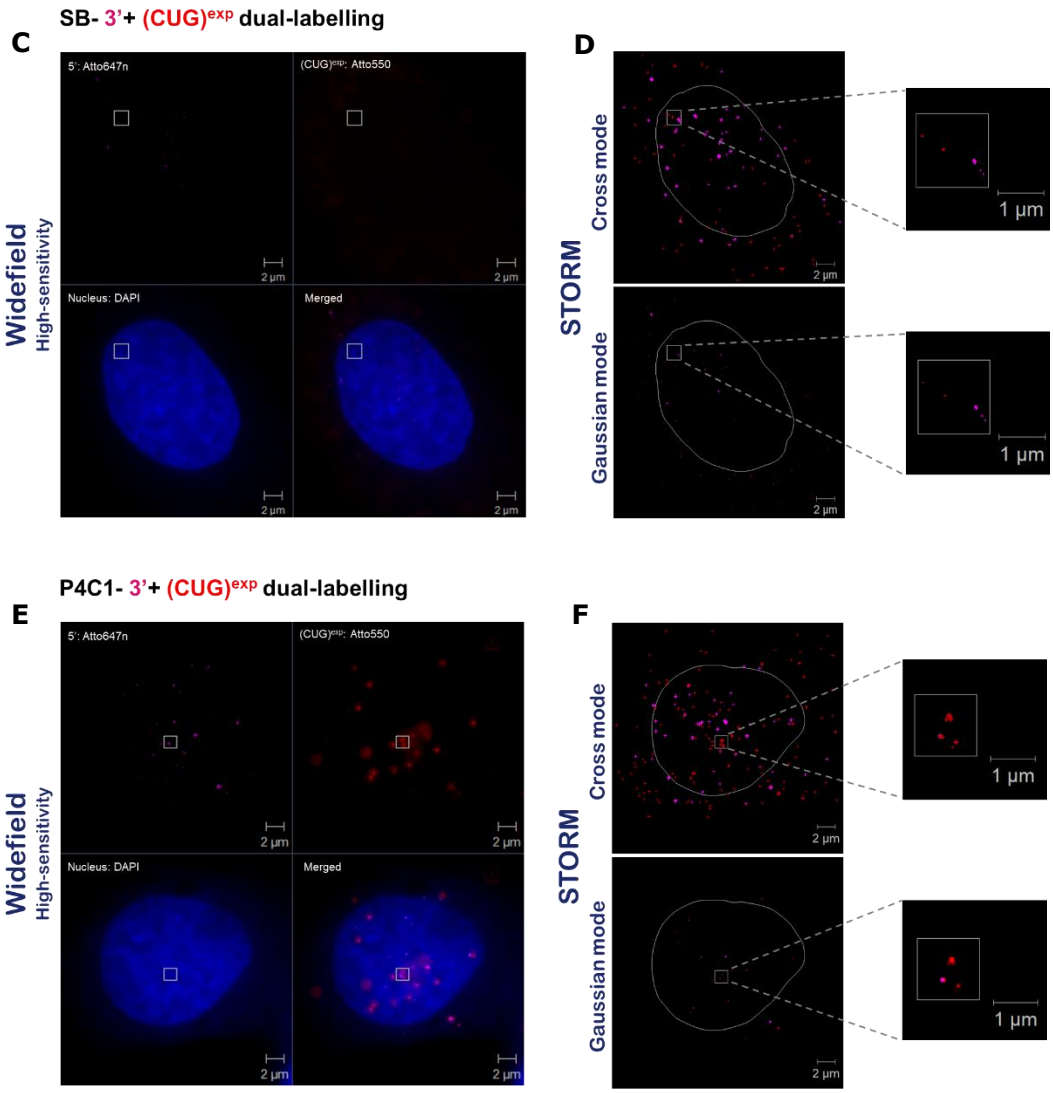




(A, C, E) Images show the conventional high-sensitivity widefield images of DM1 (KB), Non-DM (SB) and MBNL-deficient DM1 (P4C1) cells following RNA FISH using the Atto550-5' single copy probe set for the localisation of 5' end flanking regions (Details of the probes are described in the results); Atto647N-$(CAG)_8$ probe (magenta) for the localisation of $(CUG)^{exp}$ *DMPK* transcripts and DAPI (blue) to indicate the cell nuclei; (B, D, F) Images show the STORM images of the above cell lines. Individual fluorophore localisation coordinates were recorded by STORM. Under cross mode, each fluorophore signal (blinking) is shown as a cross; Under gaussian mode, fluorophore signals are reconstructed to be shown in the way by mimicking the visual effect of human eyes.

Figure 4-6 Detecting the CUG repeat region and its 3' end flanking regions in the DMPK transcript with STORM.





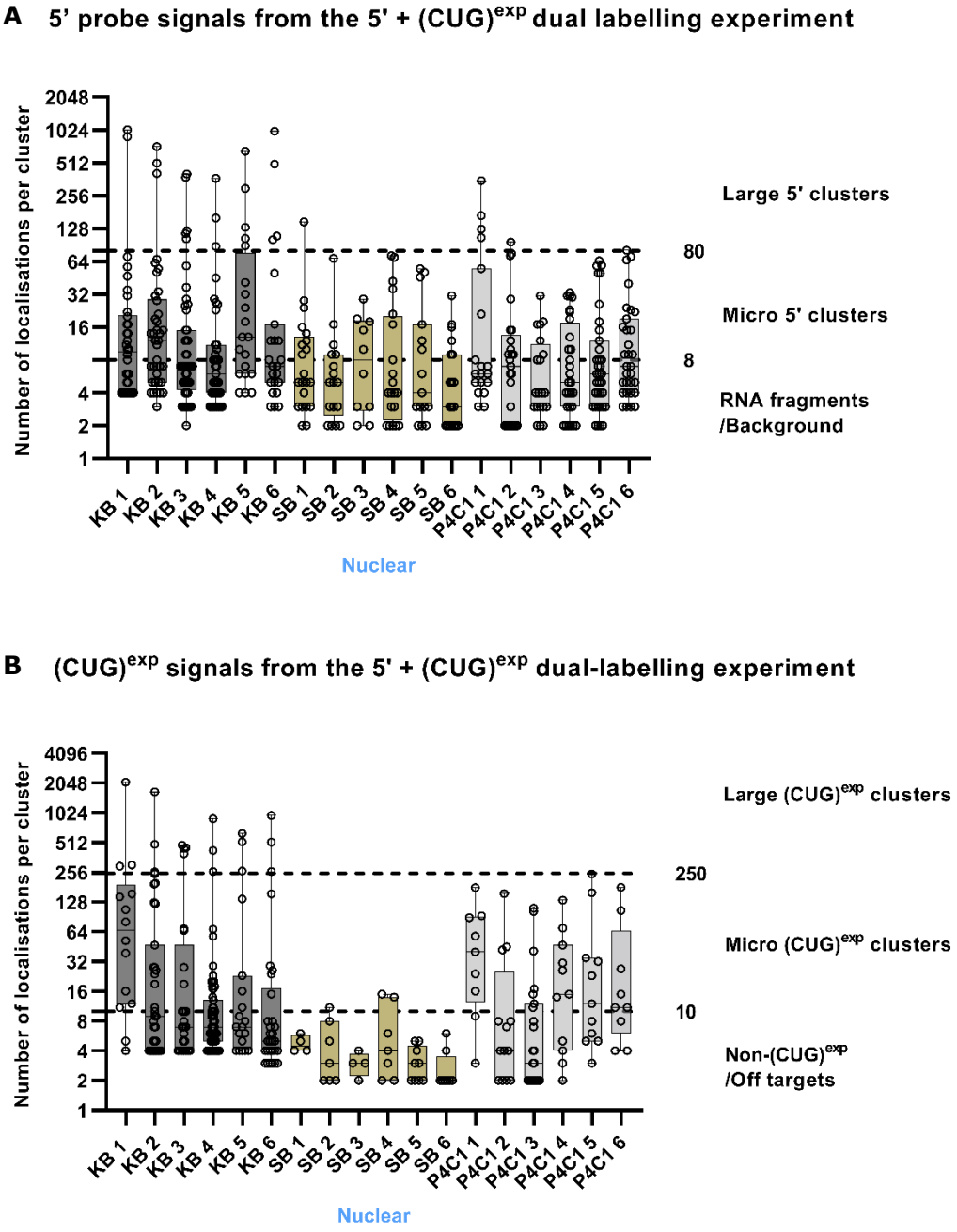
(A, C, E) Images show the conventional high-sensitivity widefield images of DM1 (KB), Non-DM (SB) and MBNL-deficient DM1 (P4C1) cells following RNA FISH using the Atto550-3' single copy probe set for the localisation of 3' end flanking regions (Details of the probes are described in the results); Atto647N-$(CAG)_8$ probe (magenta) for the localisation of $(CUG)^{exp}$ *DMPK* transcripts and DAPI (blue) to indicate the cell nucleus; (B, D, F) Images show the STORM images of the above cell lines. Individual fluorophore localisation coordinates were recorded by STORM. Under cross mode, each fluorophore signal (blinking) is shown as a cross; Under gaussian mode, fluorophore signals are reconstructed to be shown in the way by mimicking the visual effect of human eyes.

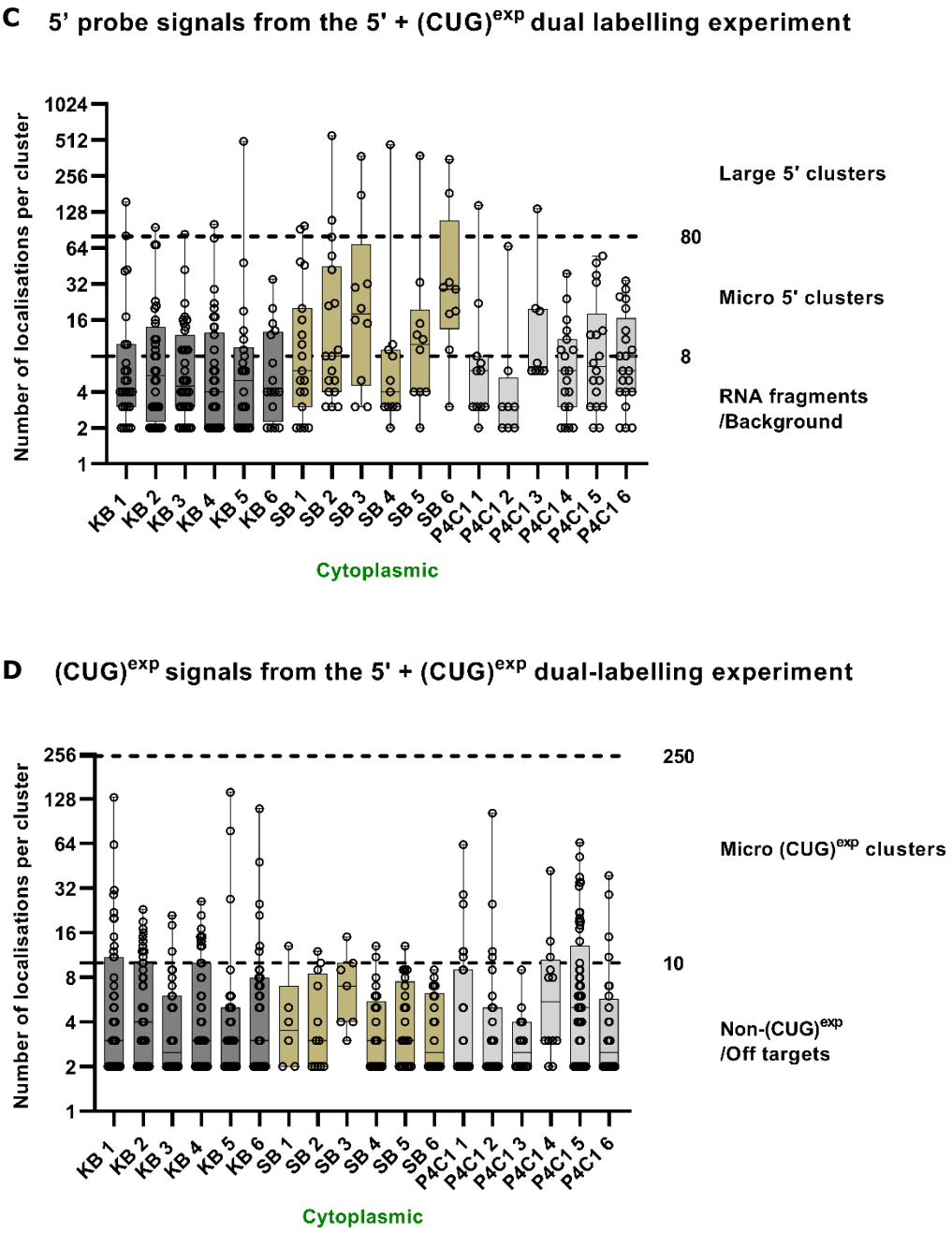
I applied the same criteria used in the $(CUG)^{exp}$ single-labelling experiments to classify the $(CUG)^{exp}$ clusters observed in the dual

labelling experiments. The $(CAG)_8$ probe complementary to the *DMPK* repeat expansion is not expected to detect the wild-type *DMPK* transcript. However, this is not the case for the probe sets complementary to the 5’ and 3’ to the repeat expansion. I utilised an additional Atto647n-labelled scramble probe as a negative control to identify false or background signals in this scenario. Analysis showed that 95% of the clusters identified with the scramble probe contained 2-7 fluorophore signals (Figure 0-4). Thus, clusters identified with the Atto647n-labelled 5' or 3' probe sets that contained the same number of fluorophore signals were considered to be derived from non-specific binding and were excluded from further result interpretation and discussion. As can be seen in Figure 4-7 and Figure 4-8, in the SB nucleus, the majority of clusters identified with the 5’ or 3’ probe sets contained 80 fluorophore signals or less. As the normal *DMPK* transcripts should be present singly in the nucleus I assume these clusters with 8 to 80 fluorophore signals (here named micro 5’ or 3’ end clusters) were derived from single *DMPK* transcripts and those with more than 80 fluorophore signals (here named large 5’ or 3’ end clusters) were considered to contain multiple *DMPK* transcripts.

The number of fluorophore signals from the 5' or 3’ probe sets varied widely in the KB nucleus, ranging from 8 to 1802 localisations per cluster (Figure 4-7) (Figure 4-8). Furthermore, the number of large or micro clusters identified with 5’ or 3’ probe sets was significantly increased in KB nuclei compared to those detected in SB nuclei (Figure 4-9 A and C) indicating elevated levels of nuclear *DMPK* transcripts.

Figure 4-7 Quantification of the number of localisations in (CUG)^{exp} clusters and in 5' end clusters.

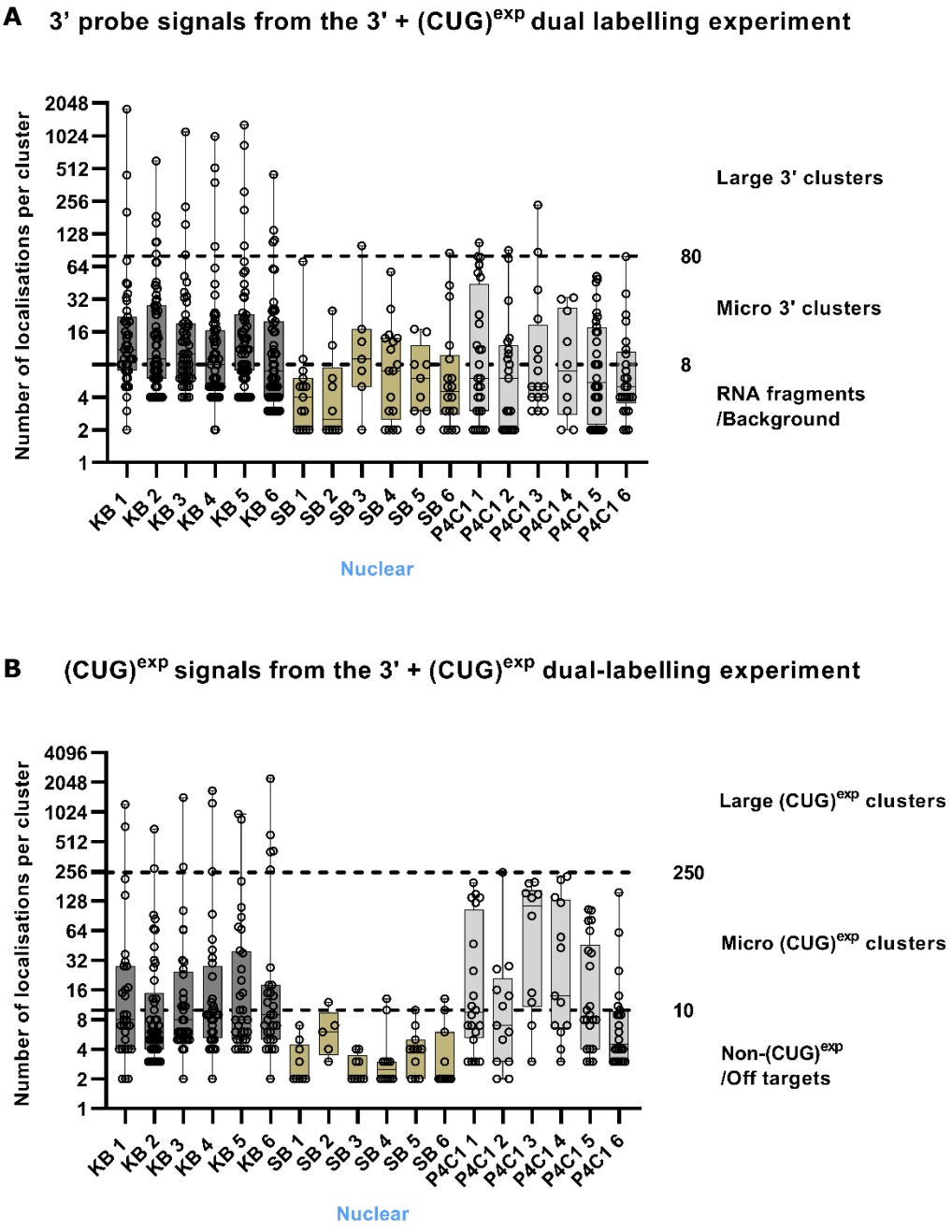




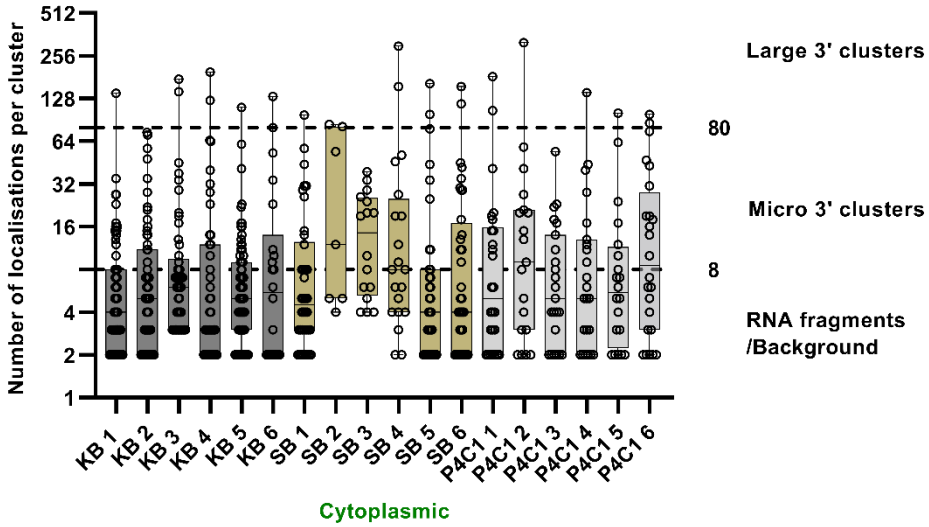
The Diagrams show the number of localisations in each cluster identified with the $(CAG)_8$ probe or the 5′ probe set in the nucleus and in the cytoplasm. The boxes represent the interquartile range (IQR) of the data, encompassing the middle 50% of observations. The whiskers, depicted as lines extending from the boxes, indicate the minimum and maximum values among all data points. $(CUG)^{exp}$ clusters containing >250 localisations are considered as large $(CUG)^{exp}$ clusters; $(CUG)^{exp}$ clusters containing 11-250 localisations are considered as micro $(CUG)^{exp}$ clusters/foci; $(CUG)^{exp}$ clusters containing $\leqslant$ 10 localisations are considered as non-$(CUG)^{exp}$ or off-targets. 5′ end clusters containing >80

localisations are considered as large 5’ end clusters; 5’ end clusters containing 8-80 localisations are considered as micro 5’ end clusters; 5’ end clusters containing <8 localisations are considered as transcript fragments or background.

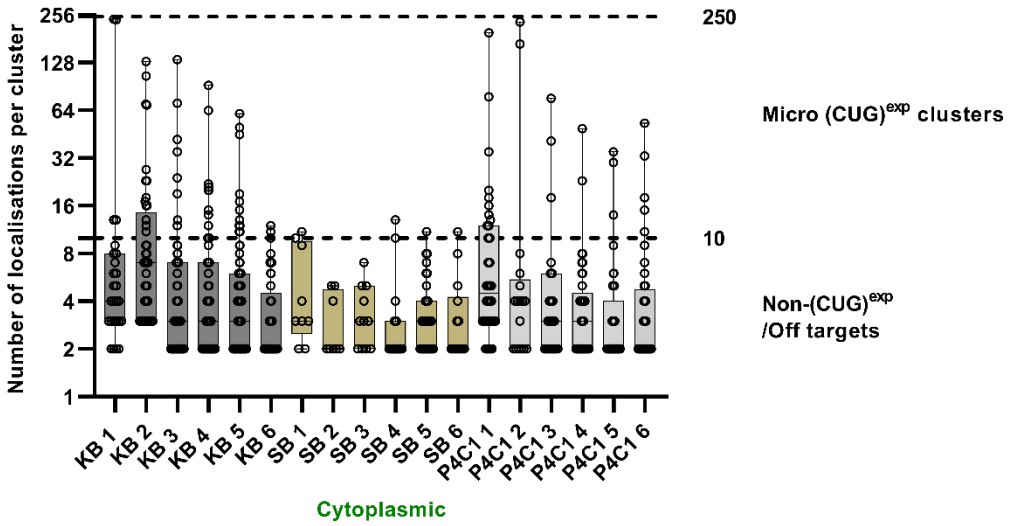
Figure 4-8 Quantification of the number of localisations in $(CUG)^{exp}$ clusters and in 3’ end clusters.



C 3' probe signals from the 3' + $(CUG)^{exp}$ dual labelling experiment



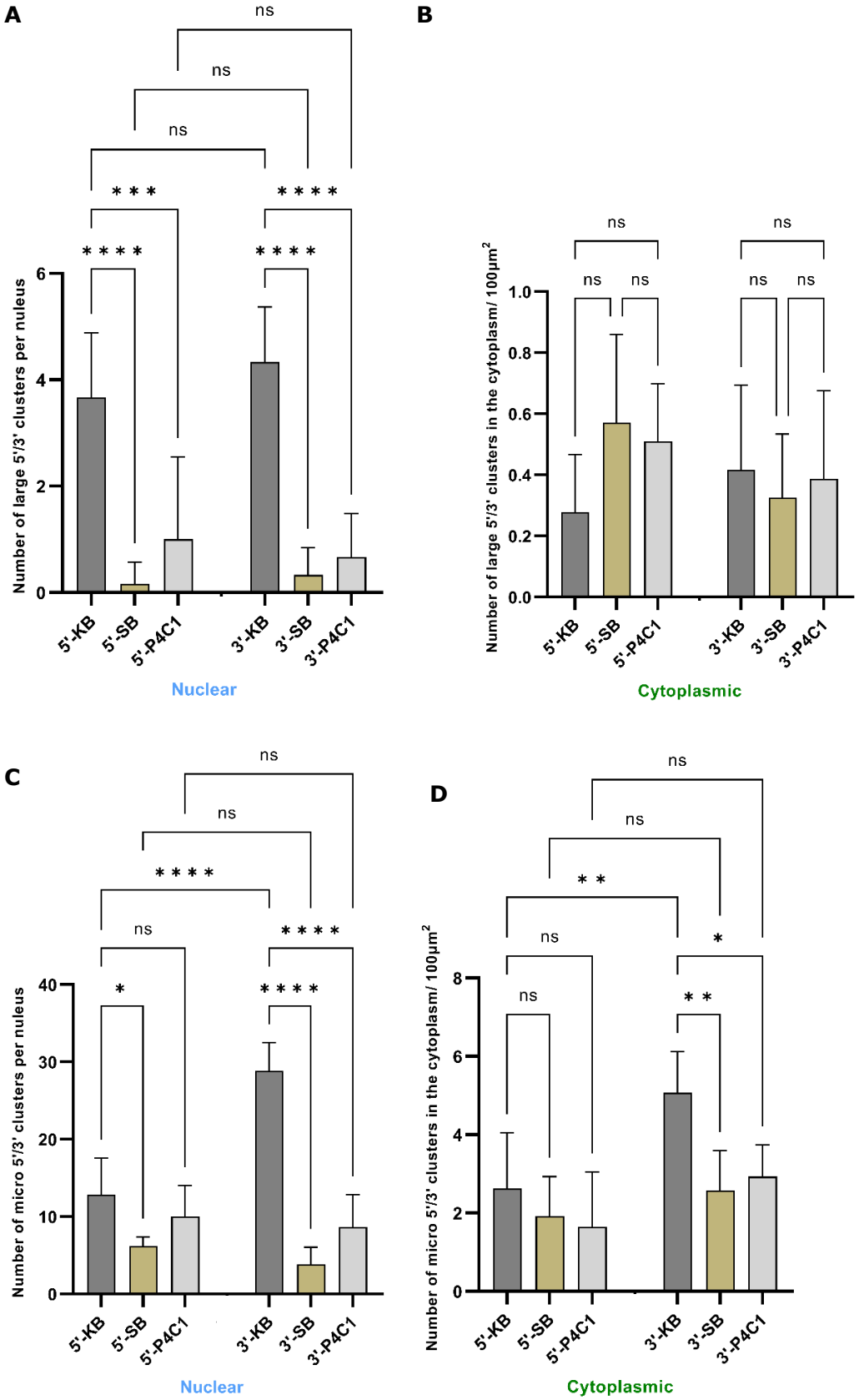
D $(CUG)^{exp}$ signals from the 3' + $(CUG)^{exp}$ dual-labelling experiment



The Diagrams show the number of localisations in each cluster identified with the $(CAG)_8$ probes or the 3' probe set in the nucleus and in the cytoplasm. The boxes represent the interquartile range (IQR) of the data, encompassing the middle 50% of observations. The whiskers, depicted as lines extending from the boxes, indicate the minimum and maximum values among all data points. $(CUG)^{exp}$ clusters containing >250 localisations are considered as large $(CUG)^{exp}$ clusters; $(CUG)^{exp}$ clusters containing 11-250 localisations are considered as micro $(CUG)^{exp}$ clusters/foci; $(CUG)^{exp}$ clusters containing $\leqslant$ 10 localisations are considered as non-$(CUG)^{exp}$ or off-targets. 3' end clusters containing >80

localisations are considered as large 3’ end clusters; 3’ end clusters containing 8-80 localisations are considered as micro 3’ end clusters; 3’ end clusters containing <8 localisations are considered as transcript fragments or background.

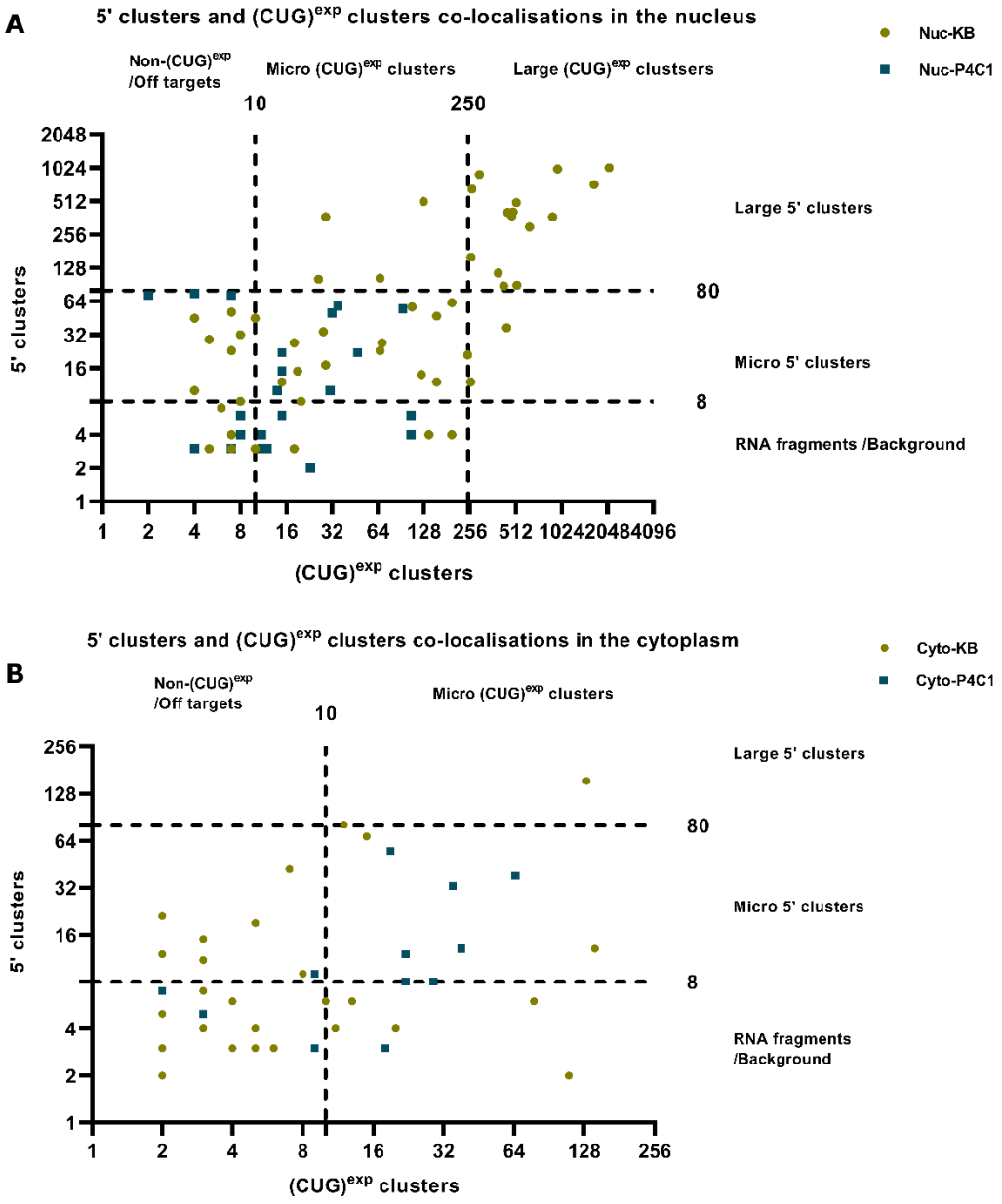
Figure 4-9 Quantification of the number of clusters identified with 5′ or 3′ probe sets in KB, SB and P4C1 (n=6).

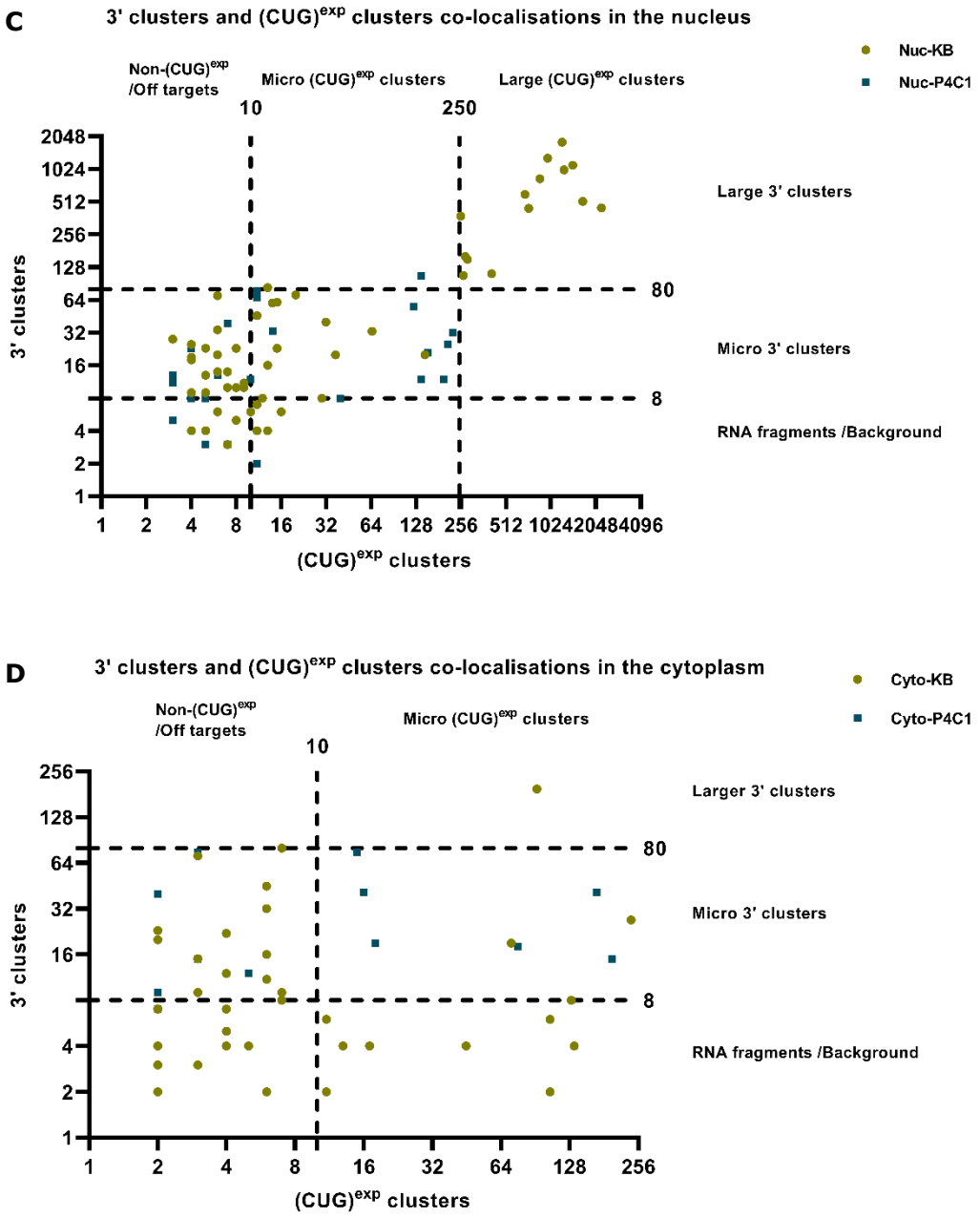


A. Histograms show the mean number of large clusters with either 5’ or 3’ probe sets per nucleus. B. Histograms show the mean number of large 5’ or 3’ end clusters per 100μm^2 cytoplasmic area. C. Histograms show the mean number of micro 5’ or 3’ end clusters per nucleus. D. Histograms show the mean number of micro 5’ or 3’ end clusters per 100μm^2 cytoplasmic area. n indicates the total number of cells analysed. Statistical significance was determined by two-way ANOVA (mean ± SD, * for P<0.05, ** for P<0.01, *** P<0.001, **** for P<0.0001) or indicated as not significant (NS) for P>0.05.

Next, I set out to determine the integrity of repeat expansion *DMPK* transcripts. I examined the extent to which the (CUG)^{exp} foci colocalised with the 5’ or 3’ end clusters. The results, presented in Figure 4-7, Figure 4-8 and Figure 4-10 demonstrated that 78.9% of the large (CUG)^{exp} foci co-localised with large 5’ end clusters, and 87.5% of the large (CUG)^{exp} foci co-localised with large 3’ end clusters in the DM1 nuclei. As large 5’ or 3’ end clusters are known to contain multiple *DMPK* transcripts, its suggests that large (CUG)^{exp} foci are typically formed by multiple mutant *DMPK* transcripts. Furthermore, the co-localisation analysis showed that the majority of micro (CUG)^{exp} foci either colocalised with micro 5’ or 3’ end clusters or did not colocalise with any genuine micro 5’ or 3’ end clusters. These findings suggest that micro (CUG)^{exp} foci consist of mutant *DMPK* transcripts which are in variable degradation states. From the most to the least intact state, a micro (CUG)^{exp} focus could be formed by a complete mutant *DMPK* transcript, an incomplete mutant *DMPK* transcripts with part of the 5’ end and/or 3’ end flanking regions, or the CUG repeats only. Therefore, I propose that micro (CUG)^{exp} foci are the intermediate degradation products of large (CUG)^{exp} foci.

Figure 4-10 Detection of (CUG)^{exp} cluster and 5′ or 3′ end cluster co-localisations.





Dots and squares indicate co-localisation events in all the KB and P4C1 cells analysed, respectively. X axis indicates the number of localisations in each colocalised $(CUG)^{exp}$ cluster; Y axis indicates the number of localisations in each colocalised 5'/3' cluster.

In MBNL-deficient DM1 (P4C1) cells, similar to SB, most of the nuclear *DMPK* transcripts were present singly (Figure 4-7) (Figure 4-8), with no significant difference found in both the number of large 5' or 3' end clusters and the number of micro 5' or 3' end

clusters in the SB and P4C1 nucleus. Notably, the number of large 5′ or 3′ end clusters was dramatically reduced in the P4C1 nucleus compared to that in the KB nucleus. However, the number of micro 5′ or 3′ end clusters did not significantly increase but remained similar or decreased in the P4C1 nucleus compared to that in the DM1 nucleus (Figure 4-9). My previous findings have suggested that the absence of MBNLs modulates the breakdown of large $(CUG)^{exp}$ foci into micro $(CUG)^{exp}$ foci, which are mainly formed by single *DMPK* transcripts and represent intermediate degradation products of mutant *DMPK* transcripts. Based on these findings, my results suggest that MBNL proteins may hold multiple *DMPK* transcripts together to form large foci, and their absence not only prevents the aggregation of multiple $(CUG)^{exp}$ *DMPK* transcripts, but also promotes their degradation and/or their transportation to the cytoplasm.

4.5 Discussion and conclusions

My analysis of micro 5' or 3' end clusters, which were likely derived from single *DMPK* transcripts, revealed a significant enrichment of 3' end clusters over 5' end clusters in KB cells, but no significant difference were observed in SB or P4C1 cells (Figure 4-9). This suggests that the 5′ → 3′ degradation of $(CUG)^{exp}$ *DMPK* mRNA proceeds more efficiently than the 3′ → 5′ degradation. Moreover, the 5′ → 3′ mRNA decay processivity may be limited by MBNL proteins sequestrated to the $(CUG)^{exp}$ *DMPK* mRNA, resulting in stabilisation of the 3′ end of $(CUG)^{exp}$ *DMPK* transcripts. In the absence of MBNLs, degradation of an entire $(CUG)^{exp}$ *DMPK* transcript appears to be unimpeded. I also observed that the large 5′ or 3′ end clusters derived from multiple *DMPK* transcripts showed no significant difference in the number of 5′ end cluster

compared to 3’ end clusters in KB nuclei. This suggests that the aggregation of multiple $(CUG)^{exp}$ *DMPK* transcripts into large RNA foci has a similar level of impact in preventing 5’ → 3’ mRNA decay and 3’ → 5’ mRNA decay.

Previous studies have shown that mRNAs are concentrated in cytoplasmic ribonucleoprotein (RNP) granules known as processing bodies (P bodies) or GW bodies where they can be either recycled for translation or degraded by mRNA decay components. In the cytoplasm of DM1, non-DM1 control and MBNL-deficient DM1 cells large 5’ or 3’ end clusters were detected (Figure 4-7) (Figure 4-8). However, there was no significant difference in the number of large cytoplasmic 5’ or 3’ end clusters amongst these cell types (Figure 4-9). Based on these findings, I hypothesise that these clusters are primarily formed by multiple normal *DMPK* mRNAs stored in P bodies. However, further analyses are required to verify my hypothesis.

Additionally, the co-localisation analysis provided evidence that micro $(CUG)^{exp}$ foci flanked by the 5’ or 3’ end were also present in the cytoplasm (Figure 4-10), which indicates that mutant *DMPK* transcripts may export to the cytoplasm when not clustered which could provide a source for RAN translation in DM1.

Using STORM technology, I were able to overcome the optical diffraction limit to gain unprecedented insights into micro $(CUG)^{exp}$ RNA foci in DM1 cells and MBNL-deficient DM1 cells. Furthermore, I used probe sets complementary to the 5' and 3' ends of the *DMPK* transcript to investigate the integrity of the transcript. The results revealed that large RNA foci are typically composed of multiple *DMPK* transcripts, while micro foci only contain single *DMPK* transcripts. It was also found that large RNA foci are unable to leave the nucleus, while micro foci can migrate to the cytoplasm,

potentially contributing to the origin of RAN translation in DM1. Moreover, my study builds upon and expands the previous notion that the depletion of MBNL1 leads to elimination of RNA foci [63, 107]. I demonstrated that the absence of MBNLs modulates the breakdown of mutant *DMPK* mRNA. In DM1, mutant *DMPK* mRNA aggregates into large, compact RNA foci that can be easily detected under conventional fluorescent microscope. However, in the absence of MBNLs, mutant *DMPK* mRNA can only form into micro RNA foci, which are less stable than large foci. This thereby promotes the degradation of mutant *DMPK* mRNA or its transportation to the cytoplasm.

Overall, my study deepens our understanding of the mechanisms underlying RNA foci dynamics and the role of MBNL proteins in DM1 pathogenesis. Moreover, the use of STORM technology provides a valuable tool for investigating RNA foci in other diseases triggered by short tandem repeat (STR) mutations.

Chapter 5 Identification of RNA decay factors involved in the degradation of $(CUG)^{exp}$ *DMPK* transcripts

5.1 Abstract

The pathogenesis of DM1 is primarily associated with the nuclear retention of CUG expansion DMPK mRNA. However, our understanding of the degradation mechanism and the enzymes involved remains limited. In light of this knowledge gap, I hypothesise that investigating the degradation mechanism of mutant DMPK mRNA and identifying the specific enzymes responsible will provide valuable insights.

To address this, I established the following research aims:

1. Identify RNA decay factors involved in the degradation of mutant DMPK mRNA.
2. Unravel the role of MBNL proteins in the mutant DMPK mRNA decay.

I examined the participation of various RNA decay factors in DM1 cells and their derivatives through shRNA lentiviral transduction. The successful knock-down of XRN2, EXOSC10, UPF1 and STAU1 produced a slight but significant increase in the foci area and the abundance of mutant DMPK mRNA in DM1 cells. Notably, a more pronounced increase was observed in MBNL-deficient DM1 cells. These findings highlight the critical role played by various RNA decay pathways in reducing the accumulation of mutant DMPK RNA. Furthermore, the sequestration of MBNLs to the mutant RNA may impede its degradation process, as the absence of MBNLs appears to enhance the efficiency of these pathways. Studying the degradation mechanism of mutant DMPK mRNA in myotonic dystrophy may also shed light on RNA degradation mechanisms in other diseases caused by short tandem repeat (STR) mutations.

5.2 Potential RNA decay factors involved in the degradation of $(CUG)^{exp}$ DMPK transcripts

It is widely reported that the fundamental mechanism underlying myotonic dystrophy pathophysiology is the sequestration of MBNL proteins by repeat expansion *DMPK* transcripts. The consensus view is that mutant transcripts sequester MBNL1 and 2, but my preliminary data indicate that it is MBNL1 and 2 that hold or retain the mutant RNA in the nucleus. Although several therapeutic approaches [126, 173, 293, 294] have appeared to eliminate nuclear foci when visualised using in situ hybridisation, they have not successfully reversed the entire pathogenesis of DM1 as it is not clear what happens to the mutant RNA transcripts after the dissolution/disappearance of nuclear foci. It is important to keep in mind that there is no evidence showing that the foci are the sites of mutant RNA decay. Thus, having a better understanding of the metabolism of mutant *DMPK* mRNA and the effect of $(CUG)^{exp}$ RNA -MBNL protein complex on mutant *DMPK* mRNA decay could have important therapeutic implications for DM.

It is well-established that most mRNAs, presumably including the wild type *DMPK* mRNA, go through deadenylation-dependent decay in the cytoplasm [232, 233]. However, due to the presence of long expanded CUG repeats in the 3' UTR and association with MBNLs, the mutant *DMPK* mRNA has prolonged nuclear residence and is therefore initially subject to nuclear RNA surveillance [225, 227]. Nuclear RNA decay is primarily conducted by exonuclease XRN2 in the 5' → 3' direction [295, 296] or/and by the nuclear exosome (with catalytic subunit EXOSC10/RRP6) in the 3' → 5' direction [215, 297-299]. Although the mutant *DMPK* transcripts are substantially nuclear retained, my studies provide the evidence that a notable portion of the mutant *DMPK* transcripts

transport to the cytoplasm. Moreover, studies by Garcia et al. (2014) also suggest that nonsense-mediated mRNA decay (NMD) pathway which principally operates in the cytoplasm, plays a conserved role in regulating CUG repeat RNA transcript levels and toxicity [250]. For example, inactivating UPF1 protein by RNAi in human DM1 patient fibroblasts causes an increase in the number of toxic RNA foci [250]. Therefore, NMD could be directly involved in the mutant *DMPK* RNA decay as the presence of expanded CUG repeats in *DMPK* mRNAs may trigger NMD due to the extended distance between the stop codon and the poly(A) signal [204, 205, 250]. Additionally, the RNA hairpin structures formed by the expanded CUG repeats in the 3'UTR of *DMPK* mRNA are very likely to be the binding sites of STAU proteins [248, 249]. Therefore, mutant *DMPK* mRNA may be subject to Staufen-mediated mRNA decay (SMD) as well. Taken together, I hypothesised that several decay pathways might play a role in the mutant *DMPK* RNA degradation.

5.3 Poly(A) tail and 5' cap structure status of the $(CUG)^{exp}$ *DMPK* transcripts

Hyperadenylation of RNA is known to be associated with nuclear retention, but the cause-consequence relationship between hyperadenylation and regulation of RNA nuclear export is still unclear [300-302]. Additionally, absence of the 7-methylguanosine (m^7G) cap at the 5' end of pre-mRNA could cause the 5'->3' degradation of mRNA by XRN exonucleases [196, 210, 303]. To compare both the 3' polyadenylation and 5' cap structure status of *DMPK* transcripts in the non-DM cells (SB, control) and DM1 patient cells (KB) and their effect on presence and stabilisation of mutant *DMPK* transcripts, I subjected RNA from KB and SB to poly(A) selection using oligo(dT) beads and 5' cap selection using 5'-Phosphate-Dependent Exonuclease. 5'-

Phosphate-Dependent Exonuclease, a processive 5'->3' exonuclease digests RNA having a 5'-monophosphate. The enzyme does not digest RNA having a 5'-triphosphate, a 5'-cap or a 5'-hydroxyl group. Thus, it can be used to isolate mature mRNA with an m^7G-cap.

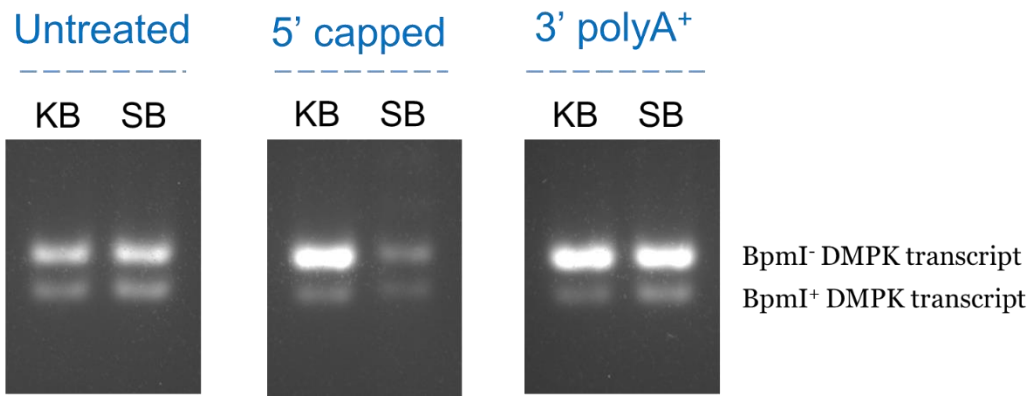
Both KB and SB are heterozygous for an rs527221 G>C single nucleotide polymorphism (SNP) located in exon 10 of *DMPK*. Moreover, A BpmI recognition site is associated with this SNP, which I refer to as the BpmI polymorphism in this thesis (Figure 3-5 A). In KB, *DMPK* transcripts with or without the BpmI recognition site can be used to distinguish the $(CUG)^{exp}$ and normal *DMPK* transcripts. To determine the abundance of BpmI+ *DMPK* transcripts in RNA processed by Poly(A) selection or 5' cap selection, I performed RT-PCR with *DMPK* primers flanking the BpmI polymorphism site followed by either BpmI digestion assay or digital PCR (see Section 3.7 for a detailed introduction to both methods).

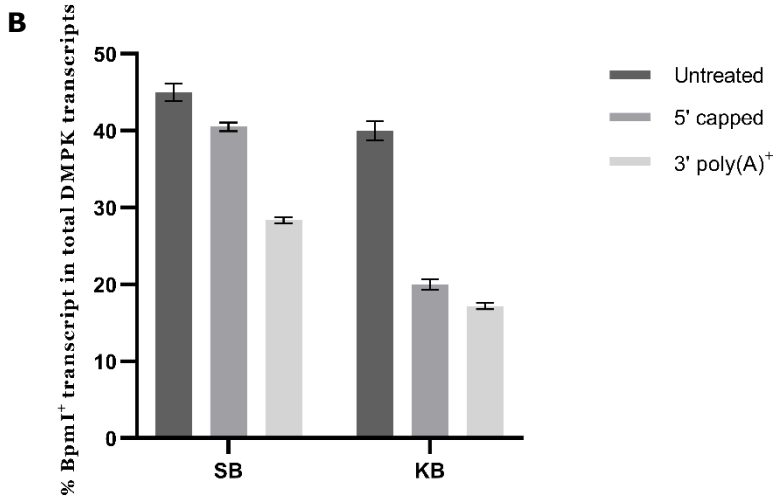
My results showed that in KB, there was less $BpmI^+$ *DMPK* transcripts recovered from the poly(A) selection or 5' cap selection, compared to SB (Figure 5-1, Figure 0-8) (Figure 5-2). However, the BpmI digestion method might underestimate the abundance of $BpmI^+$ transcripts (see Section 3.7.3 for a detailed discussion), so I used the results obtained from digital PCR method for interpretation. In SB, the poly(A) selection and 5' cap selection recovered 89.0% and 99.0% of the $BpmI^+$ *DMPK* transcripts, respectively, compared to untreated $BpmI^+$ *DMPK* transcripts. In contrast, in KB, only 62.9% and 68.7 % of the $BpmI^+$ (mutant) *DMPK* transcripts were recovered from the Poly(A) selection and 5' cap selection, respectively, compared to untreated $BpmI^+$ *DMPK* transcripts (Figure 5-2). One possible explanation for this observation would be that a significant portion of the mutant

DMPK mRNA was subject to deadenylation or decapping. My previous results indicate that the mutant *DMPK* mRNA is primarily retained in the nucleus, which could lead to its recognition as abnormal or faulty RNA, triggering RNA decay (Figure 3-8) [225, 226]. However, due to the complex hairpin structures in the CUG repeats and sequestration of MBNL (Figure 1-6, Figure 1-9), degradation may be initiated but halted to a large extent. Another explanation might be that a considerable portion of the long CUG expanded *DMPK* mRNA fractured during experimental handling and only part of a transcript harbouring a poly(A)-tail or a m^7G-cap was captured by the beads and RNA bind columns. Moreover, influence of the expanded CUG repeats on oligo(dT) binding avidity, caused by topological constraints in the expanded transcripts, could not be excluded at this point.

Figure 5-1 Assessment of poly(A) tail and 5′ cap structure status of the $(CUG)^{exp}$ DMPK transcripts via the BpmI digestion method.

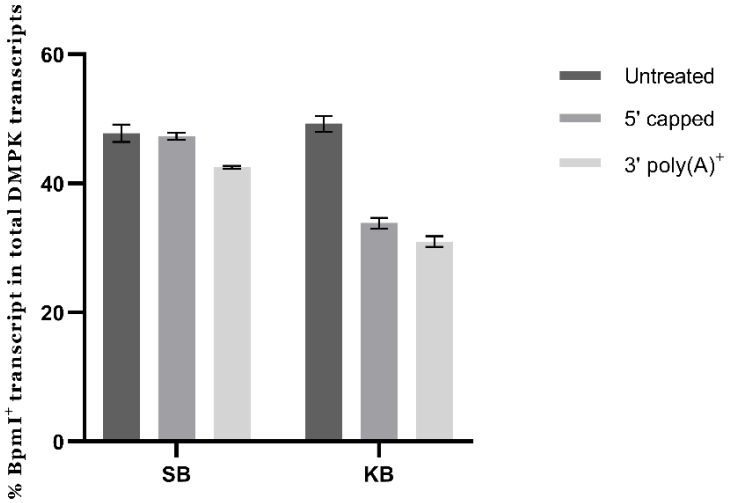
A





(A) Ethidium bromide-stained gels showing RT–PCR products from 3′ poly(A)$^+$, 5′ capped or total RNA in KB and SB cells. A fragment of *DMPK* mRNA was reverse-transcribed, amplified and digested with the BpmI restriction enzyme, and the resulting fragments were separated by gel electrophoresis (BpmI$^-$, PCR band representing *DMPK* transcripts without the BpmI recognition site and BpmI$^+$, representing with the BpmI recognition site). The experiment was performed in duplicate to ensure the reproducibility of the results. (B) Quantification of signals in (A) for the percentage of BpmI$^+$ *DMPK* transcripts in total *DMPK* transcripts in KB and SB. The BpmI$^+$ *DMPK* transcript represents the mutant *DMPK* transcript in KB.

Figure 5-2 Assessment of poly(A) tail and 5′ cap structure status of the (CUG)^{exp} DMPK transcripts via the dPCR method.

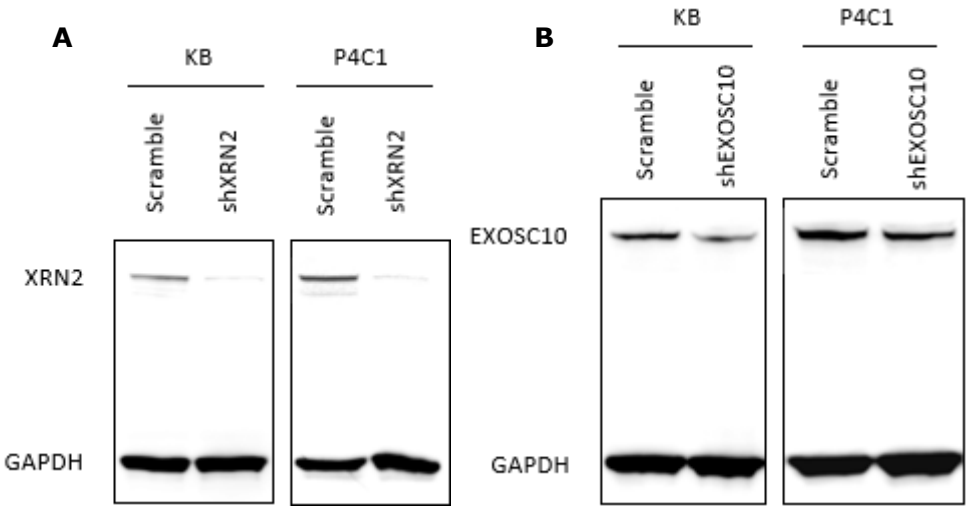


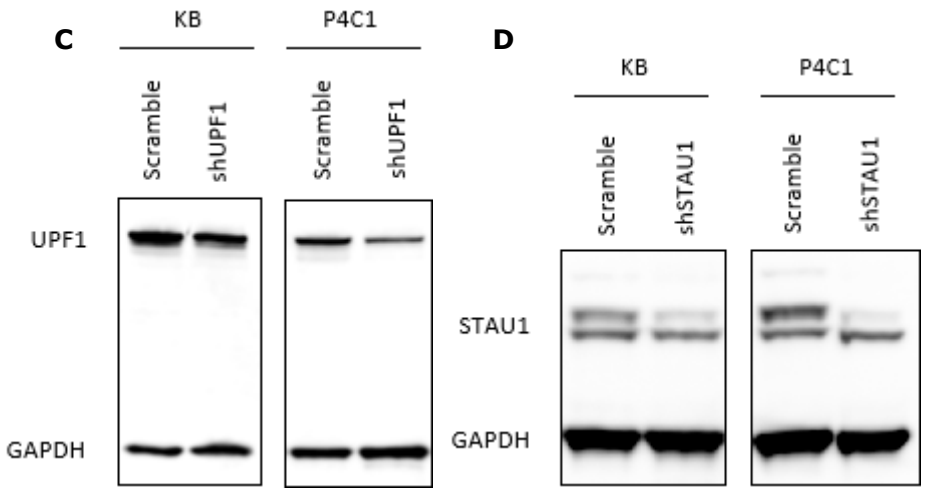
Quantification of the percentage of BpmI^{+} *DMPK* transcripts in total *DMPK* transcripts in KB and SB, as determined by dPCR with allele-specific probes. The BpmI^{+} *DMPK* transcript represents the mutant transcript in KB. The experiment was performed in triplicate to ensure the reproducibility of the results.

5.4 Knock-down of RNA decay factors

To investigate the participation of XRN2, EXOSC10, UPF1, and STAU1 in the degradation of (CUG)^{exp} *DMPK* transcripts in DM1 and MBNL-deficient DM1 cells, I individually knocked down the RNAs encoding these RNA decay factors using lentiviral shRNA-mediated knock-down. Western blot analyses were then conducted to assess the knock-down efficiency at the protein level (Figure 5-3, Figure 0-5). The results demonstrated an equivalent reduction of protein levels in both KB and P4C1 cells, with XRN2 reduced by 76.8%, EXOSC10 by 50.0%, UPF1 by 56.5% and STAU1 by 69.9% (Figure 5-4).

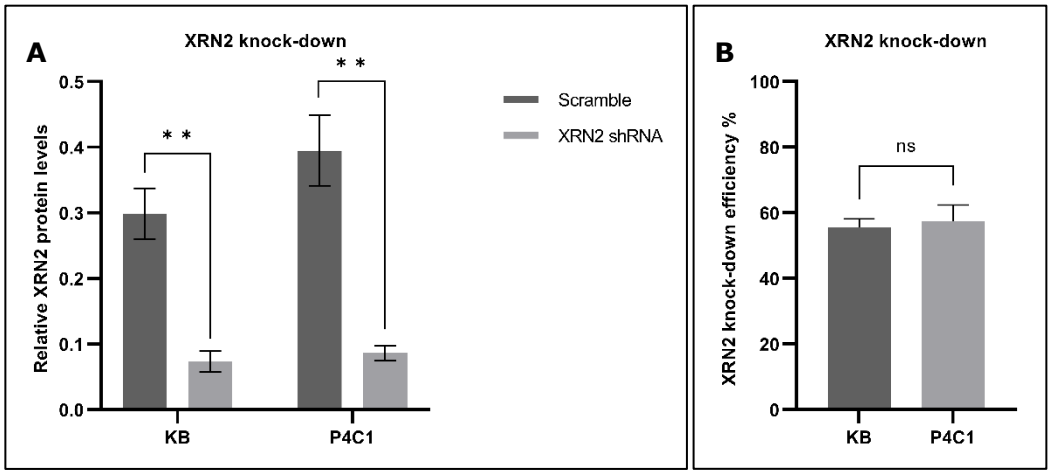
Figure 5-3 Assessment of the RNA decay factor knock-down by western blotting.

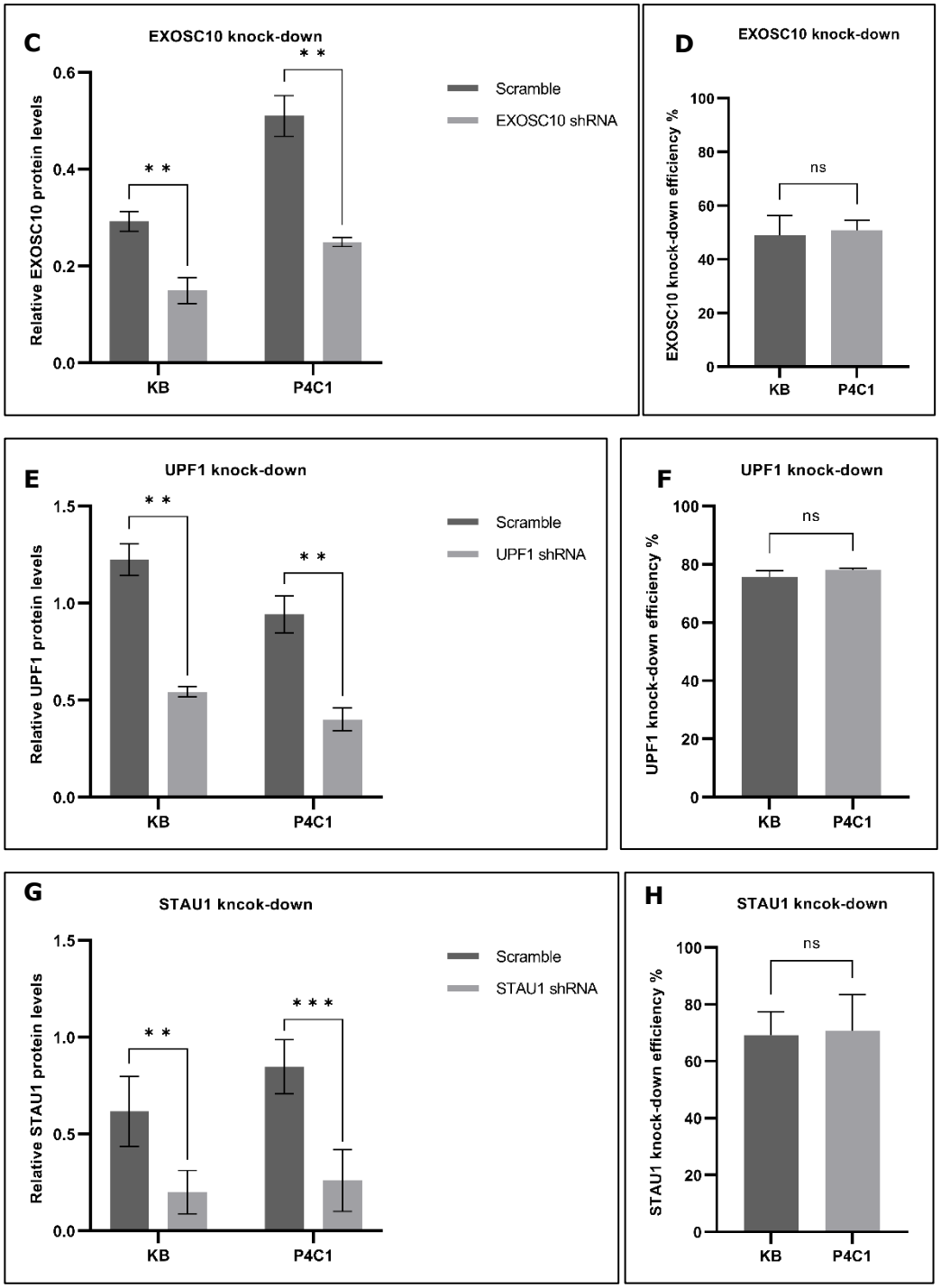




Western blot analyses showing the protein levels in KB (DM1) and P4C1 (MBNL-deficient DM1) cells following the lentiviral transfection with scramble control and shRNAs target *XRN2* (A), *EXOSC10* (B), *UPF1* (C) and *STAU1* (D). The experiment was performed in triplicate to ensure the reproducibility of the results.

Figure 5-4 Quantification of the protein levels and knock-down efficiency.

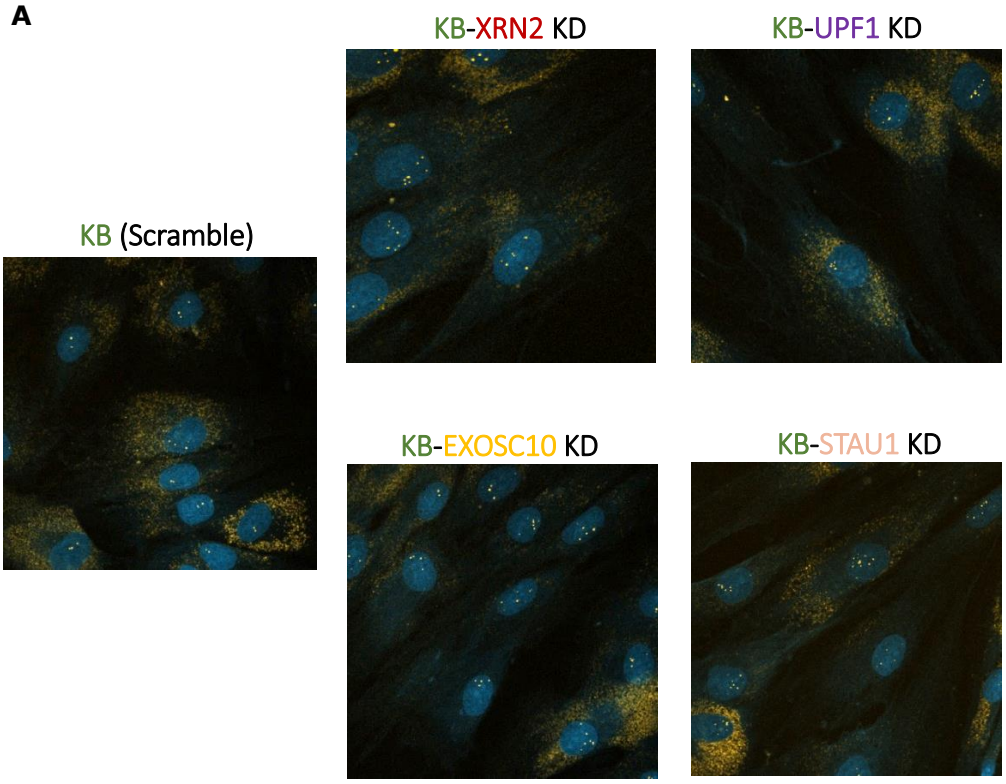


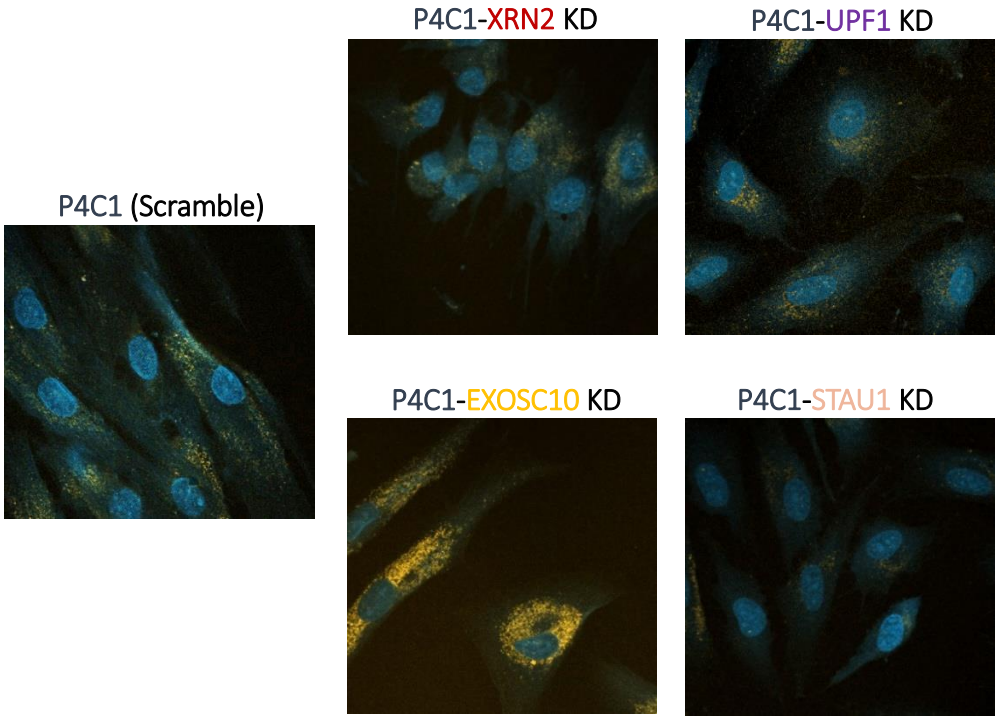


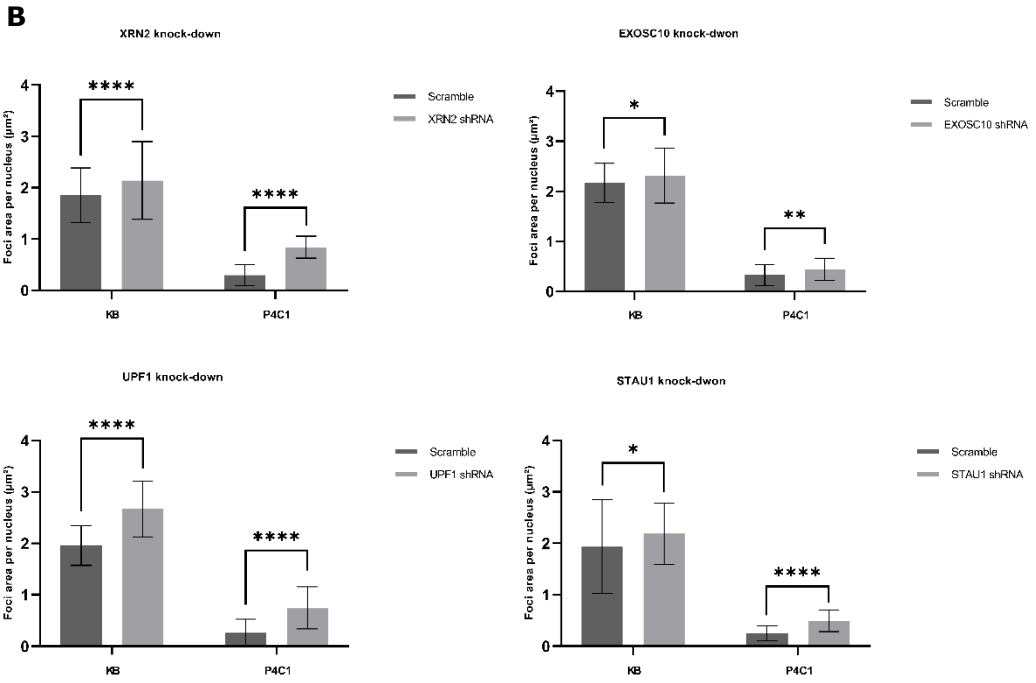
(A, C, E, G) Quantification of the relative protein levels of XRN2 (A), EXOSC10 (C), UPF1 (E) and STAU1 (G) following the lentiviral shRNA-mediated knock-down in KB and P4C1 cells, normalised against GAPDH values. (B, D, F, H) Comparisons of the knock-down efficiency of XRN2 (B), EXOSC10 (D), UPF1 (F) and STAU1 (H) between KB and P4C1. Statistical significance was determined by two-way ANOVA or two-tailed t-test (mean ± SD, n=3, * for $P<0.05$, ** for $P<0.01$, *** $P<0.001$, **** for $P<0.0001$) or indicated as not significant (NS) for $P>0.05$.

The individual knock-down of *XRN2*, *EXOSC10*, *UPF1* and *STAU1* all produced a slight but significant increase in foci area in KB and an even greater rise to that observed in P4C1 (Figure 5-5). It indicates that they all play a key role in reducing RNA foci formation and in the absence of MBNLs, they act more effectively.

Figure 5-5 Effect of the knock-down (KD) of decay factors on RNA foci.







(A) Images of KB and P4C1 cells following the lentiviral shRNA knock-down of *XRN2*, *EXOSC10*, *UPF1* and *STAU1* or transfection with corresponding scramble shRNA lentiviral particles and RNA FISH using Cy3-$(CAG)_{10}$ probe (amber) for the localisation of CUG

repeats in the mutant *DMPK* mRNA and Hoechst (blue) to indicate the cell nuclei. (B) Quantification of RNA foci area per nucleus with data compiled from the RNA FISH in above experimental cells. Statistical significance was determined by two-way ANOVA (mean ± SD, * for P<0.05, ** for P<0.01, *** P<0.001, **** for P<0.0001).

To investigate the impact of RNA decay factors knock-down on $(CUG)^{exp}$ *DMPK* (expDMPK) transcripts, nuclear and cytoplasmic RNA fractions were isolated separately using Norgen's Cytoplasmic & Nuclear RNA Purification Kit. On-column DNase digestion was performed in nuclear fractions only to eliminate contaminating genomic DNA. The quality of the RNA extracted from both fractions was assessed using Agilent Bioanalyzer (Figure 0-7). Subsequently, I utilised probe digital PCR assays to measure the abundance of exp*DMPK* transcripts. The probe was designed to bind to the exon 10 of *DMPK* transcripts, which is located within the 5' end of the CUG repeat region (Figure 5-10). For information on the probe design, see Section 3.7.2.

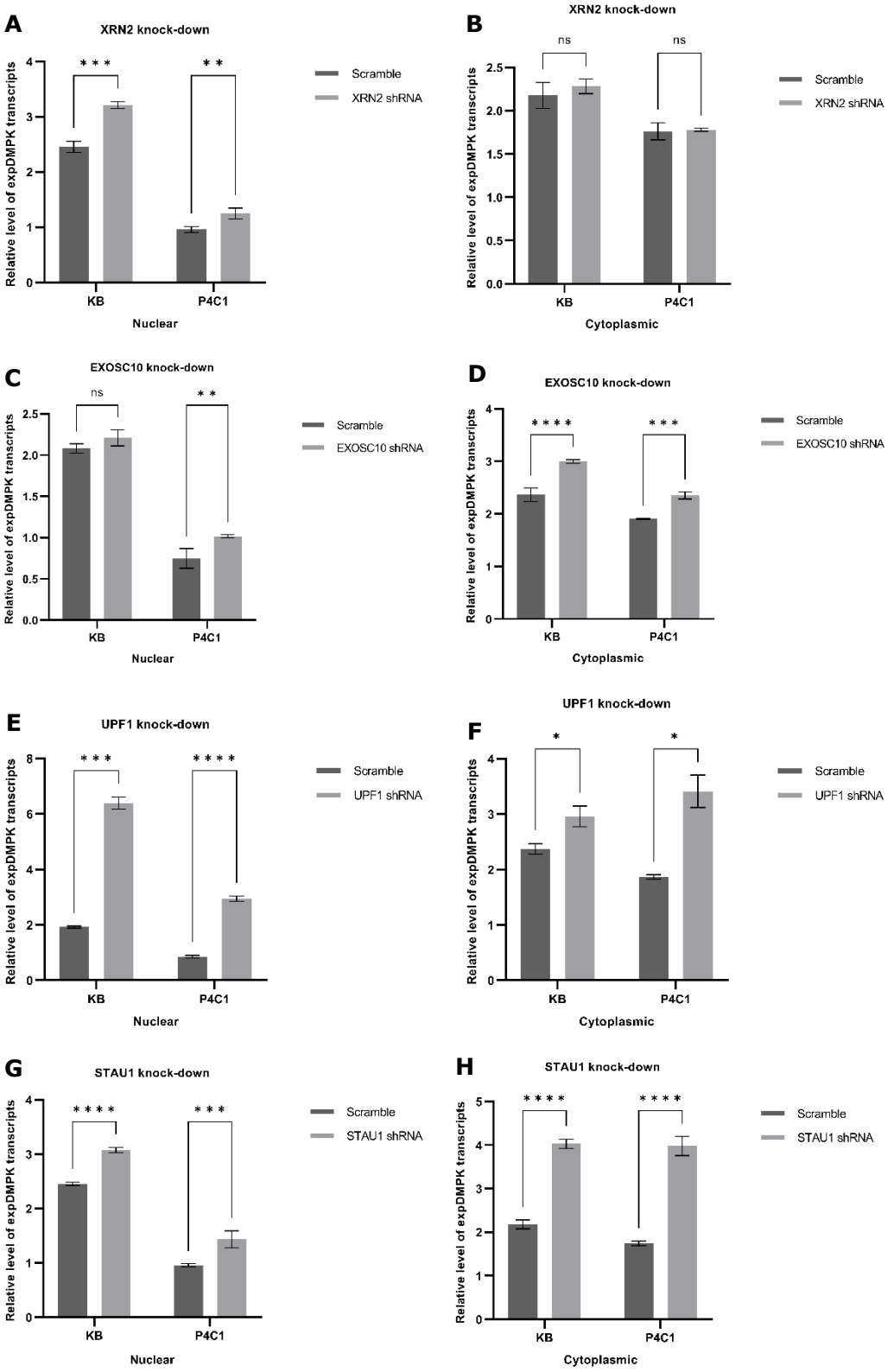
My previous results demonstrated that the 3' end of *DMPK* transcripts was more abundant than the 5' end in KB cells, while no significant difference was observed in SB or P4C1 cells (Figure 4-9). This suggests that degradation of the 3' end of exp*DMPK* transcripts may be impeded, resulting in its stabilisation. Knocking down *XRN2* resulted in a significant and equivalent degree of increase in the abundance of the 5' end of exp*DMPK* transcripts in KB and P4C1 nuclei (Figure 5-6 A), indicating that the activity of XRN2 is equally effective in degrading the 5' end of exp*DMPK* transcripts in both DM1 and MBNL-deficient DM1 cells. Together with previous findings, these results suggest that the 5' → 3' exoribonuclease activity of XRN2 may be stalled when encountering MBNL proteins at the CUG repeat region. In contrast

to the results of *XRN2* knock-down, knocking down *EXOSC10* had no effect on the abundance of the 5’ end of exp*DMPK* transcripts in KB nuclei, but it led to a significant increase in P4C1 nuclei (Figure 5-6 C). This provides direct evidence that the 3’ → 5’ exoribonuclease activity of EXOSC10 in degrading the exp*DMPK* transcripts may also be limited by MBNL proteins.

UPF1 mainly operates on the 3’UTRs of mRNAs that are directed for NMD in the cytoplasm, but recent studies offered evidence that UPF1 also facilitates the release of mRNAs from transcription sites and their export from the nucleus to the cytoplasm [244]. The knock-down of UPF1 produced a significant increase in the abundance of nuclear exp*DMPK* transcripts in both KB and P4C1 cells (Figure 5-6 E). Moreover, it also resulted in an increase in the abundance of cytoplasmic exp*DMPK* transcripts in KB and an even significantly greater increase in P4C1 cells (Figure 5-6 F). In addition to its role in Staufen-mediated decay (SMD), STAU1 has also been reported to promote the nuclear export of specific mRNAs, including $(CUG)^{exp}$ *DMPK* mRNA [249]. Similar to the results of UPF1 knock-down, knocking down *STAU1* led to a significant increase in the abundance of both nuclear and cytoplasmic exp*DMPK* transcripts in KB and an even greater increase in P4C1 (Figure 5-6 G and H).

These findings suggested that UPF1 and STAU1 are key factors in alleviating nuclear retention of $(CUG)^{exp}$ *DMPK* transcripts and facilitating their degradation in the cytoplasm.

Figure 5-6 Effect of the knock-down of RNA decay factors on the mutant DMPK transcripts.



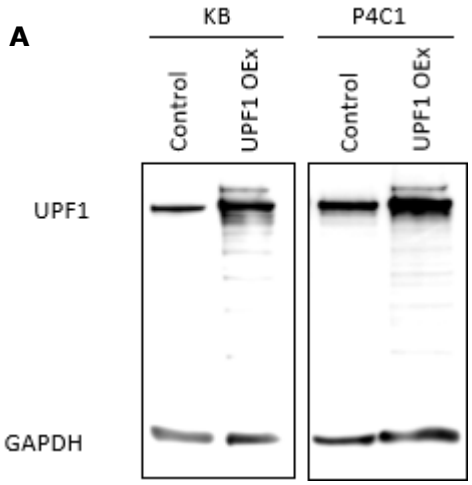
Quantification of the relative level of nuclear and cytoplasmic $(CUG)^{exp}$ *DMPK* transcripts in KB and P4C1 following the lentiviral

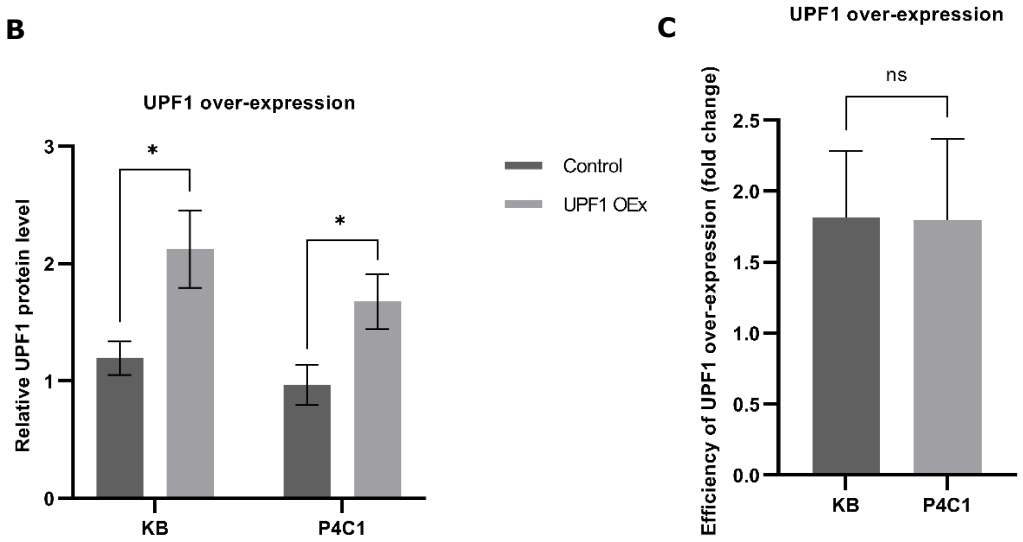
shRNA knock-down of *XRN2* (A-B), *EXOSC10* (C-D), *UPF1* (E-F) and *STAU1* (G-H) or transfection with corresponding scramble shRNA lentiviral particles by measuring the abundance of 5' end to the CUG repeats, normalised against the level of MALAT transcripts for the nuclear fraction and GAPDH transcripts for the cytoplasmic fraction. The experiment was performed in triplicate to ensure the reproducibility of the results. Statistical significance was determined by two-way ANOVA (mean ± SD, n = 3, * for $P<0.05$, ** for $P<0.01$, *** for $P<0.001$, **** for $P<0.0001$) or indicated as not significant (NS) for $P>0.05$.

5.5 UPF1 over-expression

I also successfully over-expressed UPF1 in KB and P4C1 through lentiviral gene transfer. To evaluate the effectiveness of UPF1 overexpression, I employed western blot analysis (Figure 5-7 A, Figure 0-6). The results showed that the level of UPF1 protein in infected cells was twice as high as in the control cells (Figure 5-7 B). The level of UPF1 increase was similar in both KB and P4C1 cells (Figure 5-7 C).

Figure 5-7 Assessment of the UPF1 over-expression by western blotting.

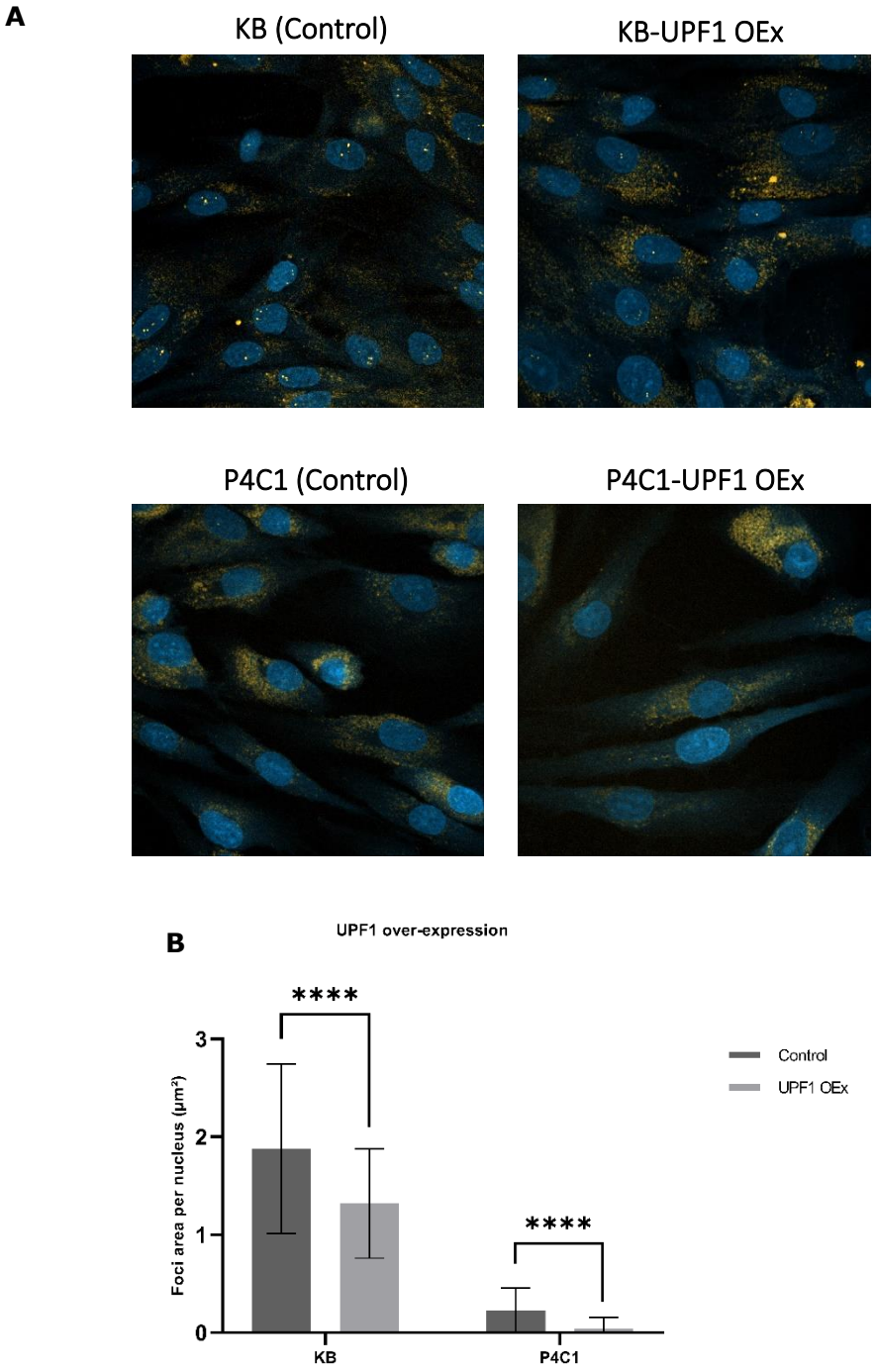




(A) Western blot analyses showing the UPF1 protein levels in KB and P4C1 following the lentiviral transfection with empty control or *UPF1* cDNA. The experiment was performed in triplicate to ensure the reproducibility of the results. (B) Quantification of the relative UPF1 protein levels, normalised against GAPDH values. (C) Comparison of the UPF1 over-expression efficiency between KB and P4C1. Statistical significance was determined by two-way ANOVA or two-tailed t-test (mean ± SD, n=3, * for $P<0.05$, ** for $P<0.01$, *** $P<0.001$, **** for $P<0.0001$) or indicated as not significant (NS) for $P>0.05$.

Followed by RNA in situ hybridisation, the results showed over-expression of UPF1 produced a significant reduction in the foci area in KB and a greater reduction in that in P4C1 (Figure 5-8).

Figure 5-8 Effect of UPF1 over-expression (OEx) on RNA foci.



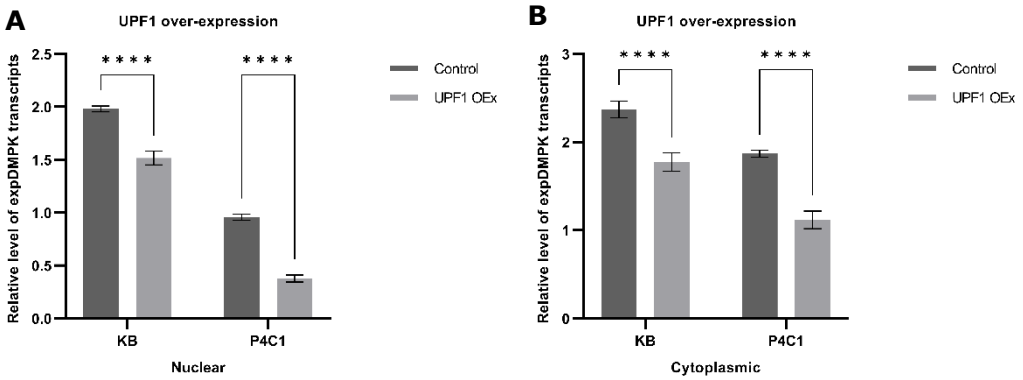
(A) Images of KB and P4C1 cells following the lentiviral transfection with empty control or *UPF1* cDNA and RNA FISH using Cy3-$(CAG)_{10}$ probe (amber) for the localisation of CUG repeats in the mutant *DMPK* mRNA and Hoechst (blue) to indicate the cell nuclei. (B) Quantification of RNA foci area per nucleus with data compiled from the RNA FISH in above experimental cells.

Statistical significance was determined by two-way ANOVA (mean ± SD, * for P<0.05, ** for P<0.01, *** P<0.001, **** for P<0.0001).

To investigate the impact of UPF1 over-expression on $(CUG)^{exp}$ *DMPK* (exp*DMPK*) transcripts, nuclear and cytoplasmic RNA fractions were isolated separately using Norgen's Cytoplasmic & Nuclear RNA Purification Kit. On-column DNase digestion was performed in nuclear fractions only to eliminate contaminating genomic DNA. The quality of the RNA extracted from both fractions was assessed using Agilent Bioanalyzer. Subsequently, I utilised probe digital PCR assays to measure the abundance of exp*DMPK* transcripts (Figure 0-7). The probe was designed to bind to the exon 10 of *DMPK* transcripts, which is located 5' to the CUG repeat region (Figure 5-10). For information on the probe design, see Section 3.7.2.

Over-expressing UPF1 resulted in a dramatic decrease in the abundance of both nuclear and cytoplasmic exp*DMPK* transcripts relative to the wild-type *DMPK* transcripts in KB with an even greater decrease in P4C1 cells (Figure 5-9), suggesting that UPF1 over-expression has a more pronounced effect on the nuclear export and cytoplasmic degradation of exp*DMPK* transcripts in MBNL-deficient DM1 cells compared to DM1 cells. This finding implies that the sequestration of MBNL protein to the mutant *DMPK* mRNA disrupts the regulation of its metabolism by causing an increase in RNA stability or hindering its decay. This disruption could contribute to the pathogenesis of DM1.

Figure 5-9 Effect of the UPF1 over-expression (OEx) on the mutant DMPK transcripts.



Quantification of the relative level of nuclear (A) and cytoplasmic (B) $(CUG)^{exp}$ *DMPK* transcripts in KB and P4C1 by measuring the abundance of 5' end to the CUG repeats, normalised against the level of MALAT transcripts for the nuclear fraction and GAPDH transcripts for the cytoplasmic fraction. The experiment was performed in triplicate to ensure the reproducibility of the results. Statistical significance was determined by two-way ANOVA (mean ± SD, n = 3, **** for $P<0.0001$).

Therefore, the observation that over-expression of UPF1 reduces the size of RNA foci and the accumulation of exp*DMPK* transcripts in DM1 cells provides support for the idea that UPF1 may be a promising target for developing therapies for DM1. However, more research would be needed to confirm this and to determine the optimal level of UPF1 expression and phosphorylation for therapeutic purposes.

5.6 Discussion and conclusions

Although both XRN2 and EXOSC10 are involved in the accumulation of RNA foci and the exp*DMPK* mRNA decay, they differ in their specific functions, EXOSC10 is a distributive 3' → 5'exonuclease, which cleaves RNA molecules by removing nucleotides one at a time from the end of the RNA molecule and releases the cleaved nucleotide before cleaving the next one [298, 304]. In contrast, XRN2 is a processive 5' → 3' exonuclease that

removes multiple nucleotides at once while moving along the RNA molecule [305, 306] (Figure 5-10). Therefore, 5’ → 3’ degradation by XRN2 is faster than 3’ → 5’ degradation by EXOSC10. Although both XRN2 and EXOSC10 appear to be stalled at expanded CUG repeat structures by MBNL proteins (Figure 4-9) (Figure 5-6 A, C and D), the stalling of XRN2 is more likely to induce its malfunction due to slow off-rates of a highly processive exoribonuclease [306], resulting in a higher accumulation of the 3' end of the transcripts.

As indicated by previous research that XRN2 is primarily located in the nucleus [222, 305], knock-down of *XRN2* in my study had a significant effect on the abundance of nuclear exp*DMPK* transcripts in KB and P4C1, but had no significant effect on the abundance of cytoplasmic exp*DMPK* transcripts in the same cell lines (Figure 5-6 B). In contrast, knock-down of *EXOSC10* had no significant effect on the abundance of nuclear exp*DMPK* transcripts in KB (Figure 5-6 C), but resulted in a significant increase in the abundance of cytoplasmic exp*DMPK* transcripts in the same cells (Figure 5-6 D). it's important to note that the abundance of exp*DMPK* transcripts was measured based on a BpmI polymorphism site located in exon 10, which is 5' to the repeat expansion region. Based on this measurement, I propose these differences in the effects of knockdown may be attributed to multiple factors. First, unlike XRN2, human EXOSC10 is not restricted to the nucleus and is also present in the cytoplasm [307]; Second, while the 5’ → 3’ degradation by XRN2 can effectively proceed to the 5' end of the repeat expansion region in both KB and P4C1 (Figure 5-6 A), the 3’ → 5’ degradation process by EXOSC10 appears to largely impeded by MBNLs at the CUG repeat region. As a result, degradation by EXOSC10 could not proceed to the 5’ end of the repeat expansion region, unless MBNLs were absent; Third, I previously found that cytoplasmic

exp*DMPK* transcripts are not aggregated like nuclear exp*DMPK* transcripts (Figure 4-7) (Figure 4-8), and a substantial portion of them lack MBNL protein binding (Figure 4-10), therefore allowing for the degradation of cytoplasmic exp*DMPK* transcripts by EXOSC10 to proceed towards the 5’ end. However, measuring the abundance of the 3’ end of exp*DMPK* transcripts pose a challenge since there is no equivalent 3’ polymorphism site that can be used. Therefore, although through the STROM experiments I found that there was a higher accumulation of the 3' end of exp*DMPK* transcripts than the 5’ end in KB, further research is needed to develop a method for measuring the abundance of the 3’ end of exp*DMPK* transcripts and to investigate the effect of *XRN2* and *EXOSC10* knock-down on the abundance of the 3’ end directly.

Overall, although both enzymes are involved in the decay of exp*DMPK* transcripts, the functional differences between XRN2 and EXOSC10 suggest that their stalling by MBNL proteins may lead to distinct effects on the decay of the transcripts.

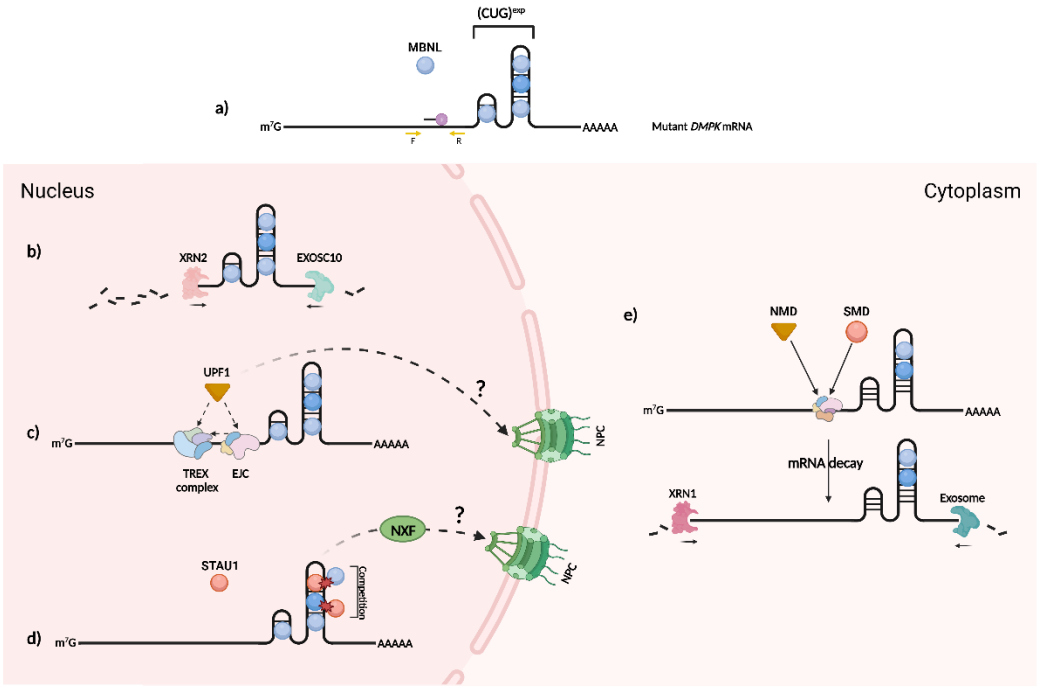
Nonsense-mediated decay (NMD) and Staufen-mediated decay (SMD) are two essential pathways for mRNA quality control that prevent the translation of defective transcripts. The efficiency and specificity of these pathways rely on key factors, including UPF1 for NMD and STAU1 for SMD. My results demonstrated that knocking down *UPF1* or *STAU1* in DM1 cells enhances RNA foci accumulation and inhibits the cytoplasmic degradation of $(CUG)^{exp}$ *DMPK* transcripts (Figure 5-6 F and H). These effects are more pronounced in MBNL-deficient DM1 cells, suggesting that both NMD and SMD pathways are essential for the degradation of mutant *DMPK* mRNA, while the sequestration of MBNLs impedes this degradation process and in the absence of MBNLs, UPF1 and STAU1 act more effectively (Figure 5-10).

Moreover, my results indicate that UPF1 and STAU1 can effectively alleviate nuclear retention of exp*DMPK* transcripts (Figure 5-6 E and G). However, the mechanisms of *DMPK* mRNA transport promoted by UPF1 and STAU1 may differ. Although the exact mechanisms are not fully understood, current studies suggest that UPF1 can indirectly promote nuclear export of mRNA by interacting with other proteins that are involved in mRNA export, such as the nuclear pore complex (NPC) or the TREX complex, which mediates the coupling of mRNA processing to mRNA export [235, 244, 308]. Alternatively, UPF1 can directly interact with the exon junction complex (EJC), deposited on mRNA [237, 243]. Hence, UPF1 appears to play a universal role in promoting nuclear export of mRNAs by nonspecific interactions (Figure 5-10). In my study, I observed that inhibition of UPF1 had a comparable effects on the mutant *DMPK* mRNA export in both KB and P4C1 (Figure 5-6 E), but over-expression of UPF1 seemed to have a more profound impact on that in P4C1 than in KB (Figure 5-9 A). Therefore, further investigation is needed to fully comprehend the underlying mechanisms involved. Furthermore, my results showed that knock-down of *STAU1* led to a significant increase in nuclear exp*DMPK* transcript levels, which was even more pronounced in MBNL-deficient DM1 cells (Figure 5-6 G and H). Unlike UPF1, STAU1 is more likely to interact directly with the CUG repeats within the mutant *DMPK* mRNA and promote its nuclear export through the assistance of NXF (nuclear export factor) family proteins [248, 249, 309] (Figure 5-10). Therefore MBNL sequestration may inhibit this event to some extent.

Taken together, my study indicates that XRN2, EXOSC10, UPF1 and STAU1 are implicated in the RNA foci accumulation and the degradation of mutant *DMPK* mRNA. UPF1 and STAU1 also play an important role in its nuclear export. My study also highlights

the critical role of MBNL proteins in regulating mutant *DMPK* mRNA metabolism. The absence of MBNLs in DM1 appears to destabilise the mutant transcripts, expediting its degradation or nuclear export. Thus, targeting these pathways and enzymes could be a potential strategy for developing therapies for DM1. For example, drugs that can promote the activity of these enzymes could be used to facilitate the degradation of the mutant *DMPK* mRNA and alleviate its toxic effects. Furthermore, my study provides insights into the mechanisms underlying the regulation of mRNA decay and transport, which could have broader implications for understanding the pathogenesis of other genetic disorders and identifying new targets for therapeutic intervention.

Figure 5-10 Schematic illustration depicting the possible mechanisms involved in the nuclear export and degradation of mutant DMPK mRNA.



(a) The positions of the digital PCR primers and probe used to quantify the levels of mutant *DMPK* mRNA are indicated. These primers and probe bind 5' to the CUG repeat region. (b) Nuclear degradation of mutant *DMPK* mRNA can occur through the action

of nuclear exonucleases such as XRN2 and EXOSC10. (c-d) UPF1 and STAU1 can facilitate the nuclear export of mutant *DMPK* mRNA. The precise mechanism by which UPF1 and STAU1 transport mRNA from the nucleus to the cytoplasm is not fully understood, but it is hypothesised to involve interactions with the nuclear pore complex and other proteins involved in mRNA transport and translation. (e) Mutant *DMPK* mRNA can also be degraded in the cytoplasm by nonsense-mediated decay (NMD) and Staufen-mediated decay (SMD) pathways.

Chapter 6 Final Conclusions

Myotonic dystrophy type 1 (DM1) is a progressive, multisystemic disorder that affects approximately 1 in 8,000 individuals

worldwide [5, 310]. It is caused by an expansion of CTG repeats in the 3' untranslated region of the *DMPK* gene, resulting in the accumulation of mutant RNA foci in affected tissues [40]. The RNA foci are abnormal aggregates of CUG repeat-containing RNA that sequester RNA-binding proteins, particularly muscleblind-like (MBNL) proteins, leading to their functional loss and causing downstream molecular and cellular defects [55, 111, 117].

MBNL proteins are key regulators of alternative splicing, and their sequestration by RNA foci disrupts their normal function, leading to mis-splicing events that contribute to DM1 pathology [73, 111]. Mis-splicing affects the expression of genes involved in muscle function, metabolism, and development, leading to the clinical symptoms of DM1, including myotonia, muscle weakness, cardiac conduction defects, and cognitive impairment [119, 311, 312]. Recent studies have shown that restoring MBNL protein function can reverse some of the molecular and cellular defects in DM1 models, highlighting the importance of understanding the role of MBNL proteins the disease pathogenesis [173, 313, 314].

The primary objectives of this doctoral thesis were threefold. Firstly, to explore the dynamics of RNA foci in DM1; secondly, to unravel the mechanisms involved in the transportation and degradation of mutant mRNA; and finally, to investigate the role of MBNL proteins in these processes.

In the first part of this thesis, I utilised the CRISPR/Cas9 system to simultaneously knock out *MBNL1* and *MBNL2* genes in a fibroblast cell line (KB) derived from a patient with DM1. Due to the essential role of MBNL proteins in cellular functions, complete deletion of both MBNL1 and 2 could potentially lead to cell death [78]. Therefore, I could not achieve complete deletion of both proteins, but I succeeded in generating two MBNL-deficient cell

lines (P4C1 and P4B6) that provided me with a useful tool for subsequent experiments. I observed that MBNL1 was either completely absent or truncated with the loss of domains encoded by several exons, including ZnF 3-4, while MBNL2 was truncated with the loss of domain encoded by constitutive exon 6 in the P4C1 and P4B6 cell lines. These findings demonstrate that the endogenous expression of both MBNL1 and MBNL2 was effectively disrupted in the generated cell lines. This result gives me confidence in the validity of using P4C1 or P4B6 cells for subsequent experiments.

The effects of *MBNL1* and *2* double knock-out on DM1 are complex and multi-faceted. While the sequestration of MBNLs by repeat expansion RNA causes the amount of free MBNLs to be insufficient in DM1 cells, further knock-out of *MBNL1* and *2* in DM1 cells exacerbates many mis-splicing events and the mis-regulation of expression of many genes implicated in the disease. Additionally, it can lead to the emergence of novel aberrant splicing events and dysregulated gene expressions. These findings suggest that there is an excess of MBNL in KB nuclei. Furthermore, the double-inactivation of MBNL1 and 2 can also cause the instability of RNA foci formed by mutant *DMPK* transcripts, facilitating the dissipation of RNA foci and the degradation or nuclear export of mutant transcripts. Subsequently, the dissipation of RNA foci could lead to the release of other RNA binding proteins that are sequestered by the toxic RNAs, potentially rescuing other aberrant alternative splicing events and dysregulated gene expressions.

Therefore, it is important to consider both the release of MBNLs and the promotion of mutant *DMPK* mRNA decay when developing therapeutic strategies for DM1. A more nuanced understanding of the complex interplay between MBNL proteins, toxic RNAs, and

RNA processing is needed for researchers to develop more effective treatments for this debilitating disease.

In the second part of this thesis, I harnessed cutting-edge techniques to unravel the elusive characteristics of $(CUG)^{exp}$ RNA foci in DM1 cells and MBNL-deficient DM1 cells. Specifically, I utilised RNA-FISH and stochastic optical reconstruction microscopy (STORM) [252, 258] to overcome the optical diffraction limit and gain unparalleled insights into the organization and distribution of these foci. Through this approach, I were able to visualise not only large RNA foci, but also micro RNA foci, and distinguished them by quantifying the number of fluorophore signals in each focus. Furthermore, I used probe sets complementary to the 5' and 3' ends of *DMPK* transcripts to investigate the integrity of mutant *DMPK* mRNA. The results revealed that large RNA foci are typically composed of multiple *DMPK* transcripts, while micro foci only contain single *DMPK* transcripts. Therefore, micro RNA foci are speculated to be the break-down products of large RNA foci. It was also found that large RNA foci are unable to leave the nucleus, while micro foci can migrate to the cytoplasm, potentially contributing to the origin of RAN translation in DM1. Moreover, my study builds upon and expands the previous notion that the depletion of MBNL1 leads to elimination of RNA foci [63, 107]. I demonstrated that the absence of MBNLs modulates the breakdown of mutant *DMPK* mRNA. In DM1, mutant *DMPK* mRNA aggregates into large, compact RNA foci that can be easily detected under conventional fluorescent microscope. However, in the absence of MBNLs, mutant *DMPK* mRNA can only form into micro RNA foci, which are less stable than large foci. This thereby promotes the degradation of mutant *DMPK* mRNA or its transportation to the cytoplasm.

Overall, my study deepens our understanding of the mechanisms underlying RNA foci dynamics and the role of MBNL proteins in DM1 pathogenesis. Moreover, the use of STORM technology provides a valuable tool for investigating RNA foci in other diseases triggered by short tandem repeat (STR) mutations.

In the last part of this thesis, I identified several factors that are involved in the nuclear export and degradation of mutant *DMPK* mRNA. My study indicates that XRN2, EXOSC10, UPF1 and STAU1 are implicated in the RNA foci accumulation and the degradation of mutant *DMPK* mRNA. UPF1 and STAU1 also play an important role in its nuclear export. My study also highlights the critical role of MBNL proteins in regulating mutant *DMPK* mRNA metabolism. The absence of MBNLs in DM1 appears to destabilise mutant transcripts, expediting its degradation or nuclear export.

It is widely reported that the fundamental mechanism underlying myotonic dystrophy pathophysiology is the sequestration of MBNL proteins by repeat expansion *DMPK* transcripts. The consensus view is that mutant transcripts sequester MBNL1 and 2, but my preliminary data indicate that it is MBNL1 and 2 that hold or retain the mutant RNA in the nucleus. Although several therapeutic approaches [126, 173, 293, 294] have appeared to eliminate nuclear foci when visualised using in situ hybridisation, they have not successfully reversed the entire pathogenesis of DM1, as it is not clear what happens to the mutant RNA transcripts after the dissolution/disappearance of nuclear foci. It is important to keep in mind that there is no evidence showing that the foci are the sites of mutant RNA decay. Thus, having a better understanding of the metabolism of mutant *DMPK* mRNA and the effect of $(CUG)^{exp}$ RNA -MBNL protein complex on mutant *DMPK* mRNA decay could have important therapeutic implications for DM.

Studying the degradation mechanism of mutant *DMPK* mRNA in myotonic dystrophy may also provide a foundation for understanding the mechanisms of RNA degradation in other diseases caused by short tandem repeat (STR) mutations. STR mutations are a common cause of many genetic disorders, including Huntington's disease, fragile X syndrome, and several types of ataxia. These mutations can result in the production of toxic RNA transcripts, similar to those observed in myotonic dystrophy. Additionally, advances in technology and techniques for studying the degradation mechanism of mutant *DMPK* transcripts may have broader applications in the field of RNA biology and quality control mechanisms. The mechanisms that regulate RNA stability and degradation are complex and involve multiple pathways and RNA-binding proteins. As I gain a better understanding of the molecular mechanisms involved in RNA processing and degradation, I may be able to develop new therapies and interventions for a wide range of diseases caused by RNA dysregulation.

In addition to the points raised above there were two technical aspects that warrant further consideration.

Firstly, the P4B6 cell line was not utilised in all experiments for two reasons. Firstly, P4C1 provides a better representation of a DM1 cell line as the *DMPK* CTG repeat lengths are more comparable to KB, whereas P4B6 exhibited generally shorter CTG repeat expansions. This point is illustrated in Section 3.11. Secondly, P4C1 demonstrated superior *MBNL1* knock-out efficiency as it completely lacked the MBNL1 protein, while a truncated version was detectable in P4B6. Given these advantages, P4C1 was chosen as the preferred superior MBNL-deficient DM1 cell line derived from KB, in order to optimise research efficiency in terms of time and resources.

Secondly, because of the long expansion of CUG repeats in the *DMPK* mRNA and the highly compact structure of RNA foci, probe labelling in large foci was frequently overly concentrated. Even though the advent of super-resolution imaging methods has demonstrated single-molecule localisation precision of a few nanometres, achieving such high localisation precision with sub-10-nm spatial resolution in biological samples remains challenging [315]. When two or more fluorophores are separated by less than 10 nm, one fluorophore in on-state can serve as donor and excite fluorophores residing in their off-state via fluorescence resonance energy transfer (FRET) [316]. This FRET event causes simultaneous blinking of fluorophores with sub-10-nm interfluorophore distances, making it impossible to localise them individually. This implies that for a group of closely populated fluorophores [i.e. large $(CUG)^{exp}$ foci], the total number of their localisations is inevitably underestimated.

Nevertheless, SMLM is a powerful tool for the precise discovery and localisation of small molecules. In my projects, it allowed me to identify micro $(CUG)^{exp}$ foci, single transcripts, and even smaller targets, while also enabling me to measure their relative lengths and sizes. These achievements cannot be attained by using confocal or average fluorescence microscopes. However, it is important to note that SMLM is not designed for high-speed, large-volume imaging. The time-consuming nature of STORM microscopy and Bayesian cluster analysis, which can only process one file at a time and requires manual operation, makes it challenging for me to examine a large number of samples for each condition.

In summary, my investigations uncovered valuable data on the RNA foci dynamics in DM1, revealing the intricate mechanisms that underlie their formation, stability, and turnover. In addition,

I extended the delineation of complex pathways involved in the transportation and degradation of the mutant *DMPK* mRNA and provided insights into the critical role played by MBNL proteins in these processes. I hope that these findings will inspire further research in this field and pave the way for the development of novel therapeutic strategies that could improve the lives of patients suffering from DM1.

References

1. Romeo, V., *Myotonic Dystrophy Type 1 or Steinert's Disease*, in *Neurodegenerative Diseases*, S.I. Ahmad, Editor. 2012, Springer US: New York, NY. p. 239-257.
2. Steinberg, H. and A. Wagner, *[Hans Steinert: 100 years of myotonic dystrophy].* Nervenarzt, 2008. **79**(8): p. 961-2, 965-70.
3. Venugopal, V. and S. Pavlakis, *Duchenne Muscular Dystrophy*, in *StatPearls*. 2023: Treasure Island (FL).
4. Harley, H.G., et al., *Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy.* Am J Hum Genet, 1993. **52**(6): p. 1164-74.
5. Thornton, C.A., *Myotonic dystrophy.* Neurol Clin, 2014. **32**(3): p. 705-19, viii.
6. Thomas, J.D., et al., *Myotonic Dystrophy and Developmental Regulation of RNA Processing.* Compr Physiol, 2018. **8**(2): p. 509-553.
7. Mateos-Aierdi, A.J., et al., *Muscle wasting in myotonic dystrophies: a model of premature aging.* Front Aging Neurosci, 2015. **7**: p. 125.
8. Ricker, K., et al., *Proximal myotonic myopathy: a new dominant disorder with myotonia, muscle weakness, and cataracts.* Neurology, 1994. **44**(8): p. 1448-52.
9. Ho, G., M. Cardamone, and M. Farrar, *Congenital and childhood myotonic dystrophy: Current aspects of disease and future directions.* World J Clin Pediatr, 2015. **4**(4): p. 66-80.
10. Fregonezi, G., et al., *Heart rate variability in myotonic dystrophy type 1 patients.* Arq Bras Cardiol, 2012. **98**(4): p. 353-60.
11. Peric, S., et al., *Heart involvement in patients with myotonic dystrophy type 2.* Acta Neurol Belg, 2019. **119**(1): p. 77-82.
12. Pelargonio, G., et al., *Myotonic dystrophy and the heart.* Heart, 2002. **88**(6): p. 665-70.
13. Gossios, T.D., et al., *An overview of heart rhythm disorders and management in myotonic dystrophy type 1.* Heart Rhythm, 2022. **19**(3): p. 497-504.
14. Bassez, G., et al., *Severe cardiac arrhythmias in young patients with myotonic dystrophy type 1.* Neurology, 2004. **63**(10): p. 1939-41.
15. Schoser, B.G., et al., *Sudden cardiac death in myotonic dystrophy type 2.* Neurology, 2004. **63**(12): p. 2402-4.
16. Turillazzi, E., et al., *Cardiac fibrosis, arrhythmia and sudden death in myotonic dystrophy type 1: could TGF-ss1 improve the predictive accuracy of patients at risk, opening new therapeutic challenges?* Int J Cardiol, 2013. **168**(5): p. 4976-8.
17. Hawkins, A.M., et al., *Respiratory dysfunction in myotonic dystrophy type 1: A systematic review.* Neuromuscul Disord, 2019. **29**(3): p. 198-212.
18. Araujo, T.L., et al., *Respiratory muscle strength and quality of life in myotonic dystrophy patients.* Rev Port Pneumol, 2010. **16**(6): p. 892-8.

19. Yetimakman, A.F., B. Bayrakci, and A.M. Esquinas, *Myotonic Dystrophy type 1, individualised respiratory care rather than standart prognostication.* J Neurol Sci, 2019. **401**: p. 125-126.

20. Mazzoli, M., et al., *Predictors of respiratory decline in myotonic dystrophy type 1 (DM1): a longitudinal cohort study.* Acta Neurol Belg, 2021. **121**(1): p. 133-142.

21. Tieleman, A.A., et al., *Gastrointestinal involvement is frequent in Myotonic Dystrophy type 2.* Neuromuscular Disorders, 2008. **18**(8): p. 646-649.

22. Ronnblom, A., H. Forsberg, and A. Danielsson, *Gastrointestinal symptoms in myotonic dystrophy.* Scand J Gastroenterol, 1996. **31**(7): p. 654-7.

23. Peterson, J.A.M. and T.A. Cooper, *Clinical and Molecular Insights into Gastrointestinal Dysfunction in Myotonic Dystrophy Types 1 & 2.* Int J Mol Sci, 2022. **23**(23).

24. LaDonna, K.A., et al., *Hard to Swallow: A Phenomenological Exploration of the Experience of Caring for Individuals With Myotonic Dystrophy and Dysphagia.* J Neurosci Nurs, 2016. **48**(1): p. 42-51.

25. Tieleman, A.A., et al., *Dysphagia is present but mild in myotonic dystrophy type 2.* Neuromuscul Disord, 2009. **19**(3): p. 196-8.

26. Winters, S.J., *Endocrine Dysfunction in Patients With Myotonic Dystrophy.* J Clin Endocrinol Metab, 2021. **106**(10): p. 2819-2827.

27. Takase, S., *[Endocrine abnormalities in myotonic dystrophy].* Rinsho Shinkeigaku, 1983. **23**(12): p. 1086-90.

28. Vaccario, M.L., et al., *[Endocrine and metabolic aspects of myotonic dystrophy].* Riv Neurol, 1977. **47**(6): p. 616-23.

29. Moxley, R.T., et al., *Whole body insulin resistance in myotonic dystrophy.* Ann Neurol, 1984. **15**(2): p. 157-62.

30. Mastrogiacomo, I., et al., *Clinical and hormonal aspects of male hypogonadism in myotonic dystrophy.* Ital J Neurol Sci, 1996. **17**(1): p. 59-65.

31. Peric, S., et al., *Hypogonadism and erectile dysfunction in myotonic dystrophy type 1.* Acta Myol, 2013. **32**(2): p. 106-9.

32. Montagnese, F., et al., *Assessing the influence of age and gender on the phenotype of myotonic dystrophy type 2.* J Neurol, 2017. **264**(12): p. 2472-2480.

33. Abe, K., et al., *Involvement of the central nervous system in myotonic dystrophy.* J Neurol Sci, 1994. **127**(2): p. 179-85.

34. Meola, G. and V. Sansone, *Cerebral involvement in myotonic dystrophies.* Muscle Nerve, 2007. **36**(3): p. 294-306.

35. Fujino, H., et al., *Cognitive impairment and quality of life in patients with myotonic dystrophy type 1.* Muscle Nerve, 2018. **57**(5): p. 742-748.

36. Peric, S., et al., *Clusters of cognitive impairment among different phenotypes of myotonic dystrophy type 1 and type 2.* Neurol Sci, 2017. **38**(3): p. 415-423.

37. Angeard, N., et al., *Cognitive profile in childhood myotonic dystrophy type 1: is there a global impairment?* Neuromuscul Disord, 2007. **17**(6): p. 451-8.

38. LoRusso, S., B. Weiner, and W.D. Arnold, *Myotonic Dystrophies: Targeting Therapies for Multisystem Disease.* Neurotherapeutics, 2018. **15**(4): p. 872-884.

39. Harley, H.G., et al., *Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy.* Nature, 1992. **355**(6360): p. 545-6.

40. Brook, J.D., et al., *Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member.* Cell, 1992. **69**(2): p. 385.

41. Friedman, J.E., *Anticipation in hereditary disease: the history of a biomedical concept.* Hum Genet, 2011. **130**(6): p. 705-14.

42. Koshy, B.T. and H.Y. Zoghbi, *The CAG/polyglutamine tract diseases: gene products and molecular pathogenesis.* Brain Pathol, 1997. **7**(3): p. 927-42.

43. Morales, F., et al., *Individual-specific levels of CTG*CAG somatic instability are shared across multiple tissues in myotonic dystrophy type 1.* Hum Mol Genet, 2023. **32**(4): p. 621-631.

44. Meola, G. and R. Cardani, *Myotonic dystrophies: An update on clinical aspects, genetic, pathology, and molecular pathomechanisms.* Biochim Biophys Acta, 2015. **1852**(4): p. 594-606.

45. Ishii, S., et al., *Small increase in triplet repeat length of cerebellum from patients with myotonic dystrophy.* Hum Genet, 1996. **98**(2): p. 138-40.

46. Jansen, G., et al., *Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in (CTG)n repeat variation and selection against extreme expansion in sperm.* Am J Hum Genet, 1994. **54**(4): p. 575-85.

47. Ricker, K., et al., *Proximal myotonic myopathy. Clinical features of a multisystem disorder similar to myotonic dystrophy.* Arch Neurol, 1995. **52**(1): p. 25-31.

48. Liquori, C.L., et al., *Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9.* Science, 2001. **293**(5531): p. 864-7.

49. Ranum, L.P., et al., *Genetic mapping of a second myotonic dystrophy locus.* Nat Genet, 1998. **19**(2): p. 196-8.

50. Meola, G. and V. Sansone, *A newly-described myotonic disorder (proximal myotonic myopathy--PROMM): personal experience and review of the literature.* Ital J Neurol Sci, 1996. **17**(5): p. 347-53.

51. Udd, B., et al., *Proximal myotonic dystrophy--a family with autosomal dominant muscular dystrophy, cataracts, hearing loss and hypogonadism: heterogeneity of proximal myotonic syndromes?* Neuromuscul Disord, 1997. **7**(4): p. 217-28.

52. Udd, B. and R. Krahe, *The myotonic dystrophies: molecular, clinical, and therapeutic challenges.* Lancet Neurol, 2012. **11**(10): p. 891-905.

53. Ranum, L.P. and J.W. Day, *Myotonic dystrophy: RNA pathogenesis comes into focus.* Am J Hum Genet, 2004. **74**(5): p. 793-804.

54. Brinegar, A.E. and T.A. Cooper, *Roles for RNA-binding proteins in development and disease.* Brain Res, 2016. **1647**: p. 1-8.

55. Goodwin, M. and M.S. Swanson, *RNA-binding protein misregulation in microsatellite expansion disorders.* Adv Exp Med Biol, 2014. **825**: p. 353-88.

56. Kelemen, O., et al., *Function of alternative splicing.* Gene, 2013. **514**(1): p. 1-30.

57. Neverov, A.D., et al., *Alternative splicing and protein function.* BMC Bioinformatics, 2005. **6**: p. 266.

58. Park, J.W., et al., *Identification of alternative splicing regulators by RNA interference in Drosophila.* Proc Natl Acad Sci U S A, 2004. **101**(45): p. 15974-9.

59. Fu, X.D. and M. Ares, Jr., *Context-dependent control of alternative splicing by RNA-binding proteins.* Nat Rev Genet, 2014. **15**(10): p. 689-701.

60. Ladd, A.N., N. Charlet, and T.A. Cooper, *The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing.* Mol Cell Biol, 2001. **21**(4): p. 1285-96.

61. Konieczny, P., E. Stepniak-Konieczna, and K. Sobczak, *MBNL proteins and their target RNAs, interaction and splicing regulation.* Nucleic Acids Res, 2014. **42**(17): p. 10873-87.

62. Kino, Y., et al., *MBNL and CELF proteins regulate alternative splicing of the skeletal muscle chloride channel CLCN1.* Nucleic Acids Res, 2009. **37**(19): p. 6477-90.

63. Dansithong, W., et al., *MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1.* J Biol Chem, 2005. **280**(7): p. 5773-80.

64. Vihola, A., et al., *Altered expression and splicing of Ca(2+) metabolism genes in myotonic dystrophies DM1 and DM2.* Neuropathol Appl Neurobiol, 2013. **39**(4): p. 390-405.

65. Ohsawa, N., et al., *ABLIM1 splicing is abnormal in skeletal muscle of patients with DM1 and regulated by MBNL, CELF and PTBP1.* Genes Cells, 2015. **20**(2): p. 121-34.

66. Wang, E.T., et al., *Transcriptome alterations in myotonic dystrophy skeletal muscle and heart.* Hum Mol Genet, 2019. **28**(8): p. 1312-1321.

67. Otero, B.A., et al., *Transcriptome alterations in myotonic dystrophy frontal cortex.* Cell Rep, 2021. **34**(3): p. 108634.

68. Du, J., et al., *RNA toxicity and missplicing in the common eye disease fuchs endothelial corneal dystrophy.* J Biol Chem, 2015. **290**(10): p. 5979-90.

69. Daughters, R.S., et al., *RNA gain-of-function in spinocerebellar ataxia type 8.* PLoS Genet, 2009. **5**(8): p. e1000600.

70. Sellier, C., et al., *Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients.* EMBO J, 2010. **29**(7): p. 1248-61.

71. Mykowska, A., et al., *CAG repeats mimic CUG repeats in the misregulation of alternative splicing.* Nucleic Acids Res, 2011. **39**(20): p. 8938-51.

72. Du, H., et al., *Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy.* Nat Struct Mol Biol, 2010. **17**(2): p. 187-93.

73. Goodwin, M., et al., *MBNL Sequestration by Toxic RNAs and RNA Misprocessing in the Myotonic Dystrophy Brain.* Cell Rep, 2015. **12**(7): p. 1159-68.

74. Masuda, A., et al., *CUGBP1 and MBNL1 preferentially bind to 3' UTRs and facilitate mRNA decay.* Sci Rep, 2012. **2**: p. 209.

75. Perbellini, R., et al., *Dysregulation and cellular mislocalization of specific miRNAs in myotonic dystrophy type 1.* Neuromuscul Disord, 2011. **21**(2): p. 81-8.

76. Wang, E.T., et al., *Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins.* Cell, 2012. **150**(4): p. 710-24.

77. Batra, R., et al., *Loss of MBNL Leads to Disruption of Developmentally Regulated Alternative Polyadenylation in RNA-Mediated Disease.* Molecular Cell, 2014. **56**(2): p. 311-322.

78. Lee, K.Y., et al., *Compound loss of muscleblind-like function in myotonic dystrophy.* EMBO Mol Med, 2013. **5**(12): p. 1887-900.

79. Kanadia, R.N., et al., *Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3.* Gene Expr Patterns, 2003. **3**(4): p. 459-62.

80. Rau, F., et al., *Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy.* Nat Struct Mol Biol, 2011. **18**(7): p. 840-5.

81. Ashwal-Fluss, R., et al., *circRNA biogenesis competes with pre-mRNA splicing.* Mol Cell, 2014. **56**(1): p. 55-66.

82. Sznajder, L.J. and M.S. Swanson, *Short Tandem Repeat Expansions and RNA-Mediated Pathogenesis in Myotonic Dystrophy.* Int J Mol Sci, 2019. **20**(13).

83. Pascual, M., et al., *The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing.* Differentiation, 2006. **74**(2-3): p. 65-80.

84. Tran, H., et al., *Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of Muscleblind-like-1 isoforms.* J Biol Chem, 2011. **286**(18): p. 16435-46.

85. Terenzi, F. and A.N. Ladd, *Conserved developmental alternative splicing of muscleblind-like (MBNL) transcripts regulates MBNL localization and activity.* RNA Biol, 2010. **7**(1): p. 43-55.

86. Kino, Y., et al., *Nuclear localization of MBNL1: splicing-mediated autoregulation and repression of repeat-derived aberrant proteins.* Hum Mol Genet, 2015. **24**(3): p. 740-56.

87. Yuan, Y., et al., *Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs.* Nucleic Acids Res, 2007. **35**(16): p. 5474-86.

88. Teplova, M. and D.J. Patel, *Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1.* Nat Struct Mol Biol, 2008. **15**(12): p. 1343-51.

89. Hale, M.A., et al., *An engineered RNA binding protein with improved splicing regulation.* Nucleic Acids Res, 2018. **46**(6): p. 3152-3168.

90. Fu, Y., et al., *MBNL1-RNA recognition: contributions of MBNL1 sequence and RNA conformation.* Chembiochem, 2012. **13**(1): p. 112-9.

91. Grammatikakis, I., et al., *Identification of MBNL1 and MBNL3 domains required for splicing activation and repression.* Nucleic Acids Res, 2011. **39**(7): p. 2769-80.

92. Wang, P.Y., et al., *Ubiquitination of MBNL1 Is Required for Its Cytoplasmic Localization and Function in Promoting Neurite Outgrowth.* Cell Rep, 2018. **22**(9): p. 2294-2306.

93. Tabaglio, T., et al., *MBNL1 alternative splicing isoforms play opposing roles in cancer.* Life Sci Alliance, 2018. **1**(5): p. e201800157.

94. Philips, A.V., L.T. Timchenko, and T.A. Cooper, *Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy.* Science, 1998. **280**(5364): p. 737-41.

95. Wang, E.T., et al., *Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins.* Genome research, 2015. **25**(6): p. 858-871.

96. Ohsawa, N., et al., *ABLIM1 splicing is abnormal in skeletal muscle of patients with DM 1 and regulated by MBNL, CELF and PTBP 1.* Genes to Cells, 2015. **20**(2): p. 121-134.

97. Wang, E.T., et al., *Transcriptome alterations in myotonic dystrophy skeletal muscle and heart.* Human molecular genetics, 2019. **28**(8): p. 1312-1321.

98. Witten, J.T. and J. Ule, *Understanding splicing regulation through RNA splicing maps.* Trends in Genetics, 2011. **27**(3): p. 89-97.

99. Batra, R., et al., *Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease.* Molecular cell, 2014. **56**(2): p. 311-322.

100. Salisbury, E., et al., *Ectopic expression of cyclin D3 corrects differentiation of DM1 myoblasts through activation of RNA CUG-binding protein, CUGBP1.* Exp Cell Res, 2008. **314**(11-12): p. 2266-78.

101. Jones, K., et al., *GSK3beta mediates muscle pathology in myotonic dystrophy.* J Clin Invest, 2012. **122**(12): p. 4461-72.

102. Lee, J.E., et al., *Systematic analysis of cis-elements in unstable mRNAs demonstrates that CUGBP1 is a key regulator of mRNA decay in muscle cells.* PLoS One, 2010. **5**(6): p. e11201.

103. Taneja, K.L., et al., *Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues.* J Cell Biol, 1995. **128**(6): p. 995-1002.

104. Wojciechowska, M. and W.J. Krzyzosiak, *Cellular toxicity of expanded RNA repeats: focus on RNA foci.* Hum Mol Genet, 2011. **20**(19): p. 3811-21.

105. Van Treeck, B. and R. Parker, *Emerging Roles for Intermolecular RNA-RNA Interactions in RNP Assemblies.* Cell, 2018. **174**(4): p. 791-802.

106. Smith, K.P., et al., *Defining early steps in mRNA transport: mutant mRNA in myotonic dystrophy type I is blocked at entry into SC-35 domains.* J Cell Biol, 2007. **178**(6): p. 951-64.

107. Sznajder, L.J., et al., *Mechanistic determinants of MBNL activity.* Nucleic Acids Res, 2016. **44**(21): p. 10326-10342.

108. Querido, E., et al., *Stochastic and reversible aggregation of mRNA with expanded CUG-triplet repeats.* J Cell Sci, 2011. **124**(Pt 10): p. 1703-14.

109. Fardaei, M., et al., *In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts.* Nucleic Acids Res, 2001. **29**(13): p. 2766-71.

110. Konieczny, P., et al., *Autoregulation of MBNL1 function by exon 1 exclusion from MBNL1 transcript.* Nucleic Acids Res, 2017. **45**(4): p. 1760-1775.

111. Paul, S., et al., *Expanded CUG repeats Dysregulate RNA splicing by altering the stoichiometry of the muscleblind 1 complex.* J Biol Chem, 2011. **286**(44): p. 38427-38.

112. Laurent, F.X., et al., *New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats.* Nucleic Acids Res, 2012. **40**(7): p. 3159-71.

113. Pettersson, O.J., et al., *DDX6 regulates sequestered nuclear CUG-expanded DMPK-mRNA in dystrophia myotonica type 1.* Nucleic Acids Res, 2014. **42**(11): p. 7186-200.

114. Sellier, C., et al., *rbFOX1/MBNL1 competition for CCUG RNA repeats binding contributes to myotonic dystrophy type 1/type 2 differences.* Nat Commun, 2018. **9**(1): p. 2009.

115. Park, S., et al., *Structural Basis for Interaction of the Tandem Zinc Finger Domains of Human Muscleblind with Cognate RNA from Human Cardiac Troponin T.* Biochemistry, 2017. **56**(32): p. 4154-4168.

116. Xia, G. and T. Ashizawa, *Dynamic changes of nuclear RNA foci in proliferating DM1 cells.* Histochem Cell Biol, 2015. **143**(6): p. 557-64.

117. Thomas, J.D., et al., *Disrupted prenatal RNA processing and myogenesis in congenital myotonic dystrophy.* Genes Dev, 2017. **31**(11): p. 1122-1133.

118. Botta, A., et al., *Effect of the [CCTG]n repeat expansion on ZNF9 expression in myotonic dystrophy type II (DM2).* Biochim Biophys Acta, 2006. **1762**(3): p. 329-34.

119. Nakamori, M., et al., *Splicing biomarkers of disease severity in myotonic dystrophy.* Ann Neurol, 2013. **74**(6): p. 862-72.

120. Klinck, R., et al., *RBFOX1 cooperates with MBNL1 to control splicing in muscle, including events altered in myotonic dystrophy type 1.* PLoS One, 2014. **9**(9): p. e107324.

121. Kim, E.Y., et al., *Distinct pathological signatures in human cellular models of myotonic dystrophy subtypes.* JCI Insight, 2019. **4**(6).

122. Zu, T., et al., *RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2.* Neuron, 2017. **95**(6): p. 1292-1305 e5.

123. van Agtmaal, E.L., et al., *CRISPR/Cas9-Induced (CTGCAG)n Repeat Instability in the Myotonic Dystrophy Type 1 Locus: Implications for Therapeutic Genome Editing.* Mol Ther, 2017. **25**(1): p. 24-43.

124. Provenzano, C., et al., *CRISPR/Cas9-Mediated Deletion of CTG Expansions Recovers Normal Phenotype in Myogenic Cells Derived from*

Myotonic Dystrophy 1 Patients. Mol Ther Nucleic Acids, 2017. **9**: p. 337-348.

125. Cardinali, B., et al., *Time-controlled and muscle-specific CRISPR/Cas9-mediated deletion of CTG-repeat expansion in the DMPK gene.* Mol Ther Nucleic Acids, 2022. **27**: p. 184-199.

126. Lo Scrudato, M., et al., *Genome Editing of Expanded CTG Repeats within the Human DMPK Gene Reduces Nuclear RNA Foci in the Muscle of DM1 Mice.* Mol Ther, 2019. **27**(8): p. 1372-1388.

127. Warf, M.B., et al., *Pentamidine reverses the splicing defects associated with myotonic dystrophy.* Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18551-6.

128. Coonrod, L.A., et al., *Reducing levels of toxic RNA with small molecules.* ACS Chem Biol, 2013. **8**(11): p. 2528-37.

129. Siboni, R.B., et al., *Actinomycin D Specifically Reduces Expanded CUG Repeat RNA in Myotonic Dystrophy Models.* Cell Rep, 2015. **13**(11): p. 2386-2394.

130. Pinto, B.S., et al., *Impeding Transcription of Expanded Microsatellite Repeats by Deactivated Cas9.* Mol Cell, 2017. **68**(3): p. 479-490 e5.

131. Aagaard, L. and J.J. Rossi, *RNAi therapeutics: principles, prospects and challenges.* Adv Drug Deliv Rev, 2007. **59**(2-3): p. 75-86.

132. Matranga, C., et al., *Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes.* Cell, 2005. **123**(4): p. 607-20.

133. Langlois, M.A., et al., *Cytoplasmic and nuclear retained DMPK mRNAs are targets for RNA interference in myotonic dystrophy cells.* J Biol Chem, 2005. **280**(17): p. 16949-54.

134. Sobczak, K., et al., *RNA interference targeting CUG repeats in a mouse model of myotonic dystrophy.* Mol Ther, 2013. **21**(2): p. 380-7.

135. Bisset, D.R., et al., *Therapeutic impact of systemic AAV-mediated RNA interference in a mouse model of myotonic dystrophy.* Hum Mol Genet, 2015. **24**(17): p. 4971-83.

136. Gomes-Pereira, M., T.A. Cooper, and G. Gourdon, *Myotonic dystrophy mouse models: towards rational therapy development.* Trends Mol Med, 2011. **17**(9): p. 506-17.

137. Sztretye, M., et al., *From Mice to Humans: An Overview of the Potentials and Limitations of Current Transgenic Mouse Models of Major Muscular Dystrophies and Congenital Myopathies.* Int J Mol Sci, 2020. **21**(23).

138. Wheeler, T.M., et al., *Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA.* Science, 2009. **325**(5938): p. 336-9.

139. Mulders, S.A., et al., *Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy.* Proc Natl Acad Sci U S A, 2009. **106**(33): p. 13915-20.

140. Wojtkowiak-Szlachcic, A., et al., *Short antisense-locked nucleic acids (all-LNAs) correct alternative splicing abnormalities in myotonic dystrophy.* Nucleic Acids Res, 2015. **43**(6): p. 3318-31.

141. Lee, J.E., C.F. Bennett, and T.A. Cooper, *RNase H-mediated degradation of toxic RNA in myotonic dystrophy type 1.* Proc Natl Acad Sci U S A, 2012. **109**(11): p. 4221-6.

142. Pandey, S.K., et al., *Identification and characterization of modified antisense oligonucleotides targeting DMPK in mice and nonhuman primates for the treatment of myotonic dystrophy type 1.* J Pharmacol Exp Ther, 2015. **355**(2): p. 329-40.

143. Batra, R., et al., *Elimination of Toxic Microsatellite Repeat Expansion RNA by RNA-Targeting Cas9.* Cell, 2017. **170**(5): p. 899-912 e10.

144. Ketley, A., et al., *CDK12 inhibition reduces abnormalities in cells from patients with myotonic dystrophy and in a mouse model.* Sci Transl Med, 2020. **12**(541).

145. Pascual-Gilabert, M., R. Artero, and A. Lopez-Castel, *The myotonic dystrophy type 1 drug development pipeline: 2022 edition.* Drug Discov Today, 2023. **28**(3): p. 103489.

146. Mullard, A., *Antibody-oligonucleotide conjugates enter the clinic.* Nat Rev Drug Discov, 2022. **21**(1): p. 6-8.

147. Zanotti, S., et al. *Repeat Dosing with DYNE-101 is Well Tolerated and Leads to a Sustained Reduction of DMPK RNA Expression in Key Muscles for DM1 Pathology in hTfR1/DMSXL Mice and NHPs*. in *MOLECULAR THERAPY*. 2022. CELL PRESS 50 HAMPSHIRE ST, FLOOR 5, CAMBRIDGE, MA 02139 USA.

148. Li, J., et al., *A Dimeric 2,9-Diamino-1,10-phenanthroline Derivative Improves Alternative Splicing in Myotonic Dystrophy Type 1 Cell and Mouse Models.* Chemistry, 2018. **24**(68): p. 18115-18122.

149. Angelbello, A.J., et al., *A Small Molecule that Binds an RNA Repeat Expansion Stimulates Its Decay via the Exosome Complex.* Cell Chem Biol, 2021. **28**(1): p. 34-45 e6.

150. Lee, J., et al., *Intrinsically cell-penetrating multivalent and multitargeting ligands for myotonic dystrophy type 1.* Proc Natl Acad Sci U S A, 2019. **116**(18): p. 8709-8714.

151. Nguyen, L., et al., *Rationally designed small molecules that target both the DNA and RNA causing myotonic dystrophy type 1.* J Am Chem Soc, 2015. **137**(44): p. 14180-9.

152. Wheeler, T.M., et al., *Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA.* Science, 2009. **325**(5938): p. 336-339.

153. Jauvin, D., et al., *Targeting DMPK with Antisense Oligonucleotide Improves Muscle Strength in Myotonic Dystrophy Type 1 Mice.* Mol Ther Nucleic Acids, 2017. **7**: p. 465-474.

154. Ait Benichou, S., et al., *Antisense oligonucleotides as a potential treatment for brain deficits observed in myotonic dystrophy type 1.* Gene Ther, 2022.

155. Jauvin, D., et al., *Targeting DMPK with antisense oligonucleotide improves muscle strength in myotonic dystrophy type 1 mice.* Molecular Therapy-Nucleic Acids, 2017. **7**: p. 465-474.

156. Chakraborty, M., et al., *Daunorubicin reduces MBNL1 sequestration caused by CUG-repeat expansion and rescues cardiac dysfunctions in a*

Drosophila model of myotonic dystrophy. Disease models & mechanisms, 2018. **11**(4): p. dmm032557.

157. Nakamori, M., et al., *Oral administration of erythromycin decreases RNA toxicity in myotonic dystrophy.* Ann Clin Transl Neurol, 2016. **3**(1): p. 42-54.

158. Hoskins, J.W., et al., *Lomofungin and dilomofungin: inhibitors of MBNL1-CUG RNA binding with distinct cellular effects.* Nucleic Acids Res, 2014. **42**(10): p. 6591-602.

159. Song, K.-Y., et al., *MBNL1 reverses the proliferation defect of skeletal muscle satellite cells in myotonic dystrophy type 1 by inhibiting autophagy via the mTOR pathway.* Cell Death & Disease, 2020. **11**(7): p. 1-16.

160. Bargiela, A., et al., *Increased Muscleblind levels by chloroquine treatment improve myotonic dystrophy type 1 phenotypes in in vitro and in vivo models.* Proceedings of the National Academy of Sciences, 2019. **116**(50): p. 25203-25213.

161. Zhang, J., et al., *The natural compound neobractatin inhibits tumor metastasis by upregulating the RNA-binding-protein MBNL2.* Cell death & disease, 2019. **10**(8): p. 1-13.

162. Louis, J.M., et al., *TNF-alpha regulates alternative splicing of genes participating in pathways of crucial metabolic syndromes; a transcriptome wide study.* Cytokine, 2020. **125**: p. 154815.

163. Louis, J.M., et al., *Global analysis of RNA–protein interactions in TNF-α induced alternative splicing in metabolic disorders.* FEBS letters, 2021. **595**(4): p. 476-490.

164. Panahi, G., et al., *High glucose induces inflammatory responses in HepG2 cells via the oxidative stress-mediated activation of NF-kappaB, and MAPK pathways in HepG2 cells.* Arch Physiol Biochem, 2018. **124**(5): p. 468-474.

165. Malakar, P., et al., *Insulin receptor alternative splicing is regulated by insulin signaling and modulates beta cell survival.* Scientific reports, 2016. **6**(1): p. 1-14.

166. Altowyan, M.S., et al., *Spiroindolone analogues bearing benzofuran moiety as a selective cyclooxygenase COX-1 with TNF-α and IL-6 inhibitors.* Saudi journal of biological sciences, 2020. **27**(5): p. 1208-1216.

167. Chen, G., et al., *Phenylbutazone induces expression of MBNL1 and suppresses formation of MBNL1-CUG RNA foci in a mouse model of myotonic dystrophy.* Scientific reports, 2016. **6**(1): p. 1-11.

168. Huang, K., et al., *Inhibition of cyclooxygenase-1 by nonsteroidal anti-inflammatory drugs demethylates MeR2 enhancer and promotes Mbnl1 transcription in myogenic cells.* Scientific reports, 2020. **10**(1): p. 1-14.

169. Herrendorff, R., et al., *Identification of Plant-derived Alkaloids with Therapeutic Potential for Myotonic Dystrophy Type I.* J Biol Chem, 2016. **291**(33): p. 17165-77.

170. Zhang, F., et al., *A flow cytometry-based screen identifies MBNL1 modulators that rescue splicing defects in myotonic dystrophy type I.* Hum Mol Genet, 2017. **26**(16): p. 3056-3068.

171. Wojciechowska, M., et al., *Small molecule kinase inhibitors alleviate different molecular features of myotonic dystrophy type 1.* RNA Biol, 2014. **11**(6): p. 742-54.

172. Wang, G.S., et al., *PKC inhibition ameliorates the cardiac phenotype in a mouse model of myotonic dystrophy type 1.* J Clin Invest, 2009. **119**(12): p. 3797-806.

173. Ketley, A., et al., *High-content screening identifies small molecules that remove nuclear foci, affect MBNL distribution and CELF1 protein levels via a PKC-independent pathway in myotonic dystrophy cell lines.* Hum Mol Genet, 2014. **23**(6): p. 1551-62.

174. Wang, M., et al., *Correction of Glycogen Synthase Kinase 3beta in Myotonic Dystrophy 1 Reduces the Mutant RNA and Improves Postnatal Survival of DMSXL Mice.* Mol Cell Biol, 2019. **39**(21).

175. Horrigan, J., et al., *A Phase 2 Study of AMO-02 (Tideglusib) in Congenital and Childhood-Onset Myotonic Dystrophy Type 1 (DM1).* Pediatr Neurol, 2020. **112**: p. 84-93.

176. Warf, M.B. and J.A. Berglund, *MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T.* RNA, 2007. **13**(12): p. 2238-51.

177. Fugier, C., et al., *Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy.* Nat Med, 2011. **17**(6): p. 720-5.

178. Wheeler, T.M., et al., *Correction of ClC-1 splicing eliminates chloride channelopathy and myotonia in mouse models of myotonic dystrophy.* J Clin Invest, 2007. **117**(12): p. 3952-7.

179. Oana, K., et al., *Manumycin A corrects aberrant splicing of Clcn1 in myotonic dystrophy type 1 (DM1) mice.* Sci Rep, 2013. **3**: p. 2142.

180. Hara, M., et al., *Identification of Ras farnesyltransferase inhibitors by microbial screening.* Proc Natl Acad Sci U S A, 1993. **90**(6): p. 2281-5.

181. Jutzi, D. and M.D. Ruepp, *Alternative Splicing in Human Biology and Disease.* Methods Mol Biol, 2022. **2537**: p. 1-19.

182. Anderson, P. and N. Kedersha, *RNA granules.* J Cell Biol, 2006. **172**(6): p. 803-8.

183. Eulalio, A., I. Behm-Ansmant, and E. Izaurralde, *P bodies: at the crossroads of post-transcriptional pathways.* Nat Rev Mol Cell Biol, 2007. **8**(1): p. 9-22.

184. Andrei, M.A., et al., *A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies.* RNA, 2005. **11**(5): p. 717-27.

185. Cougot, N., S. Babajko, and B. Seraphin, *Cytoplasmic foci are sites of mRNA decay in human cells.* J Cell Biol, 2004. **165**(1): p. 31-40.

186. Sheth, U. and R. Parker, *Decapping and decay of messenger RNA occur in cytoplasmic processing bodies.* Science, 2003. **300**(5620): p. 805-8.

187. Bashkirov, V.I., et al., *A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates.* J Cell Biol, 1997. **136**(4): p. 761-73.

188. Sheth, U. and R. Parker, *Targeting of aberrant mRNAs to cytoplasmic processing bodies.* Cell, 2006. **125**(6): p. 1095-109.

189. Sen, G.L. and H.M. Blau, *Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies.* Nat Cell Biol, 2005. **7**(6): p. 633-6.

190. Bhattacharyya, S.N., et al., *Relief of microRNA-mediated translational repression in human cells subjected to stress.* Cell, 2006. **125**(6): p. 1111-24.

191. Brengues, M., D. Teixeira, and R. Parker, *Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies.* Science, 2005. **310**(5747): p. 486-9.

192. Gatfield, D. and E. Izaurralde, *Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila.* Nature, 2004. **429**(6991): p. 575-8.

193. Stevens, A., et al., *Beta -Globin mRNA decay in erythroid cells: UG site-preferred endonucleolytic cleavage that is augmented by a premature termination codon.* Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12741-6.

194. Liu, J., et al., *Argonaute2 is the catalytic engine of mammalian RNAi.* Science, 2004. **305**(5689): p. 1437-41.

195. Song, J.J., et al., *Crystal structure of Argonaute and its implications for RISC slicer activity.* Science, 2004. **305**(5689): p. 1434-7.

196. Labno, A., R. Tomecki, and A. Dziembowski, *Cytoplasmic RNA decay pathways - Enzymes and mechanisms.* Biochim Biophys Acta, 2016. **1863**(12): p. 3125-3147.

197. Conti, E. and E. Izaurralde, *Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species.* Curr Opin Cell Biol, 2005. **17**(3): p. 316-25.

198. Couttet, P. and T. Grange, *Premature termination codons enhance mRNA decapping in human cells.* Nucleic Acids Res, 2004. **32**(2): p. 488-94.

199. Muhlrad, D. and R. Parker, *Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance.* RNA, 1999. **5**(10): p. 1299-307.

200. Welch, E.M. and A. Jacobson, *An internal open reading frame triggers nonsense-mediated decay of the yeast SPT10 mRNA.* EMBO J, 1999. **18**(21): p. 6134-45.

201. Kashima, I., et al., *Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay.* Genes Dev, 2006. **20**(3): p. 355-67.

202. Unterholzner, L. and E. Izaurralde, *SMG7 acts as a molecular link between mRNA surveillance and mRNA decay.* Mol Cell, 2004. **16**(4): p. 587-96.

203. Gatfield, D., et al., *Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways.* EMBO J, 2003. **22**(15): p. 3960-70.

204. Buhler, M., et al., *EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length.* Nat Struct Mol Biol, 2006. **13**(5): p. 462-4.

205. Amrani, N., et al., *A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay.* Nature, 2004. **432**(7013): p. 112-8.

206. Mendell, J.T., et al., *Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise.* Nat Genet, 2004. **36**(10): p. 1073-8.

207. Kaygun, H. and W.F. Marzluff, *Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1.* Nat Struct Mol Biol, 2005. **12**(9): p. 794-800.

208. Kim, Y.K., et al., *Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay.* Cell, 2005. **120**(2): p. 195-208.

209. Doma, M.K. and R. Parker, *Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation.* Nature, 2006. **440**(7083): p. 561-4.

210. Garneau, N.L., J. Wilusz, and C.J. Wilusz, *The highways and byways of mRNA decay.* Nat Rev Mol Cell Biol, 2007. **8**(2): p. 113-26.

211. Bresson, S. and D. Tollervey, *Surveillance-ready transcription: nuclear RNA decay as a default fate.* Open Biol, 2018. **8**(3).

212. Allmang, C., et al., *Functions of the exosome in rRNA, snoRNA and snRNA synthesis.* EMBO J, 1999. **18**(19): p. 5399-410.

213. Mitchell, P., et al., *The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases.* Cell, 1997. **91**(4): p. 457-66.

214. Makino, D.L., M. Baumgartner, and E. Conti, *Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex.* Nature, 2013. **495**(7439): p. 70-5.

215. Wasmuth, E.V. and C.D. Lima, *Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel.* Mol Cell, 2012. **48**(1): p. 133-44.

216. Schuch, B., et al., *The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase.* EMBO J, 2014. **33**(23): p. 2829-46.

217. Assenholt, J., et al., *Exonucleolysis is required for nuclear mRNA quality control in yeast THO mutants.* RNA, 2008. **14**(11): p. 2305-13.

218. Gudipati, R.K., et al., *Extensive degradation of RNA precursors by the exosome in wild-type cells.* Mol Cell, 2012. **48**(3): p. 409-21.

219. Schneider, C., et al., *Transcriptome-wide analysis of exosome targets.* Mol Cell, 2012. **48**(3): p. 422-33.

220. Wyers, F., et al., *Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase.* Cell, 2005. **121**(5): p. 725-37.

221. Preker, P., et al., *RNA exosome depletion reveals transcription upstream of active human promoters.* Science, 2008. **322**(5909): p. 1851-4.

222. Schmid, M. and T.H. Jensen, *Controlling nuclear RNA levels.* Nat Rev Genet, 2018. **19**(8): p. 518-529.

223. Bresson, S.M., et al., *Canonical Poly(A) Polymerase Activity Promotes the Decay of a Wide Variety of Mammalian Nuclear RNAs.* PLoS Genet, 2015. **11**(10): p. e1005610.

224. Mukherjee, N., et al., *Integrative classification of human coding and noncoding genes through RNA metabolism profiles.* Nat Struct Mol Biol, 2017. **24**(1): p. 86-96.

225. Tudek, A., M. Schmid, and T.H. Jensen, *Escaping nuclear decay: the significance of mRNA export for gene expression.* Curr Genet, 2019. **65**(2): p. 473-476.

226. Libri, D., et al., *Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation.* Mol Cell Biol, 2002. **22**(23): p. 8254-66.

227. Tudek, A., et al., *A Nuclear Export Block Triggers the Decay of Newly Synthesized Polyadenylated RNA.* Cell Rep, 2018. **24**(9): p. 2457-2467 e7.

228. Schmid, M., et al., *The Nuclear PolyA-Binding Protein Nab2p Is Essential for mRNA Production.* Cell Rep, 2015. **12**(1): p. 128-139.

229. Libri, D., *Nuclear poly(a)-binding proteins and nuclear degradation: take the mRNA and run?* Mol Cell, 2010. **37**(1): p. 3-5.

230. Meola, N. and T.H. Jensen, *Targeting the nuclear RNA exosome: Poly(A) binding proteins enter the stage.* RNA Biol, 2017. **14**(7): p. 820-826.

231. Nguyen, D., et al., *A Polyadenylation-Dependent 3' End Maturation Pathway Is Required for the Synthesis of the Human Telomerase RNA.* Cell Rep, 2015. **13**(10): p. 2244-57.

232. Muhlrad, D., C.J. Decker, and R. Parker, *Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript.* Genes Dev, 1994. **8**(7): p. 855-66.

233. Wormington, M., A.M. Searfoss, and C.A. Hurney, *Overexpression of poly(A) binding protein prevents maturation-specific deadenylation and translational inactivation in Xenopus oocytes.* EMBO J, 1996. **15**(4): p. 900-9.

234. Kawabe, Y., et al., *The RNA exosome complex degrades expanded hexanucleotide repeat RNA in C9orf72 FTLD/ALS.* EMBO J, 2020. **39**(19): p. e102700.

235. Hogg, J.R. and S.P. Goff, *Upf1 senses 3'UTR length to potentiate mRNA decay.* Cell, 2010. **143**(3): p. 379-89.

236. Lee, S.R., et al., *Target Discrimination in Nonsense-Mediated mRNA Decay Requires Upf1 ATPase Activity.* Mol Cell, 2015. **59**(3): p. 413-25.

237. Cheng, Z., et al., *Structural and functional insights into the human Upf1 helicase core.* EMBO J, 2007. **26**(1): p. 253-64.

238. Gowravaram, M., et al., *A conserved structural element in the RNA helicase UPF1 regulates its catalytic activity in an isoform-specific manner.* Nucleic Acids Res, 2018. **46**(5): p. 2648-2659.

239. Gleghorn, M.L. and L.E. Maquat, *UPF1 learns to relax and unwind.* Mol Cell, 2011. **41**(6): p. 621-3.

240. Fritz, S.E., et al., *The RNA-binding protein PTBP1 promotes ATPase-dependent dissociation of the RNA helicase UPF1 to protect transcripts*

from nonsense-mediated mRNA decay. J Biol Chem, 2020. **295**(33): p. 11613-11625.

241. Kishor, A., Z. Ge, and J.R. Hogg, *hnRNP L-dependent protection of normal mRNAs from NMD subverts quality control in B cell lymphoma.* EMBO J, 2019. **38**(3).

242. Ge, Z., et al., *Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway.* Elife, 2016. **5**.

243. Kurosaki, T. and L.E. Maquat, *Nonsense-mediated mRNA decay in humans at a glance.* J Cell Sci, 2016. **129**(3): p. 461-7.

244. Singh, A.K., et al., *The RNA helicase UPF1 associates with mRNAs co-transcriptionally and is required for the release of mRNAs from gene loci.* Elife, 2019. **8**.

245. Heber, S., et al., *Staufen2-mediated RNA recognition and localization requires combinatorial action of multiple domains.* Nat Commun, 2019. **10**(1): p. 1659.

246. Sugimoto, Y., et al., *hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1.* Nature, 2015. **519**(7544): p. 491-4.

247. Cho, H., et al., *SMG1 regulates adipogenesis via targeting of staufen1-mediated mRNA decay.* Biochim Biophys Acta, 2013. **1829**(12): p. 1276-87.

248. Gowravaram, M., et al., *Insights into the assembly and architecture of a Staufen-mediated mRNA decay (SMD)-competent mRNP.* Nat Commun, 2019. **10**(1): p. 5054.

249. Ravel-Chapuis, A., et al., *The RNA-binding protein Staufen1 is increased in DM1 skeletal muscle and promotes alternative pre-mRNA splicing.* J Cell Biol, 2012. **196**(6): p. 699-712.

250. Garcia, S.M., et al., *Identification of genes in toxicity pathways of trinucleotide-repeat RNA in C. elegans.* Nat Struct Mol Biol, 2014. **21**(8): p. 712-20.

251. Sanderson, M.J., et al., *Fluorescence microscopy.* Cold Spring Harb Protoc, 2014. **2014**(10): p. pdb top071795.

252. Whelan, D.R. and T.D. Bell, *Super-Resolution Single-Molecule Localization Microscopy: Tricks of the Trade.* J Phys Chem Lett, 2015. **6**(3): p. 374-82.

253. Mockl, L. and W.E. Moerner, *Super-resolution Microscopy with Single Molecules in Biology and Beyond-Essentials, Current Trends, and Future Challenges.* J Am Chem Soc, 2020. **142**(42): p. 17828-17844.

254. Sage, D., et al., *Quantitative evaluation of software packages for single-molecule localization microscopy.* Nat Methods, 2015. **12**(8): p. 717-24.

255. Vogelsang, J., et al., *Make them blink: probes for super-resolution microscopy.* Chemphyschem, 2010. **11**(12): p. 2475-90.

256. Liu, X., et al., *Nanographenes: Ultrastable, Switchable, and Bright Probes for Super-Resolution Microscopy.* Angew Chem Int Ed Engl, 2020. **59**(1): p. 496-502.

257. Yadav, A., C. Rao, and C.K. Nandi, *Fluorescent Probes for Super-Resolution Microscopy of Lysosomes.* ACS Omega, 2020. **5**(42): p. 26967-26977.

258. Lelek, M., et al., *Single-molecule localization microscopy.* Nat Rev Methods Primers, 2021. **1**.

259. Maruno, T., et al. *Comparison of CMOS and EMCCD Cameras for Computational Imaging and Application to Superresolution Localization Microscopy*. in *Imaging and Applied Optics Technical Papers*. 2012. Monterey, California: Optica Publishing Group.

260. Fullerton, S., et al., *Camera simulation engine enables efficient system optimization for super-resolution imaging*. SPIE BiOS. Vol. 8228. 2012: SPIE.

261. Liu, S., P. Hoess, and J. Ries, *Super-Resolution Microscopy for Structural Cell Biology.* Annu Rev Biophys, 2022. **51**: p. 301-326.

262. Durisic, N., L.L. Cuervo, and M. Lakadamyali, *Quantitative super-resolution microscopy: pitfalls and strategies for image analysis.* Curr Opin Chem Biol, 2014. **20**: p. 22-8.

263. Rubin-Delanchy, P., et al., *Bayesian cluster identification in single-molecule localization microscopy data.* Nat Methods, 2015. **12**(11): p. 1072-6.

264. Griffie, J., et al., *Topographic prominence as a method for cluster identification in single-molecule localisation data.* J Biophotonics, 2015. **8**(11-12): p. 925-34.

265. Silva, A.L. and L. Romao, *The mammalian nonsense-mediated mRNA decay pathway: to decay or not to decay! Which players make the decision?* FEBS Lett, 2009. **583**(3): p. 499-505.

266. Hamshere, M.G., et al., *Transcriptional abnormality in myotonic dystrophy affects DMPK but not neighboring genes.* Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7394-9.

267. Messina, G., et al., *Nfix regulates fetal-specific transcription in developing skeletal muscle.* Cell, 2010. **140**(4): p. 554-66.

268. Prokic, I., B.S. Cowling, and J. Laporte, *Amphiphysin 2 (BIN1) in physiology and diseases.* J Mol Med (Berl), 2014. **92**(5): p. 453-63.

269. Tanner, M.K., Z. Tang, and C.A. Thornton, *Targeted splice sequencing reveals RNA toxicity and therapeutic response in myotonic dystrophy.* Nucleic Acids Res, 2021. **49**(4): p. 2240-2254.

270. Chapuis, J., et al., *Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology.* Mol Psychiatry, 2013. **18**(11): p. 1225-34.

271. Calafate, S., et al., *Loss of Bin1 Promotes the Propagation of Tau Pathology.* Cell Rep, 2016. **17**(4): p. 931-940.

272. Spillantini, M.G. and M. Goedert, *Tau pathology and neurodegeneration.* Lancet Neurol, 2013. **12**(6): p. 609-22.

273. Hu, W., et al., *Hyperphosphorylation determines both the spread and the morphology of tau pathology.* Alzheimers Dement, 2016. **12**(10): p. 1066-1077.

274. Guo, T., W. Noble, and D.P. Hanger, *Roles of tau protein in health and disease.* Acta Neuropathol, 2017. **133**(5): p. 665-704.

275. Shahani, N. and R. Brandt, *Functions and malfunctions of the tau proteins.* Cell Mol Life Sci, 2002. **59**(10): p. 1668-80.

276. Iqbal, K., et al., *Tau pathology in Alzheimer disease and other tauopathies.* Biochim Biophys Acta, 2005. **1739**(2-3): p. 198-210.

277. Goedert, M., M.G. Spillantini, and R.A. Crowther, *Tau proteins and neurofibrillary degeneration.* Brain Pathol, 1991. **1**(4): p. 279-86.

278. Caillet-Boudin, M.L., et al., *Brain pathology in myotonic dystrophy: when tauopathy meets spliceopathy and RNAopathy.* Front Mol Neurosci, 2014. **6**: p. 57.

279. Maurage, C.A., et al., *Similar brain tau pathology in DM2/PROMM and DM1/Steinert disease.* Neurology, 2005. **65**(10): p. 1636-8.

280. Hubert, L., Jr., et al., *Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1.* Hum Mol Genet, 2011. **20**(24): p. 4822-30.

281. Jiang, L., et al., *NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing.* Nat Struct Mol Biol, 2017. **24**(10): p. 816-824.

282. Prasanth, K.V., et al., *Regulating gene expression through RNA nuclear retention.* Cell, 2005. **123**(2): p. 249-63.

283. Bajc Cesnik, A., et al., *Nuclear RNA foci from C9ORF72 expansion mutation form paraspeckle-like bodies.* J Cell Sci, 2019. **132**(5).

284. Oh, J., et al., *Widespread Alternative Splicing Changes in Metastatic Breast Cancer Cells.* Cells, 2021. **10**(4).

285. Santoro, M., et al., *Analysis of MTMR1 expression and correlation with muscle pathological features in juvenile/adult onset myotonic dystrophy type 1 (DM1) and in myotonic dystrophy type 2 (DM2).* Exp Mol Pathol, 2010. **89**(2): p. 158-68.

286. Buj-Bello, A., et al., *Muscle-specific alternative splicing of myotubularin-related 1 gene is impaired in DM1 muscle cells.* Hum Mol Genet, 2002. **11**(19): p. 2297-307.

287. McKinsey, T.A., C.L. Zhang, and E.N. Olson, *MEF2: a calcium-dependent regulator of cell division, differentiation and death.* Trends Biochem Sci, 2002. **27**(1): p. 40-7.

288. Bachinski, L.L., et al., *Altered MEF2 isoforms in myotonic dystrophy and other neuromuscular disorders.* Muscle Nerve, 2010. **42**(6): p. 856-63.

289. Ichikawa, K., et al., *Coincidence of neurofibromatosis and myotonic dystrophy in a kindred.* J Med Genet, 1981. **18**(2): p. 134-8.

290. Terracciano, C., et al., *Vitamin D deficiency in myotonic dystrophy type 1.* J Neurol, 2013. **260**(9): p. 2330-4.

291. Martorell, L., et al., *Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients.* Hum Mol Genet, 1998. **7**(2): p. 307-12.

292. Fardaei, M., et al., *Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells.* Hum Mol Genet, 2002. **11**(7): p. 805-14.

293. Langlois, M.A., et al., *Hammerhead ribozyme-mediated destruction of nuclear foci in myotonic dystrophy myoblasts.* Mol Ther, 2003. **7**(5 Pt 1): p. 670-80.

294. Chen, G., et al., *Phenylbutazone induces expression of MBNL1 and suppresses formation of MBNL1-CUG RNA foci in a mouse model of myotonic dystrophy.* Scientific Reports, 2016. **6**(1): p. 25317.

295. Amberg, D.C., A.L. Goldstein, and C.N. Cole, *Isolation and characterization of RAT1: an essential gene of Saccharomyces cerevisiae required for the efficient nucleocytoplasmic trafficking of mRNA.* Genes Dev, 1992. **6**(7): p. 1173-89.

296. Chang, J.H., et al., *Structural and biochemical studies of the 5'-->3' exoribonuclease Xrn1.* Nat Struct Mol Biol, 2011. **18**(3): p. 270-6.

297. Bonneau, F., et al., *The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation.* Cell, 2009. **139**(3): p. 547-59.

298. Januszyk, K. and C.D. Lima, *The eukaryotic RNA exosome.* Curr Opin Struct Biol, 2014. **24**: p. 132-40.

299. Wolin, S.L., S. Sim, and X. Chen, *Nuclear noncoding RNA surveillance: is the end in sight?* Trends Genet, 2012. **28**(7): p. 306-13.

300. Gudde, A., et al., *Trinucleotide-repeat expanded and normal DMPK transcripts contain unusually long poly(A) tails despite differential nuclear residence.* Biochim Biophys Acta Gene Regul Mech, 2017. **1860**(6): p. 740-749.

301. Hilleren, P. and R. Parker, *Defects in the mRNA export factors Rat7p, Gle1p, Mex67p, and Rat8p cause hyperadenylation during 3'-end formation of nascent transcripts.* RNA, 2001. **7**(5): p. 753-64.

302. Kumar, G.R. and B.A. Glaunsinger, *Nuclear import of cytoplasmic poly(A) binding protein restricts gene expression via hyperadenylation and nuclear retention of mRNA.* Mol Cell Biol, 2010. **30**(21): p. 4996-5008.

303. Davidson, L., A. Kerr, and S. West, *Co-transcriptional degradation of aberrant pre-mRNA by Xrn2.* EMBO J, 2012. **31**(11): p. 2566-78.

304. Lloret-Llinares, M. and T.H. Jensen, *Global Identification of Human Exosome Substrates Using RNA Interference and RNA Sequencing.* Methods Mol Biol, 2020. **2062**: p. 127-145.

305. Nagarajan, V.K., et al., *XRN 5'-->3' exoribonucleases: structure, mechanisms and functions.* Biochim Biophys Acta, 2013. **1829**(6-7): p. 590-603.

306. Miki, T.S. and H. Grosshans, *The multifunctional RNase XRN2.* Biochem Soc Trans, 2013. **41**(4): p. 825-30.

307. Tomecki, R., et al., *The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L.* EMBO J, 2010. **29**(14): p. 2342-57.

308. Leon, K., et al., *Zika Virus Infection Prevents Host mRNA Nuclear Export by Disrupting UPF1 Function.* bioRxiv, 2020: p. 2020.12.03.410837.

309. Tretyakova, I., et al., *Nuclear export factor family protein participates in cytoplasmic mRNA trafficking.* J Biol Chem, 2005. **280**(36): p. 31981-90.

310. Saito, K., *[Myotonic dystrophy].* Ryoikibetsu Shokogun Shirizu, 2001(34 Pt 2): p. 260-1.

311. Lopez-Martinez, A., et al., *An Overview of Alternative Splicing Defects Implicated in Myotonic Dystrophy Type I.* Genes (Basel), 2020. **11**(9).

312. Charizanis, K., et al., *Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy.* Neuron, 2012. **75**(3): p. 437-50.

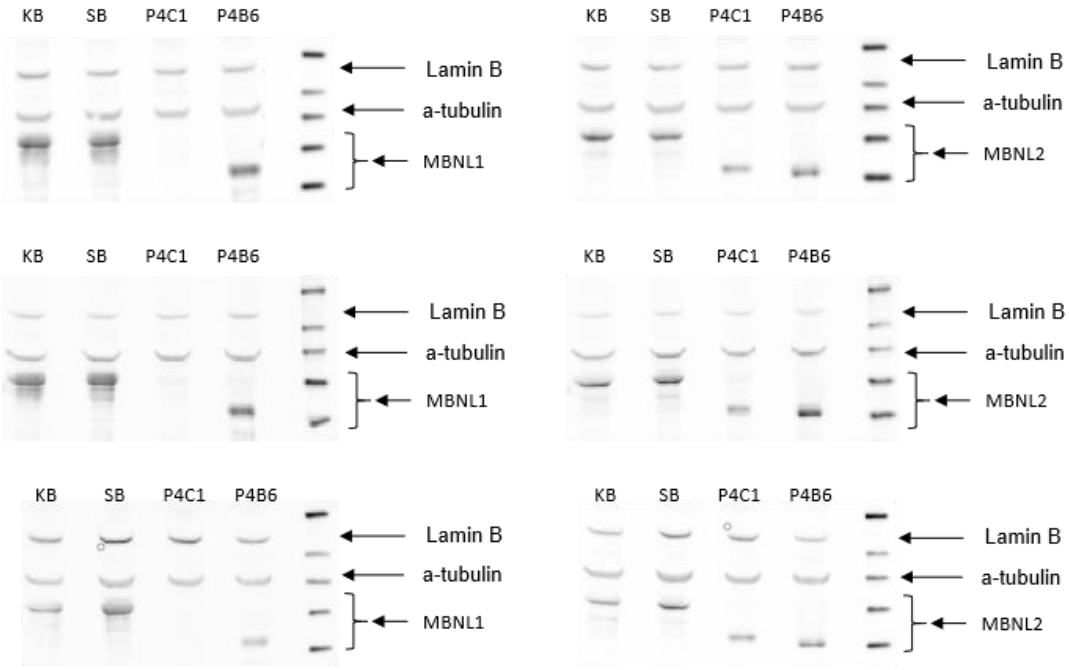
313. Bargiela, A., et al., *Increased Muscleblind levels by chloroquine treatment improve myotonic dystrophy type 1 phenotypes in in vitro and in vivo models.* Proc Natl Acad Sci U S A, 2019. **116**(50): p. 25203-25213.

314. Jahromi, A.H., et al., *A novel CUG(exp).MBNL1 inhibitor with therapeutic potential for myotonic dystrophy type 1.* ACS Chem Biol, 2013. **8**(5): p. 1037-43.

315. Helmerich, D.A., et al., *Photoswitching fingerprint analysis bypasses the 10-nm resolution barrier.* Nat Methods, 2022. **19**(8): p. 986-994.

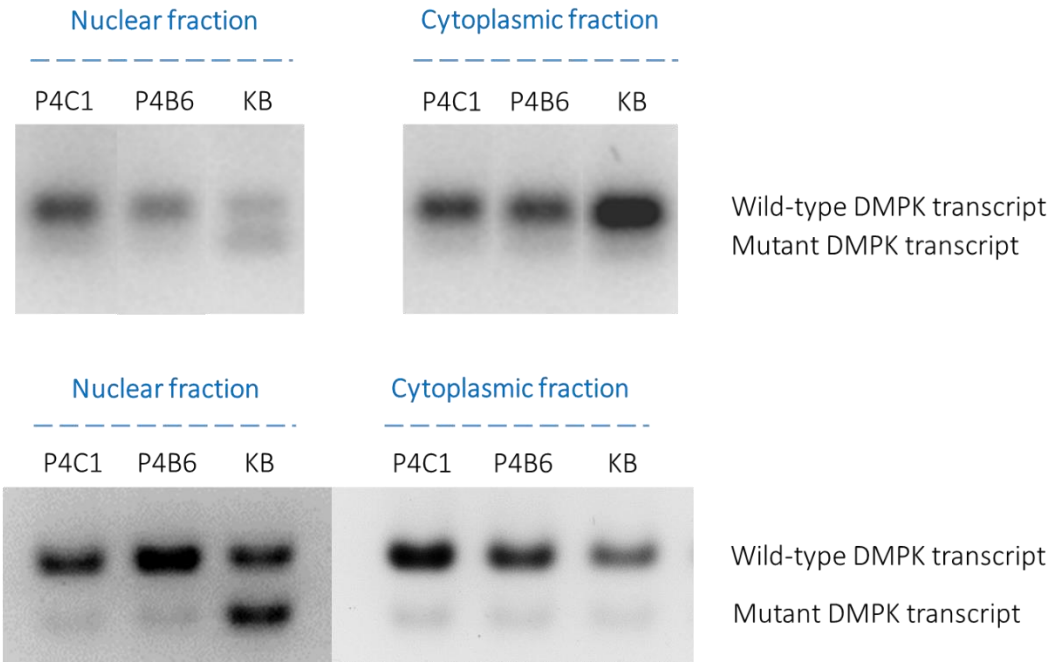
316. Scholes, G.D., *Long-range resonance energy transfer in molecular systems.* Annu Rev Phys Chem, 2003. **54**: p. 57-87.

Figure 0-1 Assessment of double knock-out of MBNL1 and 2 by Western blotting in triplicate.



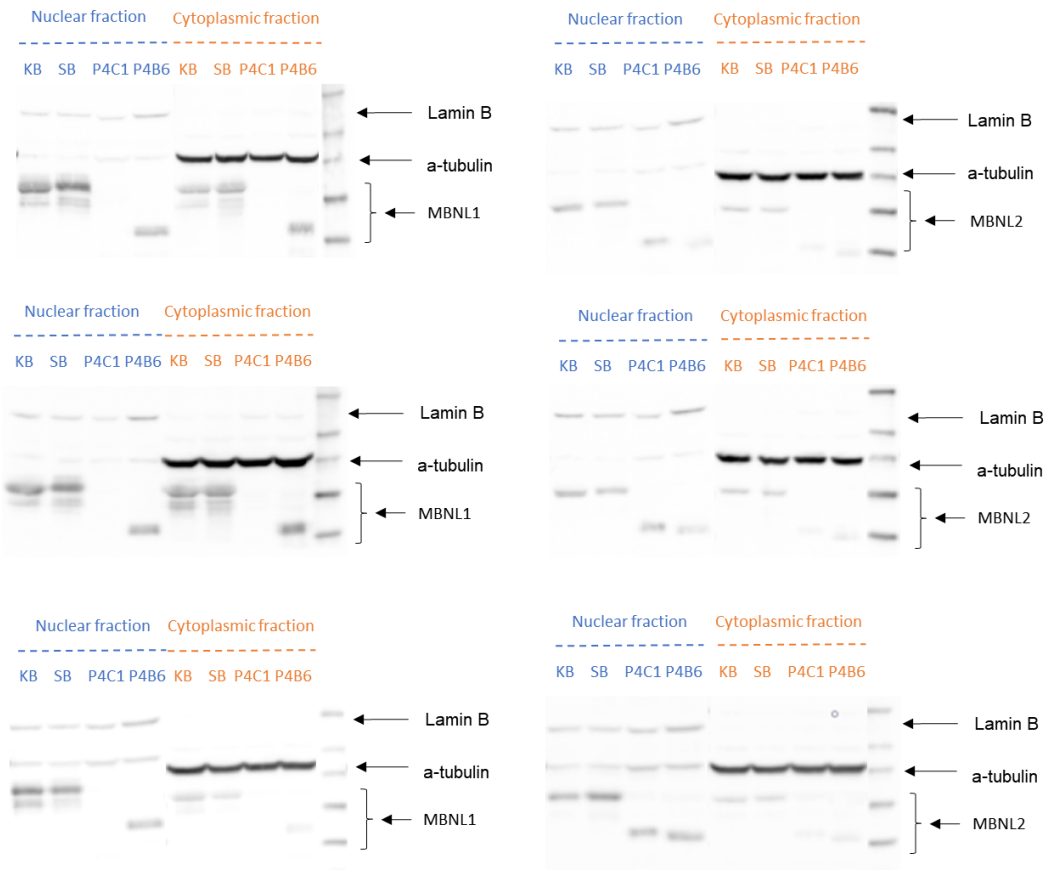
Total protein extracts from KB (a DM1 patient cell line), SB (a non-DM control cell line), as well as P4C1 and P4B6 (MBNL-deficient DM1 cell lines) were analysed by western blot to detect the expression of MBNL1 and MBNL2 proteins. α-tubulin was used as a loading control. The experiment was performed in triplicate to ensure the reproducibility of the results.

Figure 0-2 Assessment of the effect of MBNL1 and 2 double knock-out on the (CUG)^{exp} DMPK transcript via BmpI digestion method (in duplicate).



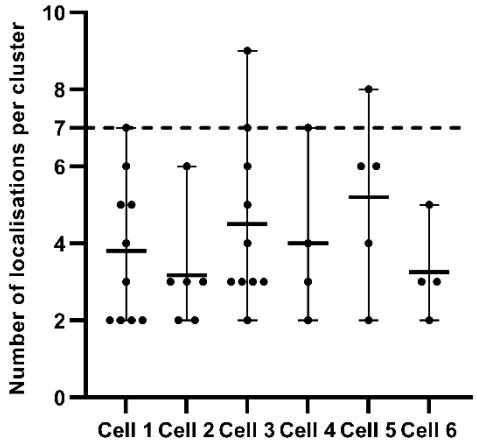
Ethidium bromide-stained gels showing RT–PCR products from nuclear and cytoplasmic RNA in KB, P4C1 and P4B6 cells. A fragment of *DMPK* mRNA was reverse-transcribed, amplified and digested with the BpmI restriction enzyme, and the resulting fragments were separated by gel electrophoresis. The experiment was performed in duplicate to ensure the reproducibility of the results.

Figure 0-3 Subcellular Localisation of MBNL1 and 2 proteins detected by Western blotting in triplicate.



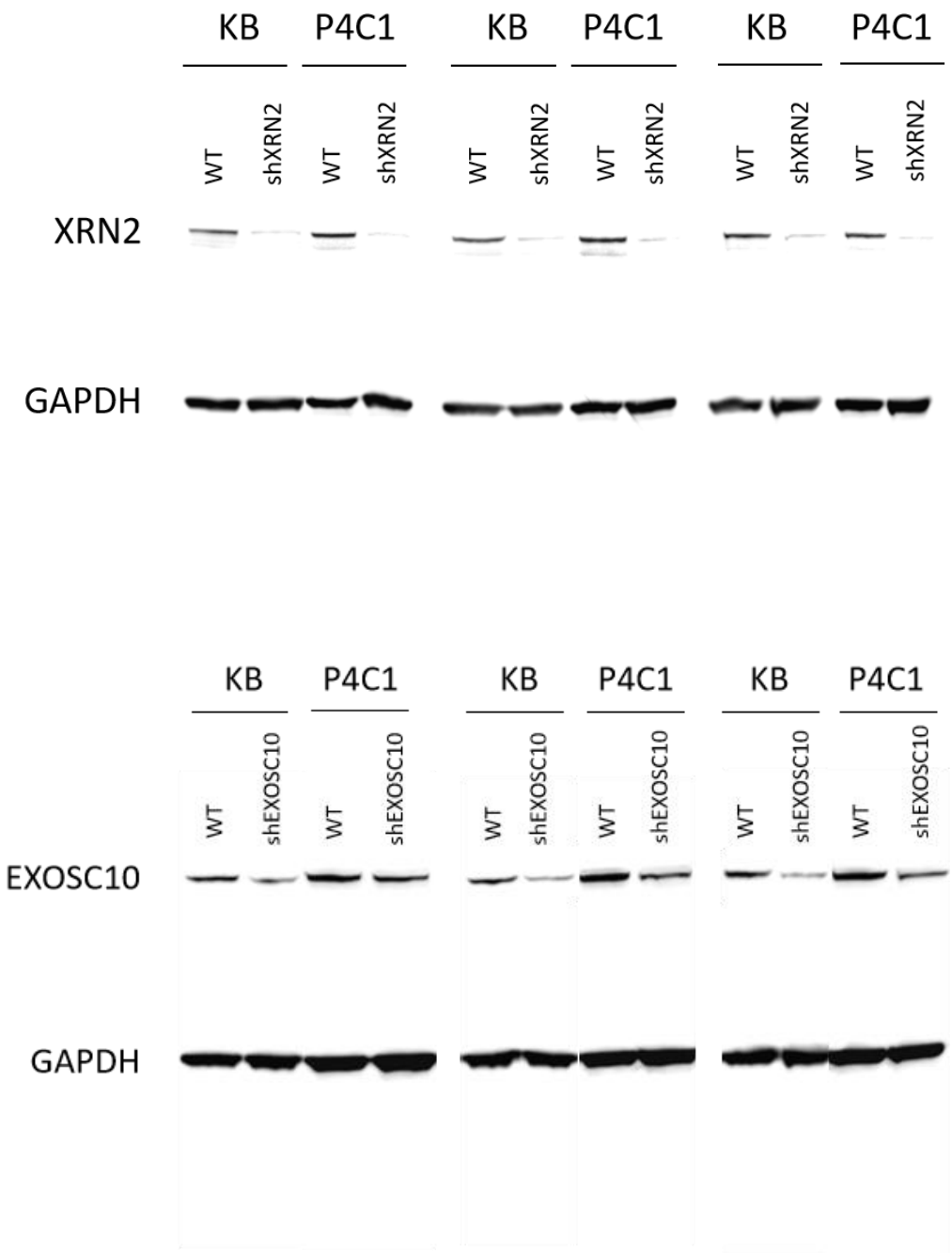
Western blot analysis was performed to investigate the subcellular distribution of MBNL1 and MBNL2 proteins in different cell lines. Nuclear and cytoplasmic protein fractions were isolated from KB, SB, P4C1, and P4B6 cells, and protein expression levels were probed using specific antibodies. Lamin-B and α-tubulin were used as loading controls for the nuclear and cytoplasmic fractions, respectively. The experiment was performed in triplicate to ensure reproducibility of the results.

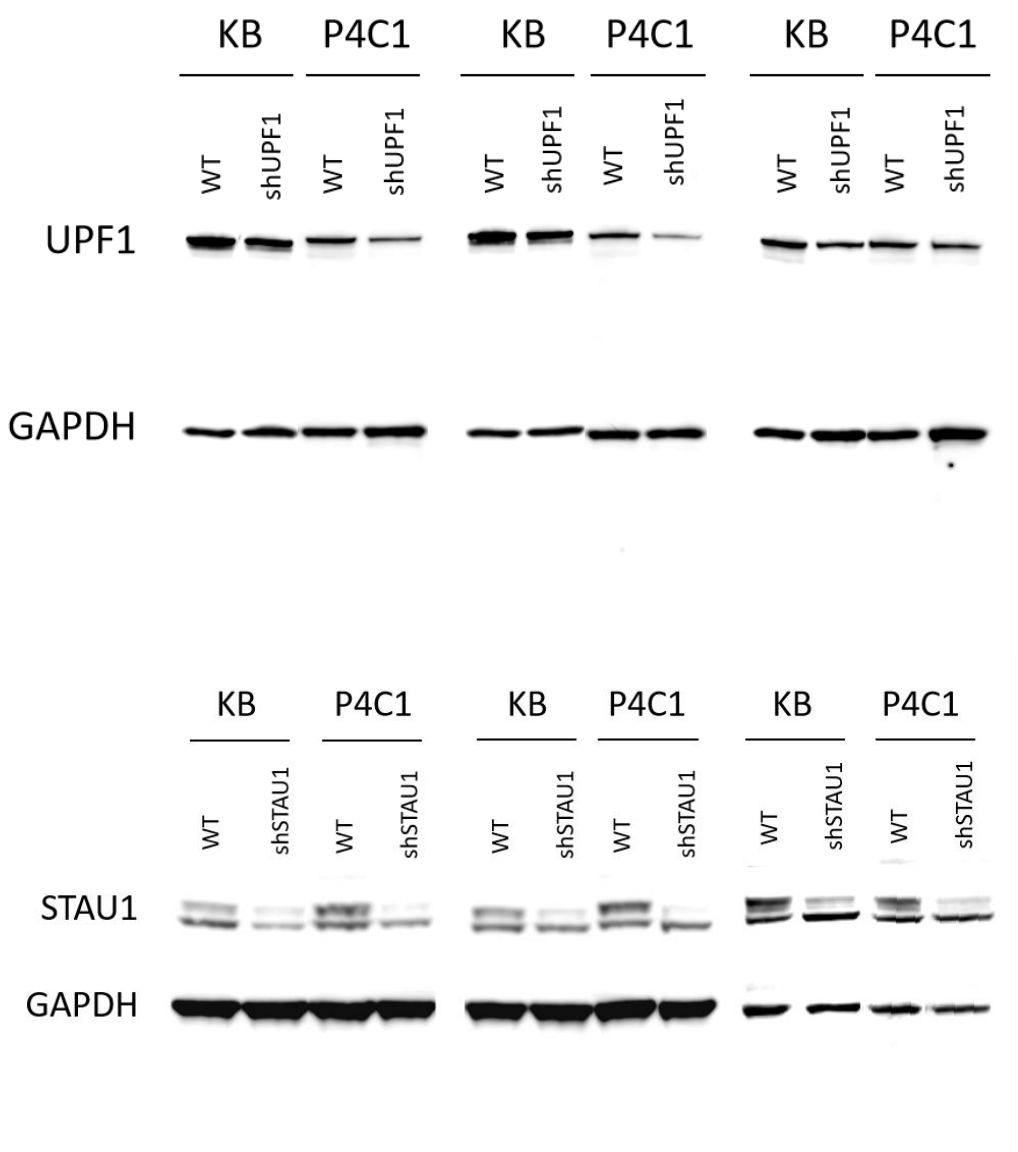
Figure 0-4 Signal detection in STORM microscopy by using the scramble probe.



The diagram illustrates the number of localisations in each cluster (represented by a dot) that were detected by using the Atto647n scramble probe under the same imaging conditions as the Atto647n-labelled 5’/3’ single-copy probe sets.

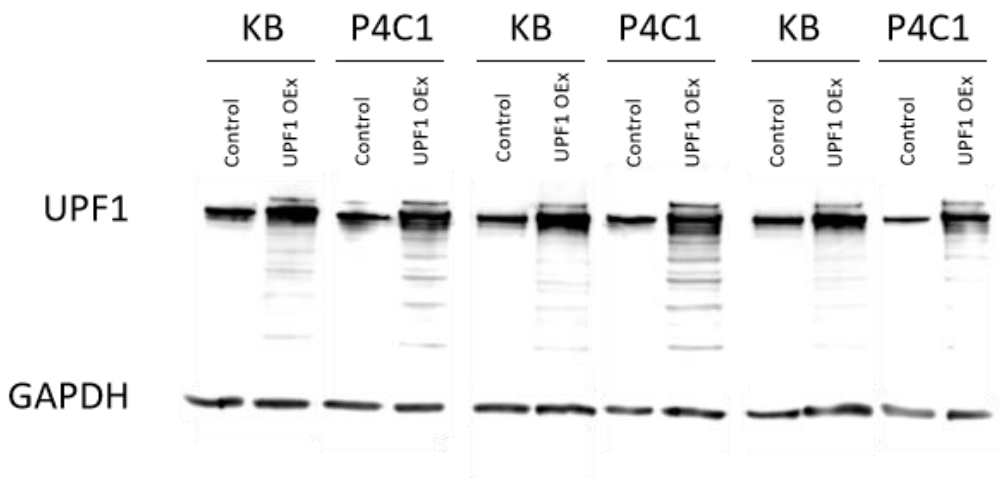
Figure 0-5 Assessment of the knock-down efficiency of RNA decay factors by Western blotting in triplicate.





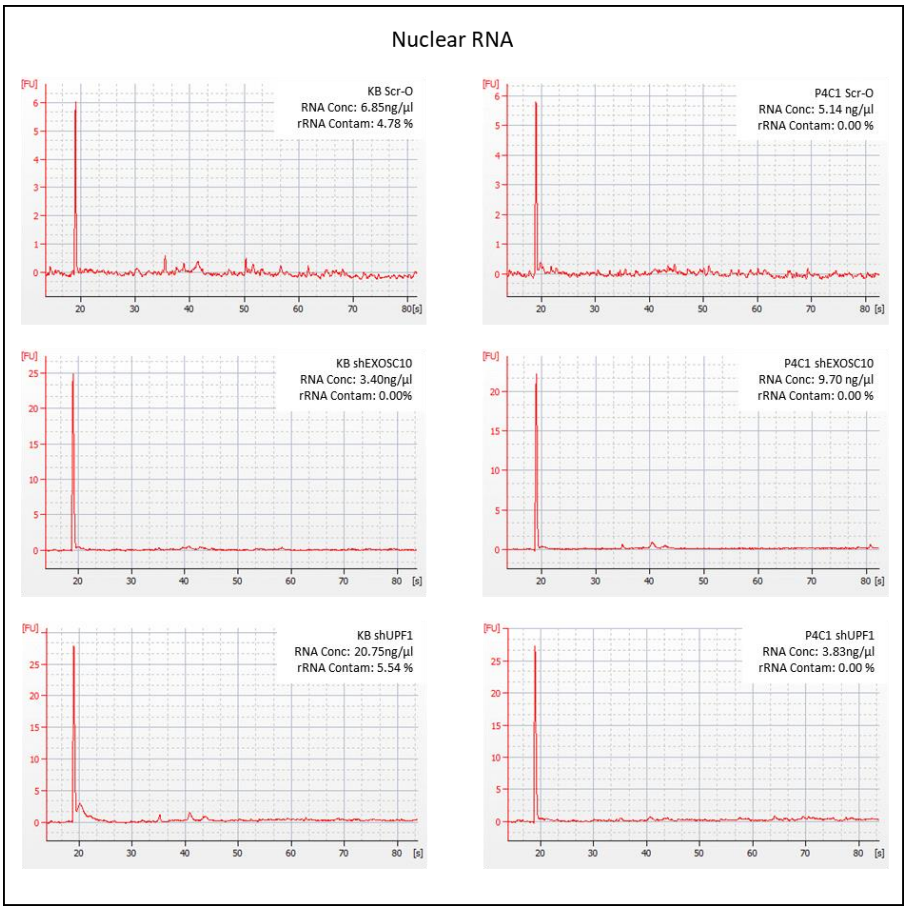
Western blotting images of samples blotted with anti-XRN2, anti-EXOSC10, anti-UPF1 and anti-STAU1 antibody and re-blotted with anti-GAPDH antibody on the same membrane. The experiment was performed in triplicate to ensure reproducibility of the results.

Figure 0-6 Assessment of the efficiency of UPF1 over-expression by Western blotting in triplicate.

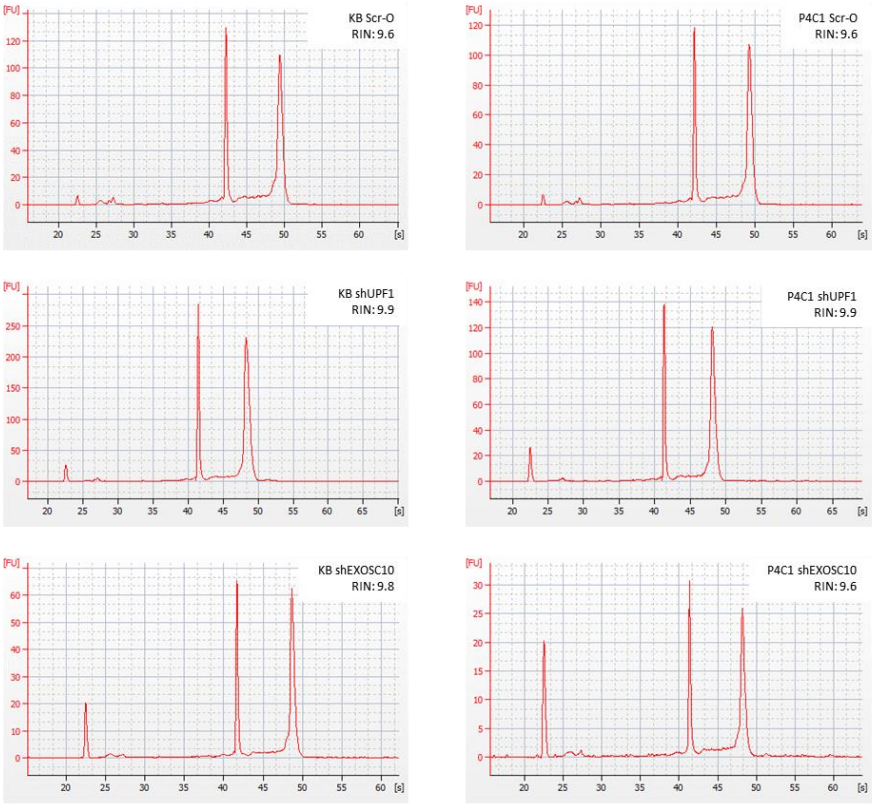


Western blotting images of samples blotted with anti-UPF1 antibody and re-blotted with anti-GAPDH antibody on the same membrane. The experiment was performed in triplicate to ensure reproducibility of the results.

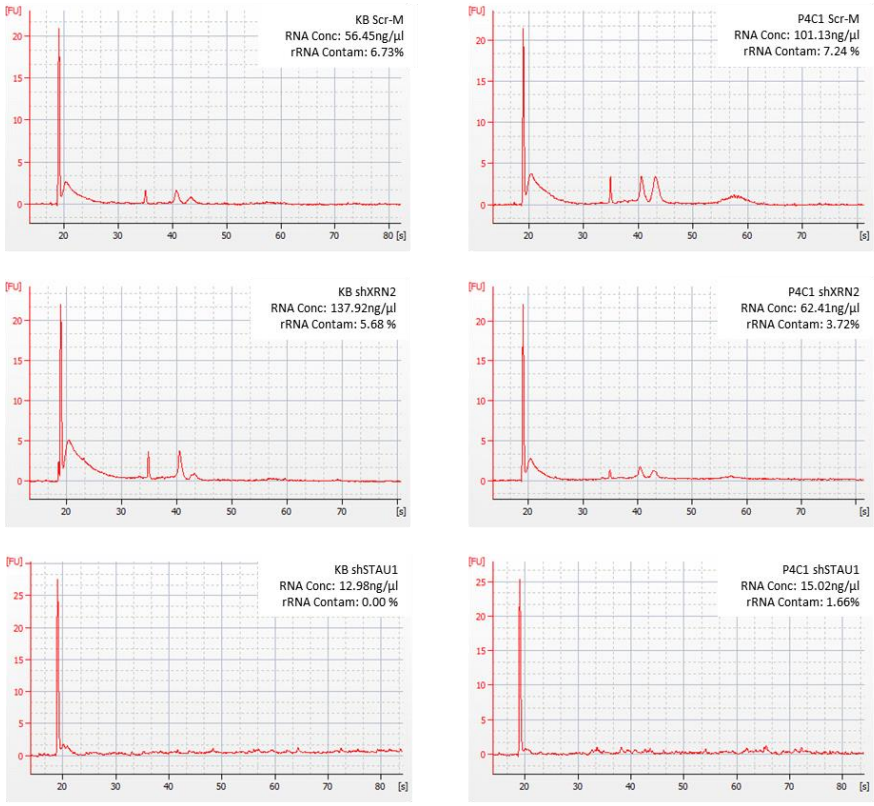
Figure 0-7 Assessment of nuclear and cytoplasmic RNA quality for RNA samples utilised in Chapter 5.

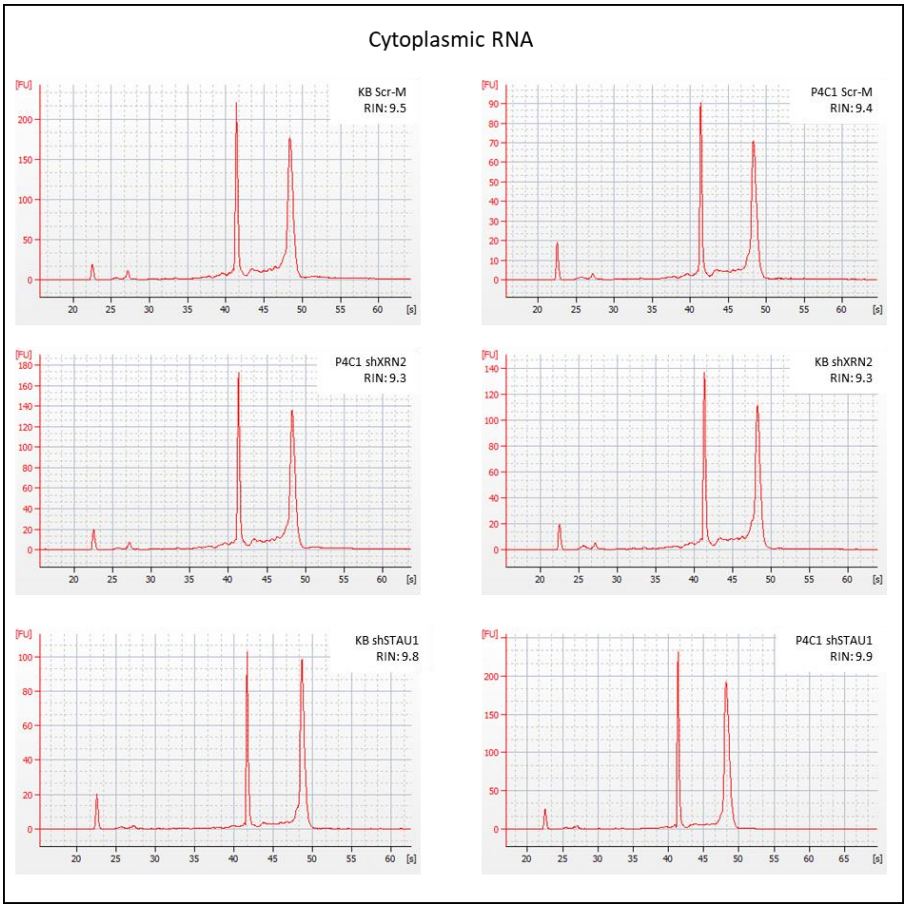


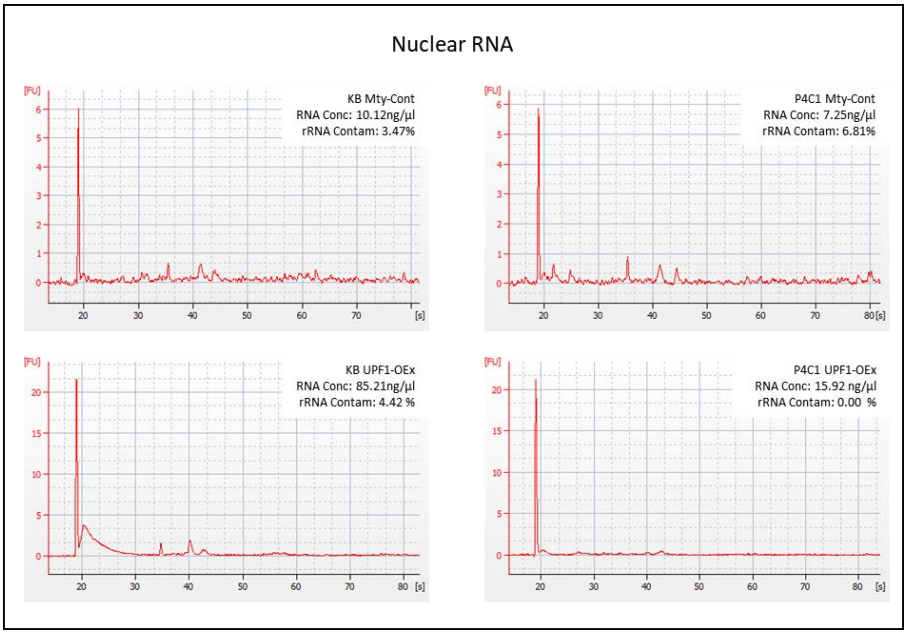
Cytoplasmic RNA

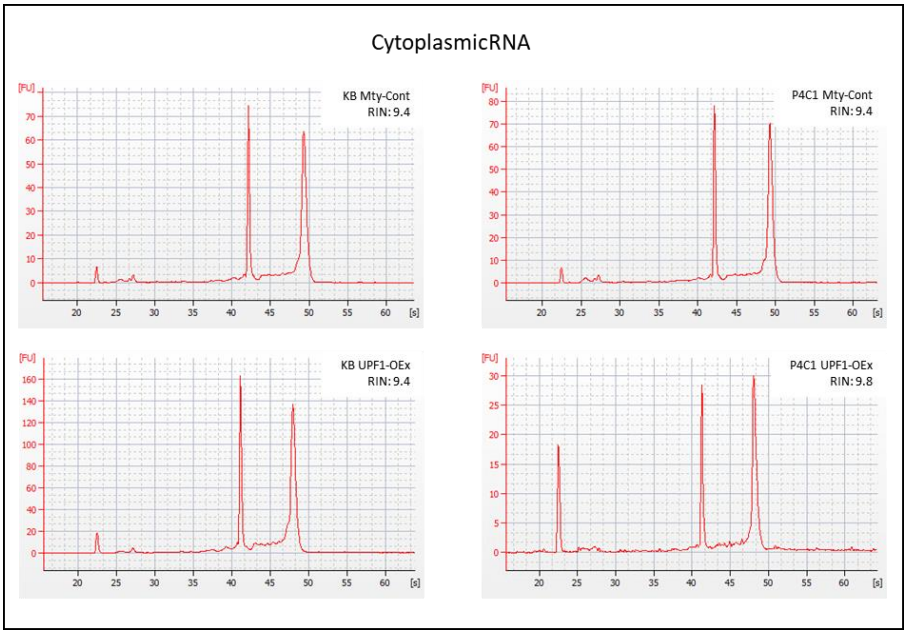


Nuclear RNA



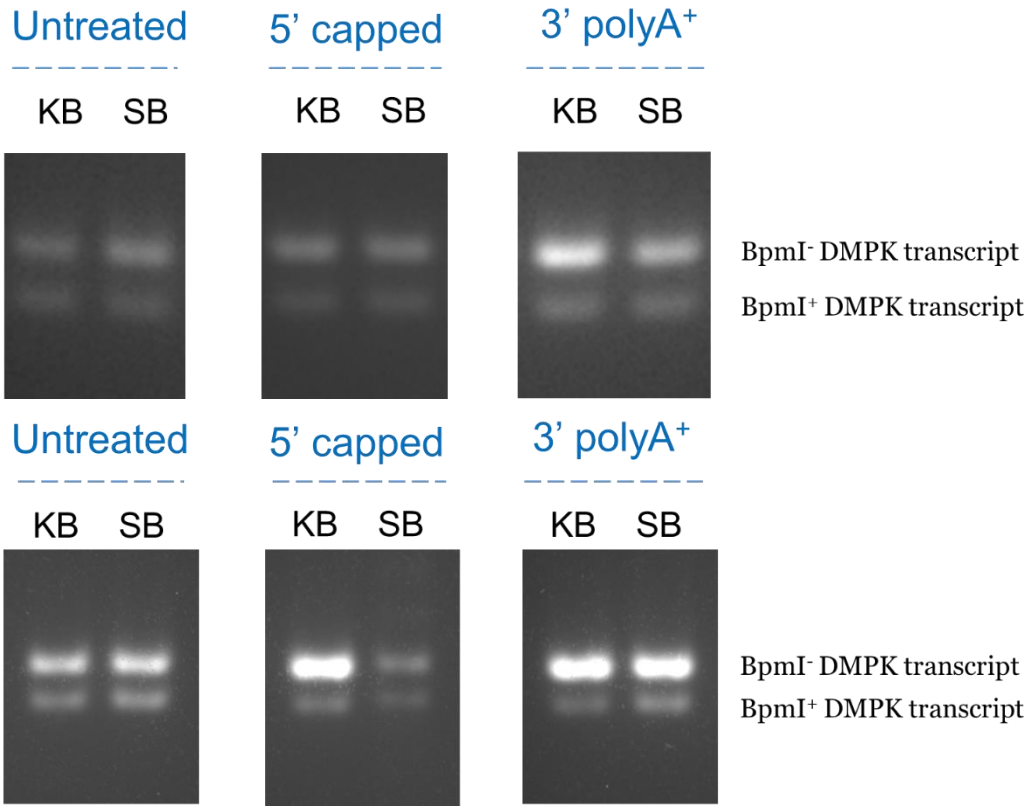

Cytoplasmic RNA
KB Scr-M
RIN: 9.5
P4C1 Scr-M
RIN: 9.4
P4C1 shXRN2
RIN: 9.3
KB shXRN2
RIN: 9.3
KB shSTAU1
RIN: 9.8
P4C1 shSTAU1
RIN: 9.9


Nuclear RNA
KB Mty-Cont
RNA Conc: 10.12ng/µl
rRNA Contam: 3.47%
P4C1 Mty-Cont
RNA Conc: 7.25ng/µl
rRNA Contam: 6.81%
KB UPF1-OEx
RNA Conc: 85.21ng/µl
rRNA Contam: 4.42 %
P4C1 UPF1-OEx
RNA Conc: 15.92 ng/µl
rRNA Contam: 0.00 %



Electropherogram of purified RNA from the cytoplasm and the nucleus. The x-axis shows migration time (in seconds), and the y-axis shows fluorescence intensity (in relative fluorescence units, RFU). The two prominent peaks in the electropherogram correspond to 18s and 28s ribosomal RNA (rRNA) molecules. KB Scr-O and P4C1 Scr-O served as scramble controls for shRNA-mediated knock-down of *UPF1* and *EXOSC10* in KB and P4C1, respectively. KB Scr-M and P4C1 Scr-M served as scramble controls for shRNA-mediated knock-down of *XRN2* and *STAU1* in KB and P4C1, respectively. KB Mty-Cont and P4C1 Mty-Cont served as empty controls for lentivirus-mediated UPF1 over-expression in KB and P4C1, respectively.

Figure 0-8 Assessment of poly(A) tail and 5′ cap structure status of the (CUG)^{exp} DMPK transcript via the BpmI digestion method.



Ethidium bromide-stained gels showing RT–PCR products from 3′ poly(A)$^+$, 5′ capped or total RNA in KB and SB cells. A fragment of *DMPK* mRNA was reverse-transcribed, amplified and digested with the BpmI restriction enzyme, and the resulting fragments were separated by gel electrophoresis (BpmI$^-$, PCR band representing *DMPK* transcripts without the BpmI recognition site and BpmI$^+$, representing with the BpmI recognition site). The experiment was performed in duplicate to ensure the reproducibility of the results.

Figure 0-9 Probes positioning in the DMPK cDNA.

>ENST00000291270.9 DMPK-201 cDNA:protein_coding
Exon 1:
AGAGAGGCCCAGGGGACAGCCAGGGACAGGCAGACATGCAGCCAGGGCTCCAGGGCCTGGACAGGGGCTGCCAGGCCCTGTGACAGGAG
GACCCCGAGCCCCCGGCCCGGGGAGGGGCCATGGTGCTGCCTGTCCAACATGTCAGCCGAGGTGCGGCTGAGGCGGCTCCAGCAGCTGGTG
TTGGACCCGGGCTTCCTGGGGCTGGAGCCCCTGCTCGACCTTCTCCTGGGCGTCCACCAGGAGCTGGGCGCCTCCGAACTGGCCCAGGACAA
GTACGTGGCCGACTTCTTGCAGTGGG
Exon 2:
CGGAGCCCATCGTGGTGAGGCTTAAGGAGGTCCGACTGCAGAGGGACGACTTCGAGATTCTGAAGGTGATCGGACGCGGGGCGTTCAGCGA
G
Exon 3:
GTAGCGGTAGTGAAGATGAAGCAGACGGGCCAGGTGTATGCCATGAAGATCATGAACAAGTGGGACATGCTGAAGAGGGGCGAG
Exon 4:
GTGTCGTGCTTCCGTGAGGAGAGGGACGTGTTGGTGAATGGGGACCGGCGGTGGATCACGCAGCTGCACTTCGCCTTCCAGGATGAGAACTA
CCTG
Exon 5:
TACCTGGTCATGGAGTATTACGTGGGCGGGGACCTGCTGACACTGCTGAGCAAGTTTGGGGAGCGGATTCCGGCCGAGATGGCGCGCTTCTA
CCTGGCGGAGATTGTCATGGCCATAGACTCGGTGCACCGGCTTGGCTACGTGCACAG
Exon 6:
GGACATCAAACCCGACAACATCCTGCTGGACCGCTGTGGCCACATCCGCCTGGCCGACTTCGGCTCTTGCCTCAAGCTGCGGGCAGATGGAAC
G
Exon 7:
GTGCGGTCGCTGGTGGCTGTGGGCACCCCAGACTACCTGTCCCCCGAGATCCTGCAGGCTGTGGGCGGTGGGCCTGGGACAGGCAGCTACG
GGCCCGAGTGTGACTGGTGGGCGCTGGGTGTATTCGCCTATGAAATGTTCTATGGGCAGACGCCCTTCTACGCGGATTCCACGGCGGAGACC
TATGGCAAGATCGTCCACTACAAG
Exon 8:
GAGCACCTCTCTCTGCCGCTGGTGGACGAAGGGGTCCCTGAGGAGGCTCGAGACTTCATTCAGCGGTTGCTGTGTCCCCCGGAGACACGGCT
GGGCCGGGGTGGAGCAGGCGACTTCCGGACACATCCCTTCTTCTTTGGCCTCGACTGGGATGGTCTCCGGGACAGCGTGCCCCCCTTTACACC
GGATTTCGAAGGTGCCACCGACACATGCAACTTCGACTTGGTGGAGGACGGGCTCACTGCCATGGTGAGCGGGGGCGGG
Exon 9:
GAGACACTGTCGGACATTCGGGAAGGTGCGCCGCTAGGGGTCCACCTGCCTTTTGTGGGCTACTCCTACTCCTGCATGGCCCTCAG
Exon 10:
GGACAGTGAGGTCCCAGGCCCCACACCCATGGAACTGGAGGCCGAGCAGCTGCTTGAGCCACACGTGCAAGCGCCCAGCCTGGAGCCCTCG
GTGTCCCCACAGGATGAAACA
Exon 11:
GCTGAAGTGGCAGTTCCAGCGGCTGTCCCTGCGGCAGAGGCTGAGGCCGAGGTGACGCTGCGGGAGCTCCAGGAAGCCCTGGAGGAGGAG
GTGCTCACCCGGCAGAGCCTGAGCCGGGAGATGGAGGCCATCCGCACGGACAACCAGAACTTCGCCAG
Exon 12:
TCAACTACGCGAGGCAGAGGCTCGGAACCGGGACCTAGAGGCACACGTCCGGCAGTTGCAGGAGCGGATGGAGTTGCTGCAGGCAGAGGG
AGCCACAG
Exon 13:
CTGTCACGGGGGTCCCCAGTCCCCGGGCCACGGATCCACCTTCCCAT
Exon 14:
CTAGATGGCCCCCCGGCCGTGGCTGTGGGCCAGTGCCCGCTGGTGGGGCCAGGCCCCATGCACCGCCGCCACCTGCTGCTCCCTGCCAGG
Exon 15:
GTCCCTAGGCCTGGCCTATCGGAGGCGCTTTCCCTGCTCCTGTTCGCCGTTGTTCTGTCTCGTGCCGCCGCCCTGGGCTGCATTGGGTTGGTGG
CCCACGCCGGCCAACTCACCGCAGTCTGGCGCCGCCCAGGAGCCGCCCGCGCTCCCTGAACCCTAGAACTGTCTTCGACTCCGGGGCCCCGTT
GGAAGACTGAGTGCCCGGGGCACGGCACAGAAGCCGCGCCCACCGCCTGCCAGTTCACAACCGCTCCGAGCGTGGGTCTCCGCCCAGCTCCA
GTCCTGTGATCCGGGCCCGCCCCCTAGCGGCCGGGGAGGGAGGGGCCGGGTCCGCGGCCGGCGAACGGGGCTCGAAGGGTCCTTGTAGCC
GGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGGGGGGGATCACAGACCATTTCTTTCTTT
CGGCCAGGCTGAGGC CCTGACGTGGATGGGCAAACTGCAGGCCTGGGAAGGCAGCAAGCCGGGCCGTCCGTGTTCCATCCTCCACGCACCC
CCACCTATCGTTGGTTCGCAAAGTGCAAAGCTTTCTTGTGCATGACGCCCTGCTCTGGGGAGCGTCTGGCGCGATCTCTGCCTGCTTACTCGGG
AAATTTGCTTTTGCCAAACCCGCTTTTTCGGGGATCCCGCGCCCCCCTCCTCACTTGCGCTGCTCTCGGAGCCCCAGCCGGCTCCGCCCGCTTCG
GCGGTTTGGATATTTATTGACCTCGTCCTCCGACTCGCTGACAGGCTACAGGACCCCCAACAACCCCAATCCACGTTTTGGATGCACTGAGACC
CCGACATTCCTCGGTATTTATTGTCTGTCCCCACCTAGGACCCCCACCCCCGACCCTCGCGAATAAAAGGCCCTCCATCTGCCCAAA

The figure illustrates the positions of 5' and 3' single-copy probes employed in Super-Resolution Single Molecule Localization Microscopy (SMLM). The *DMPK* cDNA is depicted, with the 5' single-copy probes highlighted in yellow and the 3' single-copy probes highlighted in blue (Probes spanning two exons are also underlined). The CTG repeat region is specifically indicated in red.