

The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Investigating genome engineering of myotonic dystrophy cells and small molecules as potential therapeutic agents

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

Marta López Morató, BSc

The University of Nottingham

November, 2018

Table of contents

List of Figures **xii**

List of Tables **xv**

Abstract **i**

1 Introduction **1**

1.1 History of Myotonic Dystrophy 1

1.2 The Phenotype of Myotonic Dystrophy 3

1.2.1 Congenital DM1 4

1.2.2 Childhood DM1 5

1.2.3 Classical/Adult onset DM1 5

1.2.4 Minimal/Late onset DM1 6

1.2.5 DM2 7

1.3 The Genetic Basis of Myotonic Dystrophy 8

1.3.1 Genetic basis of DM1 8

1.3.2 Genetic basis of DM2 10

1.4 Molecular Pathomechanism of Myotonic Dystrophy 12

1.4.1 *DMPK* and *CNBP* haploinsufficiency 12

1.4.2 Alteration of chromatin structure 13

1.4.3 RNA-induced toxicity 15

1.4.4 RAN translation 17
1.5 The Molecular and Cell Biology of Myotonic Dystrophy 19
1.5.1 CELF1 in DM 20
1.5.2 Muscleblind in DM 22
1.5.3 Aberrant alternative splicing in DM 23
1.5.4 Deregulation of microRNAs 26
1.6 Animal Models for Myotonic Dystrophy 27
1.6.1 Inactivation of genes of the DM1 or DM2 locus 28
1.6.2 Transgenic models of toxic RNA gain-of-function . . . 28
1.6.3 Mouse models for abnormal alternative splicing 30
1.6.4 Mouse models of DM2 30
1.7 Therapies for Myotonic Dystrophy 31
1.7.1 Targeting the DNA repeats 31
1.7.2 Targeting the toxic RNA 33
1.7.3 Targeting protein-RNA binding 36
1.7.4 Altering levels of RNA binding proteins 38
1.7.5 Targeting the mispliced genes 39
1.7.6 Additional targets 39
1.7.7 Kinase Inhibitors 40
1.8 CDK12 in Myotonic Dystrophy 41
1.9 Genome Engineering using CRISPR/Cas9 System 43
1.10 Inducible Gene Expression Systems 52
1.11 Aims of Study 55

2 Materials and Methods **56**

2.1 Tissue culture 56

2.1.1 Cell lines 56
2.1.2 Media, solutions and antibiotics 56
2.1.3 Cell culture 57
2.1.4 Cell passaging 57
2.1.5 Cell count 57
2.1.6 Cryopreservation of cells 58
2.1.7 Defrosting cells from liquid nitrogen 58
2.1.8 Compound treatment of DM cells 58
2.1.9 MTT assay 59
2.1.10 RNA Fluorescence *In Situ* Hybridization 59
2.1.11 Immunocytochemistry (ICC) 60
2.1.12 Co-localisation experiments (ICC + FISH) 61
2.2 Microbiology and Cloning Techniques 62
2.2.1 Media 62
2.2.2 Antibiotics 62
2.2.3 Restriction digestion 62
2.2.4 Vector-Insert Ligation 63
2.2.5 pGEM-T-easy Vector 63
2.2.6 Transformation 64
2.2.7 Preparation of Glycerol Stocks 65
2.2.8 Miniprep Plasmid DNA Isolation 65
2.2.9 Maxiprep and Endotoxin-free Maxiprep DNA Isolation . 65
2.2.10 Analysis of Plasmid DNA Integrity and Concentration . 66
2.3 Molecular Biology Techniques 66
2.3.1 PCR Methods 66
2.3.2 Gel Electrophoresis 67

2.3.3 Gel Extraction and Purification 68
2.3.4 DNA Sequencing 68
2.3.5 DNA Extraction from Cells 68
2.3.6 RNA Extraction from Cells 68
2.3.7 Quantification of DNA and RNA 69
2.3.8 Reverse Transcriptase PCR 69
2.3.9 Alternative Splicing Assays 70
2.4 Genome Engineering by CRISPR/Cas9 System 71
2.4.1 Target Genes 71
2.4.2 gRNAs Design 71
2.4.3 *In vitro* Assessment of the gRNA and Cas9 Activities . . 72
2.4.4 Construction of the Cas9 Plasmid 74
2.4.5 pScramble_Cas9 Plasmid 75
2.4.6 Construction of the Repair Template for Homologous Recombination 76
2.5 Transgene Generation 79
2.5.1 Solutions, Media and Antibiotics 79
2.5.2 Titration of a Kill Curve for Hygromycin 79
2.5.3 Transfection of the Cells 80
2.5.4 Antibiotic Selection of Positive Clones 80
2.5.5 Colony Pick Up 80
2.5.6 Colony PCR of Human Fibroblasts 81
2.5.7 Identification of Mutations in the Target Site 81
2.6 Protein Analysis 82
2.6.1 Solutions, Buffers and Antibodies 82
2.6.2 Protein Extraction from Cells 83

2.6.3 Protein Quantification 83
2.6.4 Detection of *CDK12* on Western Blot 84
2.7 Microscopy 85
2.7.1 Brightfield Microscopy 85
2.7.2 Micro Widefield Plate Reader 85
2.7.3 Confocal Microscopy 86
2.8 Electronic Databases 86

3 Characterization of the effect of kinase inhibitors on DM cells 87
3.1 Aim 87
3.2 Introduction 87
3.3 Investigation of the localization of foci in the nucleus of DM cells 89
3.3.1 Co-localization of foci with SC35 speckles 89
3.3.2 Co-localization of foci with CDK12 91
3.3.3 Co-localization of CDK12 with SC35 speckles 96
3.4 Effect of the kinase inhibitors on repeat expansion foci 100
3.5 Assessment of the toxicity of the compounds 103
3.6 Effect of the kinase inhibitors on SC35 speckles 104
3.7 Effect of the kinase inhibitors on MBNL1 107
3.8 Effect of the kinase inhibitors on Alternative Splicing 113
3.9 Effect of newly synthesized kinase inhibitors in DM2 cells . . . 118
3.10 Summary 120

4 *CDK12* CRISPR/Cas9-mediated gene knock out in DM1 cells 123
4.1 Aim 123
4.2 Introduction 123
4.3 Analysis of *in vitro* cleavage efficiency of the gRNAs 125

4.4 Transfection of cells and expansion of the clones 128
4.5 Determining of the efficiency of genome editing 133
4.6 Assessment of *CDK12* knock out by RT-PCR 138
4.7 *CDK12* knock out quantification by Western Blot 141
4.8 Effect of *CDK12* genome editing on the mutant transcript 145
4.9 Effect of *CDK12* overexpression on the mutant transcript 150
4.10 Summary 152

5 Experiments to control foci expression in DM cells by CRISPR/Cas9-mediated knock in of an inducible promoter 156

5.1 Aim 156
5.2 Introduction 156
5.3 Analysis of *in vitro* cleavage efficiency of the gRNAs 160
5.4 Transfection Experiment 1 161
5.4.1 Transfection of the cells 161
5.4.2 Analysis of transfection efficiency by GFP detection . . 163
5.4.3 Expansion of the clones 164
5.4.4 Determining of the efficiency of genome editing in Transfection experiment 1 165
5.4.5 Detection of mutant transcripts by FISH in the CRISPR-generated clones 172
5.5 Transfection Experiment 2 176
5.5.1 Transfection of the cells 176
5.5.2 Analysis of transfection efficiency by GFP detection . . 177
5.5.3 Expansion of the clones 178
5.5.4 Determining of the efficiency of genome editing in Transfection experiment 2 179

5.6 Transfection Experiment 3 183
5.6.1 Transfection of the cells 183
5.7 Summary 184

6 Final Discussion and Conclusion **186**

6.1 The localization of foci in DM1 and DM2 cells 189
6.2 Effect of kinase inhibitors on DM cells 192
6.3 CRISPR-genome engineering of *CDK12* in DM1 cells 194
6.3.1 Role of CDK12 in myotonic dystrophy 198
6.4 Use of CRISPR/Cas9 system to engineer a disease model for myotonic dystrophy 199
6.5 Final Conclusion 202

References **204**

Appendices **237**

List of Figures

1.1 Individuals with congenital myotonic dystrophy. 5

1.2 Members of a family affected with adult onset DM. 6

1.3 Representation of the DM-causing genes. 15

1.4 DM1-proposed disease mechanisms of CUG repeat expansions. 20

1.5 The balance of MBNL1 and CUGBP1 regulate a subset of splicing events that are reversed in DM1. 26

1.6 Schematic of the structure of the human CDK12 protein. 42

1.7 Overview of the CRISPR adaptive immunity. 45

1.8 Types of CRISPR systems. 47

1.9 DNA repair pathways after the introduction of DSBs. 49

1.10 Outcomes of HDR-, NHEJ- and MMEJ-mediated knock-in. . . . 50

1.11 Examples of cell types and organisms that have been successfully engineered using the CRISPR system. 51

1.12 Schematic representation of inducible gene expression systems. 53

1.13 Tet-One design 54

2.1 Details of the oligonucleotides used for the *in vitro* assay and the steps followed in order to transcribe the gRNAs *in vitro*. 73

2.2 Map of the pCas9_gRNA_CDK12 plasmid. 75

2.3 Map of the pCas9_gRNA_ZNF9 plasmid. 75

2.4 Map of the pScramble_Cas9 plasmid. 76

2.5 Map of the pCDK12_HDR plasmid. 77

2.6 Map of the pZNF9_HDR plasmid. 78

2.7 Schematic of MMEJ-mediated gene targeting strategy to knock in the inducible promoter in DM2 cells. 79

3.1 Co-localization or association of nuclear foci with SC35 speckles in DM1 and DM2 cells. 90

3.2 Quantification of the number of foci independent or co-localizing with SC35 speckles in DM1 and DM2 cells. 91

3.3 Optimization of different fixative methods for the combination of FISH with ICC in order to study the co-localization of foci with CDK12 protein. 93

3.4 Co-localization of nuclear foci with CDK12 in DM1 but not in DM2 cells. 96

3.5 Co-localization of SC35 speckles with CDK12 spots in non-DM, DM1 and DM2 cells. 97

3.6 Pearson's Correlation Coefficient formula. 98

3.7 Pearson's Correlation Coefficient for SC35 and CDK12 co-localization. 99

3.8 Comparison of the number and area of SC35 speckles and CDK12 spots between non-DM, DM1 and DM2 cells. 100

3.9 Foci reduction in the DM2 cell line (KAGO Telo) following treatment with 6 kinase inhibitors. 102

3.10 *In sit*u hybridization images of DM2 cells (KAGO Telo) treated with Dinaciclib and SNS-032. 103

3.11 MTT assay for Dinaciclib and SNS-032. 104

3.12 Quantification of SC35 speckles in DM1, DM2 and non-DM cells after treatment with Dinaciclib and SNS-032 for 24 hours. . 106

3.13 Quantification of SC35 speckles in DM1 cells after treatment with Dinaciclib and SNS-032 for 48 hours. 107

3.14 Fluorescence images of DM1 and DM2 cells treated with DMSO for 24 hours. 108

3.15 FISH images and MBNL1 ICC after treatment of DM1 cells with Dinaciclib and SNS-032 for 24 hours. 109

3.16 FISH images and MBNL1 ICC after treatment of DM2 cells with Dinaciclib and SNS-032 for 24 hours. 110

3.17 Effect of Dinaciclib on MBNL1 protein expression in DM1 (KB) and DM2 (KAGO) cells. 111

3.18 Effect of SNS-032 on MBNL1 protein expression in DM1 (KB) and DM2 (KAGO) cells. 112

3.19 Splicing analysis of *INSR* and *PHKA1* genes after treatment of DM2 cells with Dinaciclib. 113

3.20 Splicing analysis of *SERCA1* and *NCOR* genes after treatment of DM2 cells with Dinaciclib. 114

3.21 Splicing analysis of the *MBNL1* gene after treatment of DM2 cells with Dinaciclib. 114

3.22 Splicing analysis of *NFIX* and *ITGA6* genes after treatment of DM2 cells with Dinaciclib. 115

3.23 Analysis of the ratio of the isoforms of *INSR*, *PHKA1*, *SERCA1*, *NCOR*, *NFIX* and *ITGA6* in DM2 cells treated with DMSO. . . 116

3.24 Analysis of the alternative splicing of several genes in DM2 cells treated with 5 μM Dinaciclib and DMSO. 117

3.25 Foci reduction in KAGO Telo (DM2) cells following treatment with newly synthesized kinase inhibitors. 119

4.1 Overview of the *in vitro* assay protocol. 126

4.2 *In vitro* assay of the predicted gRNAs targeting *CDK12*. 127

4.3 Analysis of GFP expression after electroporation of KB Telo MyoD cells with pmaxGFP. 132

4.4 Diagram of *CDK12* genomic DNA and cDNA showing the location of the set of primers used in this chapter. 134

4.5 Analysis of colonies arising from the pCas9_gRNA_CDK12 + hygromycin linear marker transfections in KB Telo MyoD cells. 135

4.6 RT-PCR amplification of *CDK12* and *GAPDH* RNA in various cell clones. 138

4.7 RT-PCR analysis of *CDK12* expression in CRISPR-engineered DM1 cells. 139

4.8 Exon skipping analysis by RT-PCR of CRISPR-mutated DM1 cells. 140

4.9 Alignment of the genomic DNA and the cDNA sequences of the clones 2-G3 and 2-H4. 141

4.10 CDK12 western blot images and protein levels in the CRISPR-modified clones. 142

4.11 CDK12 immunostaining in control cells (KB) and 4 CRISPR-modified cell clones. 143

4.12 Quantification of the number of CDK12 spots in control cells (KB) and 4 CRISPR-modified cell clones. 144

4.13 Statistical analysis of the number, area and intensity of foci per cell in the CRISPR-generated clones. 147

4.14 Relationship between the number of foci per cell and the area of foci in the CRISPR-generated clones. 148

4.15 Structure of the plasmid used to transiently overexpress CDK12 in DM1 and DM2 cells. 150

4.16 Representative images of foci in DM1 and DM2 cells transfected with a plasmid to overexpress CDK12. 151

4.17 Transient overexpression of CDK12 increases the number of foci per cell in DM cells. 152

5.1 Schematic of the integration event and control of the expression of the mutant transcript in DM2 cells. 158

5.2 Map of the pTet-One plasmid (Clontech). 159

5.3 *In vitro* assay of the predicted gRNAs targeting *CNBP*. 160

5.4 Expression of GFP in DM2 cells transfected with pmaxGFP in Transfection Experiment 1. 164

5.5 Genotyping analysis of CRISPR/Cas9-mediated knock-in at intron 1 of *CNBP* in DM2 cells in Transfection Experiment 1. . . . 167

5.6 Genotyping analysis of CRISPR/Cas9-mediated knock-in at intron 1 of *CNBP* in DM2 cells transfected without the repair template in Transfection Experiment 1. 168

5.7 Characterisation of the mutations created after the Cas9 cleavage in intron 1 of the *CNBP* gene in Transfection Experiment 1. . . . 169

5.8 Number of foci per cell in colonies from replicate 1, engineered with CRISPR/Cas9. 173

5.9 Number of foci per cell in colonies from replicates 2 and 3, engineered with CRISPR/Cas9. 174

5.10 Microscopy images of colonies 1-A1, 1-A4 and 2-D1 after performing *in situ* hybridization. 175

5.11 Expression of GFP in DM2 cells transfected with pmaxGFP in Transfection Experiment 2. 178

5.12 Genotyping analysis of CRISPR/Cas9-mediated knock-in at intron 1 of *CNBP* in DM2 cells in Transfection Experiment 2. . . . 180

5.13 Characterisation of the mutations created after the Cas9 cleavage in intron 1 of the *CNBP* gene in Transfection Experiment 2. . . . 181

List of Tables

1.1 Summary of DM1 phenotypes, clinical manifestations, CTG length and age of onset. 7

1.2 Comparison of the clinical manifestations between DM1 and DM2. 8

1.3 Summary table of the six kinase inhibitors used in this study . . 41

2.1 Sequences of the primers used to analyse the alternative splicing after treatment of DM2 cells with 5 μM of Dinaciclib. 70

2.2 gRNA sequences targeting *CDK12* 71

2.3 gRNA sequences targeting *CNBP* 72

4.1 Off-target effects predicted for gRNA 2 targeting *CDK12*. 128

4.2 Concentrations of the reagents transfected into KB Telo MyoD cells in order to mutate *CDK12*. 131

4.3 Number of colonies picked from each transfection event. 132

4.4 Mutations identified in DM1 cells transfected with pCas9_gRNA_CDK12 and the hygromycin linear marker. . . . 136

4.5 Summary of the number of hygromycin-resistant colonies and alleles mutated after CRISPR/Cas9-genome editing of the *CDK12* gene in DM1 cells. 137

4.6 Efficiency of Cas9 cleavage and hygromycin integration into the target site of *CDK12* in DM1 cells. 137

4.7 Foci number, area and intensity values after performing *in situ* hybridization in the CRISPR-generated clones. 146

4.8 Data from the type of mutations found in the target site, CDK12 mRNA and protein levels and number of foci per cell for each of the colonies screened in this chapter. 149

5.1 Off-target effects predicted for the gRNA 2 targeting *CNBP*. . . 161

5.2 Concentrations of the reagents transfected into DM2 cells in Transfection Experiment 1. 163

5.3 Number of colonies picked up from each transfection event in the Transfection Experiment 1. 165

5.4 List of the colonies screened and the mutations present in each of them in the Transfection Experiment 1. 171

5.5 Summary of the number of Hygromycin-resistant colonies and alleles mutated after CRISPR/Cas9-genome editing of the *CNBP* gene in Transfection Experiment 1. 172

5.6 Efficiency of the Cas9 cleavage and recombination event in Transfection Experiment 1. 172

5.7 Concentrations of the reagents transfected into the cells in Transfection Experiment 2. 177

5.8 Number of colonies picked up from each transfection event in Transfection Experiment 2. 178

5.9 List of the colonies screened and the mutations present in each of them in Transfection Experiment 2. 181

5.10 Summary of the number of Hygromycin-resistant colonies and alleles mutated after CRISPR/Cas9-genome editing of the *CNBP* gene in Transfection Experiment 2. 182

5.11 Efficiency of the Cas9 cleavage and pTetOne integration in Transfection Experiment 2. 182

5.12 Concentrations of the reagents transfected into the cells in Transfection Experiment 3. 183

List of Abbreviations

ALS	Amyotrophic Lateral Sclerosis
AONs	Antisense Oligonucleotides
BSA	Bovine Serum Albumin
CDK	Cyclin-Dependent Kinase
CDM	Congenital Myotonic Dystrophy
CELF1	CUGBP Elav-Like Family Member 1
CLCN1	Chloride Channel 1
CNBP	Cellular Nucleic acid-Binding Protein
CNS	Central Nervous System
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTD	C-Terminal Domain
cTNT	Cardiac Troponin T
CUGBP1	CUG-Binding Protein
DAPI	Dihydrochloride
DEPC	Diethylpyrocarbonate
dNTP	Deoxynucleotide Triphosphate
DM	Dystrophia Myotonica
DMEM	Dulbecco's Modified Eagle's Medium
DMPK	Dystrophia Myotonica Protein Kinase
DMSO	Dimethyl Sulfoxide
EDTA	Ethylenediaminetetraacetic acid
FBS	Fetal Bovine Serum
FISH	Fluorescence *In Situ* Hybridization
FTD	Frontotemporal Dementia
FXTAS	Fragile X-associated Tremor Ataxia Syndrome
GFP	Green Fluorescent Protein
gRNA	Guide RNA

HDAC	Histone Deacetylase
HDR	Homology-Directed Repair
HSA	Human Skeletal Actin
ICC	Immunocytochemistry
INSR	Insulin Receptor
IRES	Internal Ribosome Entry Site
ITGA6	Integrin Subunit Alpha 6
LB	Luria Broth
LHA	Left Homology Arm
MBNL	Muscleblind-like
MCS	Multiple Cloning Site
miRNA	microRNA
MMEJ	Microhomology-Mediated End Joining
MyoD	Myogenic Differentiation
NCBI	National Centre for Biotechnology Information
NCOR	Nuclear receptor Co-Repressor 1
NHEJ	Non-homologous End Joining
NFIX	Nuclear Factor I/X
NLS	Nuclear Localization Signal
ORF	Open Reading Frame
PAM	Protospacer Adjacent Motif
PBS	Phosphate Buffered Saline
PCC	Pearson's Correlation Coefficient
PFA	Paraformaldehyde
PHKA1	Phosphorylase Kinase Regulatory Subunit Alpha 1
PITCH	Precise Integration into Target Chromosome
PKC	Protein Kinase C
PROMM	Proximal Myotonic Myopathy
PVDF	Polyvinylidene Fluoride

RAN	Repeat-Associated Non-ATG
RHA	Right Homology Arm
RIPA buffer	Radioimmunoprecipitation assay buffer
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SC35	Serine and Arginine Rich Splicing Factor 2
SCA8	Spinocerebellar Ataxia Type 8
SERCA	Sarcoplasmic/endoplasmic reticulum Ca2ATPase
siRNA	small interfering RNA
TAE	Tris-Acetate-EDTA
TALEN	Transcription Activator-Like Nucleases
TBS-T	Tris Buffered Saline and Tween 20
TNNT2	Cardiac Troponin T
UTR	Untranslated Region
ZNF9	Zinc Finger 9
ZFN	Zinc-Finger Nuclease

Acknowledgements

First of all I would like to thank my supervisor, David Brook, for the great opportunity to work in his lab and grow as an independent scientist and for his guidance and advice throughout my entire PhD. Also, thanks to my second supervisor, Chris Hayes, for his comments and suggestions, especially regarding the chemistry aspects. In addition, I would like to thank all members of David's lab for their help and the good moments lived together, but especially to Marzena Wojciechowska, for her experienced advice and encouragement, to Ami Ketley, for her help and guidance, and to Naveed Malik, for his support and help and for the lovely company, especially in the coffee breaks. And I would like to thank the University of Nottingham for the funding provided.

Second, I am very grateful for all the friends I've made in Nottingham. Thanks to Elena, who besides being a friend since the beginning, has been an excellent flatmate all these years. Also, thanks to Judith, Mamen, Carol, Pablo, Nuria, Teresa, Angela and many others. I've met great people who have become very good friends, and I will always remember all the good experiences lived here. And of course, I am grateful to my friends from Spain, especially to my close friends, Alba, Maria, Blanca, Tamara, and many others, who have been supporting me all this time.

And last but not least, I am very thankful to my family for all the support and encouragement, and for believing in me all these years, but especially to my mum, Neus, who regardless of the distance has been taking care of me every time I've needed her.

Thank you all, I couldn't have done this without you!

Abstract

Myotonic dystrophy (DM) is a dominantly inherited multisystem disease. There are 2 types of DM: DM1, caused by a CTG repeat expansion in the 3'-UTR of the *DMPK* gene, and DM2, caused by a CCTG repeat expansion in intron 1 of the *CNBP* gene. In both types of DM the transcribed RNA, containing the repeat expansions, accumulate in the nucleus and sequester alternative splicing factors, such as MBNL1 proteins, causing the missplicing of many genes and contributing to the disease pathogenesis. These aggregations accumulate in the nucleus in the form of foci and represent a key target for therapeutic development.

In the first part of this thesis I studied the relationship of foci with SC35 speckles and CDK12 protein in the nuclei of DM1 and DM2 cell lines. I also examined a subset of kinase inhibitors, which target mainly CDK12, for their ability to eliminate foci in DM2 cells. Additional effects of the two most potent inhibitors, Dinaciclib and SNS-032, were studied in both DM1 and DM2 cells. Both inhibitors reduce the number of SC35 speckles in the nucleus, cause a redistribution of MBNL1 and reduce the levels of MBNL1 protein, but do not restore the alternative splicing of the genes analysed. Additionally, attempts were made to knockdown *CDK12* in DM1 cells using the CRISPR/Cas9 system. The system achieved a 45% efficiency of cleavage of the target site in hygromycin-resistant clones. Nevertheless, the mutations introduced in the gene did not eliminate the CDK12 protein and they had an unexpected effect on foci production. Transient overexpression of CDK12 increased the number of foci in both DM1 and DM2 cells. Thus far, results suggest that CDK12 is involved in the transcription of the mutant transcripts.

The second part of this thesis focuses on the development of an inducible cell model for DM2 using the CRISPR/Cas9 system. The approach used here aims to overcome the difficulties associated with cloning of very unstable long repeat expansions. The exogenous inducible promoter was integrated into two clones, although in one clone it was not in the right orientation and in the other, the end of the fragment was missing. Efficiency of the CRISPR system was around 57 to 72%, and many hygromycin-resistant clones showed reduced numbers of foci per cell, suggesting that the mutations in intron 1 of *CNBP* compromise the transcription of the gene and the production of the mutant transcript.

Thus, the experiments described in this thesis have made a contribution to improve understanding of DM disease pathogenesis and provide platform for further studies.

Chapter 1

Introduction

1.1 History of Myotonic Dystrophy

The earliest reports of myotonic dystrophy (DM) come from the late 19th century. In 1876, Dr. Julius Thomsen described the main symptoms of the disease he suffered: the myotonia, its hereditary nature and the benign nature of the disease, so soon this disease became known as myotonia congenita or 'Thomsen's disease'. Myotonia is an involuntary contraction of the skeletal muscle followed by a delayed relaxation, due to the repetitive firing of action potentials. Due to the genetic heterogeneity of the disease, during the next years a series of atypical Thomsen's disease were reported, and the last quarter of the 19th century and the first quarter of the 20th century were marked for the discovery, delineation and establishment of the main features of the disorder (Harper, 2001).

Myotonic dystrophy is also known as 'Steinert's disease' due to the detailed study doctor Steinert made of the disease back in 1909, in which he described nine cases of affected people showing the main symptoms of the disease (reviewed in (Steinberg and Wagner, 2008)). Following the first descriptions of myotonic dystrophy, a few years later the disease was recognised as a generalized systemic disorder and additional symptoms such as cataracts, gonadal atrophy, mental retardation, cardiorespiratory problems, changes in the muscle and anticipation were associated with the disorder (Harper, 2001).

The descriptive studies of patients affected with myotonic dystrophy provided knowledge of its clinical features, pathology and inheritance, but still the factors involved in its pathogenesis were unknown. In the second half of the 20th century the experimental studies provided more knowledge about the pathogenesis of the disorder, and in the 1980s, thanks to the development of molecular genetics and gene mapping, it became possible to isolate the specific gene and mutation responsible for the disease (Harper, 2001). It was in 1992 when an expanded trinucleotide (CTG) repeat sequence in the chromosome 19, which was unstable and prone to expand in successive generations, was described as the cause of myotonic dystrophy (Brook et al., 1992; Fu et al., 1992; Mahadevan et al., 1992). The next few years following the discovery of the mutation were important to improve the understanding of the genetic and mutational aspects of the disease, however since then the progress to understand the disease pathogenesis has been very slow. The reason is that myotonic dystrophy is a multisystem disease in which the gene function can be altered at different levels such as DNA, RNA and protein, and also other genes may be involved in the disease pathogenesis (Harper, 2001).

For many years myotonic dystrophy was considered a unique progressive muscular dystrophy. However, the discovery of patients with a clinically atypical myotonic dystrophy and the lack of the mutation on chromosome 19 responsible for Steinert's disease were the first steps to the discovery and establishment of a new type of the disease, originally called PROMM in Europe (Proximal Myotonic Myopathy) (Ricker et al., 1994), or myotonic dystrophy with no CTG repeat expansion or myotonic dystrophy type 2 (DM2) in the United States (Thornton et al., 1994). Ricker *et al.* were the first ones in describing this new type of myotonic dystrophy in 1994, and in the next years many patients and families with this type of the disease were also reported and the symptoms described (Ricker et al., 1994). Some of the differences between PROMM and myotonic dystrophy that led to consider them as two different disorders include a more proximal weakness in PROMM compared to a distal weakness in myotonic dystrophy. Myotonia is milder in PROMM, but patients show muscle pain, whereas myotonic dystrophy patients usually don't. The involvement of

the CNS is less marked in patients with PROMM (Harper, 2001). In 2000 the International Myotonic Dystrophy Consortium established a new nomenclature for myotonic dystrophy, so that Steinert's disease is now termed as myotonic dystrophy type 1 (DM1) and PROMM is termed as myotonic dystrophy type 2 (DM2) (IDMC, 2000).

In 2001 Liquori *et al.* discovered and mapped the mutation responsible for DM2, which is caused by a CCTG repeat expansion in intron 1 of the *CNBP* gene (formerly known as *ZNF9* gene) on chromosome 3 (Liquori et al., 2001).

Through the years patients with different symptoms have been reported, and some of them can not be classified neither as DM1 or DM2 patients (Harper, 2001). Myotonic dystrophy is not the only disease characterized by myotonia, and different disorders have been identified, such as myotonia congenita, caused by mutations in muscle chloride ion channel; paramyotonia congenita, caused by mutations in muscle sodium ion channel; or chondrodystrophic myotonia, caused by a mutation in the chromosome 1 (Harper, 2001). This work will focus on DM1 and DM2, which will be described next, whereas characteristics of the other types of myotonia won't be considered further in this thesis.

1.2 The Phenotype of Myotonic Dystrophy

Myotonic dystrophy is a dominantly inherited multisystem disease and the most common adult muscular dystrophy. The prevalence of DM1 is estimated to be 1/8000, and is more common in populations of Europe, quite frequent in Japan and rare in India, Taiwan and Sub-Sahara Africa (Ashizawa and Epstein, 1991; Suominen et al., 2011). On the other hand, the prevalence of DM2 is not well established due to the large phenotypic variability and a large number of patients with mild symptoms that remain undiagnosed. However, DM2 prevalence is estimated to be lower than DM1 in European populations, except in Germany, Poland and Finland (Bachinski et al., 2003; Liquori et al., 2003; Suominen et al., 2008, 2011).

Common symptoms to both types of the disease include myotonia, mus-

cular dystrophy and muscle weakness, cataracts, endocrine disorders, cardiac conduction defects and male infertility caused by testicular atrophy (Harper, 2001). The severity and characteristics of these symptoms vary between DM1 and DM2. Distal muscle weakness is predominant in DM1 whereas in DM2 is proximal. Myotonia is more severe in DM1, with DM2 patients showing a mild myotonia. Cardiac involvement is also milder in DM2 patients. Central nervous system is more involved in DM1 compared to DM2, with DM1 patients suffering from intellectual disability, psychological dysfunction, personality traits or excessive sleepiness (Harper, 2001). A summary of the clinical differences between DM1 and DM2 can be found in Table 1.2

A significant difference between DM1 and DM2 is the absence of a congenital form in DM2 and the fact that the repeat expansions may contract from one generation to the next one, unlike in DM1 where the offspring shows larger repeat expansions due to the instability of the repeats. These larger expansions in the offspring correlate with a more severe DM phenotype and the symptoms tend to appear at an earlier age, a phenomenon called anticipation (Day et al., 2003). DM2 is an adult-onset disease, nonetheless DM1 can be classified in four categories based on the age of onset, the severity of the symptoms and the clinical features: congenital, childhood, classical or adult onset and minimal or late onset. These four types are summarized in Table 1.1.

1.2.1 Congenital DM1

The congenital form of DM1 (CDM) is a fetal-onset disorder with more severe symptoms and involvement of the CNS, associated with expansions of more than 1000 CTG repeats (Tsilfidis et al., 1992). Such large repeat expansions are frequently generated during oogenesis, thus CDM is mainly maternal inherited, although transmission of CDM by the father has been reported (Zeesman et al., 2002). It presents before birth as polyhydramnios and reduced fetal movements. At the time of birth clinical features include generalized muscle weakness, neonatal hypotonia, facial diplegia, cerebral atrophy, ventricular enlargement and feeding or respiratory difficulties. Later on in life CDM patients exhibit

learning disabilities, oropharyngeal weakness, producing a characteristic tented appearance of the upper lip (termed tented or fish-shaped), marked dysarthria, impairment of communication skills, behavioural problems, autism and develop the degenerative features of the disease. Infant mortality due to respiratory problems is high (Turner and Hilton-Jones, 2010).



Figure 1.1: **Individuals with congenital myotonic dystrophy. Left.** Child with the typical V-shaped upper lip. **Right.** Adult with congenital DM. Taken from (Turner and Hilton-Jones, 2010).

1.2.2 Childhood DM1

Childhood DM is developed in children after the first year but before age 10. In this case the sex of the parent does not influence the development of the disease. Affected children present intelligence impairment, with IQ in the range of 50 to 70 (Steyaert et al., 1997) and learning abnormalities (Steyaert et al., 2000), a range of psychiatric problems, including anxiety and mood disorder, facial weakness and retarded motor development (Douniol et al., 2012).

1.2.3 Classical/Adult onset DM1

The classical form of the disease develops during the second, third or fourth decade. The most common symptoms are myotonia and distal muscle weakness. Muscle weakness progresses slowly, particularly affecting long finger flexors and ankle dorsiflexors, leading to difficulties when performing tasks requiring fine dexterity (Krishnan and Kiernan, 2006). Myotonia can also difficult daily

activities such as using tools, household equipment or doorknobs. It causes problems with talking, chewing and swallowing due to tongue or facial muscle weakness. Cardiac abnormalities are common, including arrhythmias and conduction blocks that contribute to mortality of the disease (Chebel et al., 2005). Patients develop cataracts and daytime sleepiness is common. Type 2 diabetes is also common in DM1 patients, caused by insulin resistance. Other endocrine abnormalities include testicular atrophy and hypotestosteronism (Meola and Cardani, 2015).



Figure 1.2: **Members of a family affected with adult onset DM.** A mother affected with DM1 in the centre, with her daughter on the left and her son on the right. Both daughter and son have more CTG repeats than the mother, and therefore present more severe symptoms appeared at an earlier age, due to the anticipation event. Taken from (Turner and Hilton-Jones, 2010).

1.2.4 Minimal/Late onset DM1

Small CTG expansions in the range of 50 to 100 are associated with mild myotonia, weakness and cataracts, usually developed after age 40. Many individuals with this type of mutation are undiagnosed, as some can be asymptomatic (Arsenault et al., 2006).

Table 1.1: **Summary of DM1 phenotypes, clinical manifestations, CTG length and age of onset.** Modified from (Meola and Cardani, 2015).

Phenotypes	Clinical findings	CTG length	Age of onset
Congenital	Infantile hypotonia Respiratory failure Learning disability Cardiorespiratory complications	>1000	Birth
Childhood onset	Facial weakness Myotonia Low IQ Conduction defects	50 to 1000	1 to 10 years
Classical/Adult onset	Weakness Myotonia Cataracts Conduction defects Insulin resistance Respiratory failure	50 to 1000	10 to 30 years
Minimal/Late onset	Mild myotonia Cataracts	50 to 100	20 to 70 years
Pre-mutation	None	38 to 49	N/A

1.2.5 DM2

Myotonic dystrophy type 2 usually appears in adult life with a large phenotypic variability, which includes early-onset cataracts, muscle pain and weakness, varying grip myotonia and thigh muscle stiffness, appearing between ages 20 to 70 (Thornton et al., 1994; Day et al., 2003). Myotonia is often less apparent than in DM1 and it is more difficult to identify (Meola, 2000). Cardiac abnormalities are less frequent than in DM1, and they include first-degree atrio-ventricular and bundle branch block. Nonetheless, severe arrhythmias, pacemaker implantation and sudden death have been described in a small subset of patients (Sansone et al., 2013). The involvement of the central nervous system is controversial, as a few DM2 individuals with mental retardation have been reported, but it may be accidental or an infrequent disease consequence (Ricker et al., 1994; Day et al., 2003). Nonetheless, mild cognitive dysfunctions have been reported, such as daytime sleepiness, psychological dysfunction, neuropsychological changes and abnormal personality traits (Maurage et al., 2005). Other clinical manifestation typical of DM2 patients include glucose intolerance, excessive sweating, hypogonadism and dysphagia, and they worsen over time (Thornton et al., 1994; Ricker et al., 1994).

Table 1.2: **Comparison of the clinical manifestations between DM1 and DM2.** Modified from (Meola and Cardani, 2015).

Clinical Features	**DM1**	**DM2**
Epidemiology	Widespread	European
Age of onset	0 to adult	8 to 60
Anticipation	Frequent	Exceptional
Congenital form	Present	Absent
Life expectancy	Reduced	Normal range
Cataracts	Frequent	Present in minority
Myotonia	Evident in adult-onset	Present in <50%
Muscle weakness	Disabling at age 50	Onset after age 50 to 70
Facial and jaw weakness	Always present	Usually absent
Respiratory muscle weakness	Always later	Exceptional
Distal limb muscle weakness	Maybe present	Late
Proximal limb muscle weakness	Absent or mild	Most disabling symptom
Sternocleidomastoid weakness	Face, temporal, distal hands and legs	Usually absent
Myalgic pain	Absent	Present in >50%
Tremors	Absent	Prominent in many
Behavioural change	Early in most	Not apparent
Cognitive disorders	Prominent	Not apparent
Hypersomnia	Prominent	Infrequent
Cardiac arrhythmias	Frequent	From absent to severe
Male hypogonadism	Manifest	Subclinical in most
Diabetes	Frequent	Infrequent

1.3 The Genetic Basis of Myotonic Dystrophy

Myotonic dystrophy is a microsatellite expansion disorder, in which a short DNA sequence expands to a pathogenic range and causes the disease. Both types of myotonic dystrophy are caused by a similar mutation but in different genes and chromosomes, and this influences the similarities and differences between them.

1.3.1 Genetic basis of DM1

The *DMPK* gene encodes the protein DMPK (DM1 protein kinase). Alternative splicing produces six different DMPK isoforms, with all isoforms having an N-terminal domain, a kinase domain and a coiled region (Groenen et al., 2000). The most abundant isoform is an 80 KDa protein expressed in skeletal muscle, cardiac muscle and smooth muscle (Lam et al., 2000). DMPK is important for

skeletal muscle proper structure and function, and it plays a role in myogenesis (Carrasco et al., 2002). Downregulation of *DMPK* can cause muscle atrophy, and this is one of the main symptoms of myotonic dystrophy (Reddy et al., 1996). Cardiac expression of DMPK also suggests a role of the protein in cardiac muscle function. Indeed, mice lacking *Dmpk* developed cardiac conduction defects, similar to those observed in DM1 patients (Berul et al., 1999). The metabolic alteration observed in DM1 patients can also be related to DMPK. The protein is highly expressed in muscle, which is a key target for insulin-dependent regulation of glucose metabolism (Zisman et al., 2000). Mice lacking *Dmpk* show impaired insulin-induced glucose uptake and decreased insulin sensitivity in cardiac and skeletal muscles (Llagostera et al., 2007).

DM1 is caused by the aberrant expansion of the CTG triplet repeat in the 3'-UTR (untranslated region) of the *DMPK* gene on chromosome 19q13.3. Unaffected individuals have between 5 and 37 CTG repeats. Repeat lengths of 38 to 50 are considered a premutation and affected individuals have from 50 up to several thousand CTG repeats (Brook et al., 1992; Mahadevan et al., 1992). Patients with premutations are asymptomatic or present very mild symptoms. Repeat expansions larger than 1000 CTGs are associated with congenital myotonic dystrophy (Schoser and Timchenko, 2010). The CTG repeats are highly somatically unstable, leading to high variability of repeat length in different tissues (Ashizawa et al., 1993). They also tend to expand from one generation to another, and the symptoms tend to become more severe and appear at an earlier age (Redman et al., 1993), a phenomenon known as genetic anticipation. However, sometimes the expanded repeat contracts during offspring, but this represents less than 5% of transmissions (Ashizawa et al., 1994; Puymirat et al., 2009).

The CTG repeat expansion length shows a general but not absolute correlation with severity of the disease and age of onset. Based on Hamshere's finding, the correlation is observable in blood, in patients with repeat lengths shorter than 400 CTGs. This does not apply in other organs such as muscle, where repeat length does not correlate to the degree of weakness (Hamshere et al., 1999). However, Morales *et al.* showed that the progenitor allele length is a modifier of

the age of onset, with no need of a threshold above which the repeat length contributes towards age at onset. Furthermore, the level of somatic instability also contributes to the age of onset (Morales et al., 2012). Recently, a small number of DM1 unrelated families with interrupted CTG repeats have been reported. The most common interruptions are CCG or CGG triplets and the individuals affected tend to have milder symptoms and variable phenotypes. Interruptions within the CTG repeats appear to increase stability of the repeat tract and reduce anticipation. However, foci are still present in muscle tissue and splicing defects are observable (Musova et al., 2009; Braida et al., 2010; Santoro et al., 2013; Botta et al., 2017).

1.3.2 Genetic basis of DM2

In 1989 Rajavashisth and collaborators, looking for a *trans*-acting factor that regulates cholesterol homeostasis, identified a cDNA that encoded a 19-kDa protein with 7 conserved repeats of 14 amino acids containing the CCHC zing finger sequence. These repeats were very similar to the finger domains of the family of retroviral nucleic acid-binding proteins (NBPs), thus the protein was designated as cellular nucleic acid-binding protein (CNBP). Because of the presence of the zinc finger domains, this gene has also been known for many years as zinc finger 9 (ZNF9) (Rajavashisth et al., 1989). Currently, the Human Genome Organization Gene Name Committee has adopted CNBP as the official name of the gene (Calcaterra et al., 2010). A few years later after the discovery of *CNBP*, the human gene was mapped to the long arm of chromosome 3 (3q13.3-q24). The gene is structured in 5 exons with introns of variable size, with intron 1 the longest (Lusis et al., 1990).

CNBP is known to bind to RNA and single-stranded DNA, but not double-stranded DNA, and it prefers nucleic acids with G-rich unpaired stretches. It promotes rearrangements of nucleic acid secondary structure, acting as a chaperone and thus could play a role in controlling transcriptional and translational processes through chromatin or RNA structural remodelling (Lombardo et al., 2007; Armas et al., 2008). Some of the elements to which CNBP binds

are regulatory elements of several genes (Rajavashisth et al., 1989), and it is also involved in the regulation of cap-dependent and cap-independent translation by activation of IRES-mediated translation (Sammons et al., 2010). Furthermore, CNBP binds to the 5' UTR of TOP-containing mRNAs, which are involved in the translation of proteins associated with translational apparatus, such as poly(A)-binding proteins, ribosomal proteins and elongation factors eEF1A and eEF2 (Pellizzoni et al., 1997).

In 2001 the *CNBP* gene was linked to myotonic dystrophy type 2. Liquori *et al.* demonstrated that DM2 is caused by a CCTG tetranucleotide repeat expansion in intron 1 of the *CNBP* gene on chromosome 3q21. Expanded alleles range from 75 up to 11000 CCTG repeats, with a mean repeat length of around 5000 (Liquori et al., 2001). Usually the wild type allele is interrupted by one or more ACTG, TCTG or GCTG motifs, whereas the expanded alleles are typically uninterrupted (Liquori et al., 2001; Bachinski et al., 2003). The DM2 repeat expansion shows somatic instability, increasing in length over time. Contrary to DM1, in DM2 there is no genetic anticipation, and in fact sometimes the mutation contracts during offspring, being shorter in the children. This could explain some of the differences observed in DM2, such as the lack of a congenital form (Day et al., 2003).

The role of CNBP in DM2 pathogenesis is not clear yet. It is also not clear if the level of CNBP protein is reduced in DM2 patients or not. Two different works examined the levels of CNBP mRNA and protein in DM2 cells and didn't find significant changes (Botta et al., 2006; Margolis et al., 2006). However, a different study showed that RNA-binding activity and CNBP protein levels are diminished in the cytoplasm of DM2 myoblasts and muscle biopsies from DM2 patients. Furthermore, the rate of global protein synthesis was found to be reduced in these DM2 samples (Huichalaf et al., 2009). Still further work is required to fully understand the pathomechanism of DM2 and the effect of the mutations on the *CNBP* gene.

1.4 Molecular Pathomechanism of Myotonic Dystrophy

Originally three hypotheses were proposed to explain the pathomechanism of DM: (1) the repeat expansion reduces the expression of the *DMPK* or *CNBP* genes, resulting in haploinsufficiency; (2) the repeat expansion alters the chromatin structure affecting *DMPK* and neighbouring genes such as *SIX5* or *DMWD*; and (3) RNA-induced toxicity. Experiments to confirm the first two hypotheses didn't recreate the main symptoms of myotonic dystrophy, so currently the RNA gain-of-function model is considered the most accepted and suitable model to explain the DM pathogenesis.

1.4.1 *DMPK* and *CNBP* haploinsufficiency

To explain how a mutation in a non-translated part of a gene could cause myotonic dystrophy, the *DMPK* haploinsufficiency model of pathogenesis was proposed. The first attempts to quantify the *DMPK* expression from cell lines and tissues from DM patients gave confusing results due to the use of PCR to amplify large CTG repeats, which is not very reliable. Two groups found a complete absence of *DMPK* expression in RNA extracted from DM muscle samples (Carango et al., 1993; Novelli et al., 1993). Two other groups found a reduction in the *DMPK* transcript level. Fu *et al.* found a reduction in the mutant allele, whereas Hofmann-Radvanyi *et al.* found a reduction of the wild type *DMPK* allele (Fu et al., 1993; Hofmann-Radvanyi et al., 1993). However, an increase in the level of *DMPK* RNA from DM sources has also been reported by Sabourin *et al.*, who attributed the increase to mRNA instability or an altered translational control mechanism (Sabouri et al., 1993). However, a few years later the same group generated myoblast clones overexpressing the 3'UTR of *DMPK* and found an inhibition of muscle terminal differentiation, and furthermore currently is generally accepted that overexpression of *DMPK* is not caused by the repeat expansion (Sabourin et al., 1997).

The generation of mouse models of abnormal *DMPK* expression gave

insight of the contribution of *DMPK* haploinsufficiency to the disease. A *Dmpk* homozygous knockout mice produced a late-onset progressive skeletal myopathy with mice having progressive weakness and structural abnormalities of skeletal muscle fibres. These results led to the conclusion that DMPK is important for the maintenance of the structural integrity of skeletal muscle fibres and its reduction may contribute to DM pathology (Reddy et al., 1996). Heterozygous *Dmpk* knockout mice exhibited a mild variable phenotype (Reddy et al., 1996) and another mouse where the endogenous *Dmpk* was disrupted and human *DMPK* was overexpressed also showed a mild phenotype (Jansen et al., 1996). Neither of these mouse models exhibited myotonia, which is one of the main symptoms of DM, or other characteristic symptoms such as cataracts, atrophy or male-infertility, leading to the conclusion that neither *DMPK* haploinsufficiency or gain-of-function is the only requirement for the development of myotonic dystrophy, although it contributes to the disease pathogenesis.

In the case of DM2 there is conflicting data regarding *CNBP* reduction caused by the CCTG repeats. Some studies report no effect of the CCTG repeats on the *CNBP* transcript or on the protein level (Margolis et al., 2006; Botta et al., 2006). However, recent papers have reported an abnormal expression of the *CNBP* gene caused by the CCTG repeat expansion (Huichalaf et al., 2009; Raheem et al., 2010). A *Cnbp* heterozygous knockout mice presented symptoms resembling those of DM (Chen et al., 2007), suggesting that *CNBP* haploinsufficiency contributes to the phenotype of myotonic dystrophy, nonetheless it does not appear to be the major determinant of the disease, so an additional hypothesis for the pathomechanism of DM1 and DM2 was needed.

1.4.2 Alteration of chromatin structure

Another hypothesis to explain the DM pathomechanism was that the repeat expansion could affect the expression of adjacent genes. The first report of an altered chromatin structure came from Wang *et al.*, who reported an increased nucleosome formation caused by the CTG repeats. Nucleosomes mediate general transcriptional repression, suggesting that the repeat expansion promotes the

repression of the transcription through the reorganisation of chromatin (Wang et al., 1994). Indeed, a condensed chromatin structure was found downstream of the CTG repeat expansion, which could affect transcription due to limited access of transcription factors (Otten and Tapscott, 1995). The gene *SIX5* (formerly *DMAHP*) is located downstream the *DMPK* gene and the repeat expansion overlaps with the 5' end of the promoter region of the *SIX5* gene. Observations of reduced transcription across CTG repeats led to the suggestion that expression of *SIX5* could be reduced. *SIX5* was found to be decreased in myoblasts, muscle and myocardium from DM1 patients (Thornton et al., 1997). *SIX5* is similar to the *Drosophila* homeodomain-containing protein gene *sine oculis* (SO) that is implicated in eye development. The expression of *SIX5* was examined in the human eye and it was found to be expressed in the adult cornea, lens, uvea and retina, suggesting that changes in its expression could be responsible for the pathological changes leading to cataract formation in DM patients (Winchester et al., 1999). The expression of *SIX5* has been studied in DM biopsies and cell lines, and it was found to be reduced as well, compared to controls (Korade-Mirnics et al., 1999; Alwazzan et al., 1999).

A few mouse models to study the contribution of *SIX5* to the disease pathogenesis have been generated. Disruption of *SIX5* demonstrated that formation of cataracts is inversely related to *SIX5* dosage. Furthermore, both homozygous and heterozygous mice showed increased steady-state levels of the Na^+/K^+-ATPase α-1 subunit and decreased *Dm15* mRNA levels, the mouse orthologue of human *DMPK*. Altered ion homeostasis may also contribute to the formation of cataracts, and these mice confirm the implication of *SIX5* in the development of cataracts in DM patients (Sarkar et al., 2000; Klesert et al., 2000).

DMWD (Dystrophia Myotonica-Containing WD Repeat Motif protein) is a gene located upstream the *DMPK* gene, and it is highly expressed in the brain and testis. The levels of the DM-associated *DMWD* allele were reduced in the cytoplasm by 20 to 50%, whereas they remained unchanged in the nuclear fraction. These results suggest that the export of *DMWD* transcripts might also be affected in *cis* by the repeat expansion (Alwazzan et al., 1999). Its tissue expression led to consider that haploinsufficiency of *DMWD* may be responsible

for the mental retardation and the testis atrophy observed in DM patients.

In DM2, the *CNBP* gene is not flanked by other close genes, and haploinsufficiency of *DMWD* and *SIX5* genes in DM1 contributes to the DM phenotype but is not enough to recreate all the symptoms of the disease, so the model of the altered chromatin structure was dispelled as the main pathomechanism of the disease.

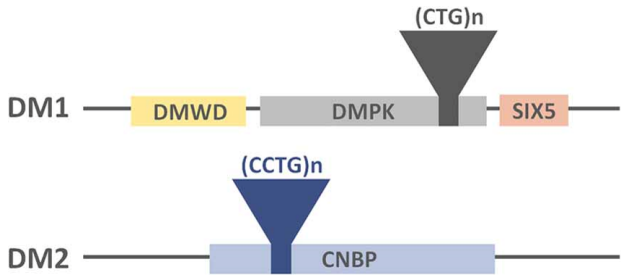


Figure 1.3: **Representation of the DM-causing genes.** The neighboring genes and the location of the repeats are shown in the picture. Taken from (Mateos-Aierdi et al., 2015).

1.4.3 RNA-induced toxicity

According to the RNA gain-of-function model the mutant RNA transcribed from the expanded allele is sufficient to induce symptoms of the disease. The expanded RNA transcript was proposed to interfere with transcription and post-transcriptional modifications, such as splicing, adenylation or cap formation, and export of the mutant transcript to the cytoplasm for its degradation (Krahe et al., 1995).

The first evidence of RNA toxicity came from Wang *et al.* who showed that the mutant transcript changed in *trans* the processing of the wild type DMPK transcript (Wang et al., 1995). The mutant transcript was also seen to accumulate in the nucleus of DM cells forming foci. Nonetheless, its export to the cytoplasm didn't appear to be impaired by the presence of the repeats, as DMPK mRNA was found also in the cytoplasm (Taneja et al., 1995). However, other studies reported a complete block of the export of the mutant transcript to the cytoplasm. Analysis of the mutant and wild type DMPK transcripts by northern blot and *in situ* hybridization revealed that the transcripts were retained in the nucleus of

myoblasts and fibroblasts where they formed stable, dynamic clusters associated with the nuclear matrix (Davis et al., 1997). Similar conclusion was obtained in another study in which allele specific RT-PCR analysis of nuclear and cytoplasmic RNA revealed a nuclear retention of expanded DMPK transcripts. This retention appears to have a critical threshold between 80 and 400 CTG repeats (Hamshere et al., 1997). In 1999 it was reported that a mutant DMPK 3'UTR with just 57 CTG repeats decreased expression of DMPK protein and resulted in the accumulation of discrete foci in the nucleus of C2C12 cells. Another mutant transcript with 200 CTG repeats also produced foci and inhibited myogenic differentiation of C2C12 myoblasts, acting as a riboregulator to inhibit myogenesis (Amack et al., 1999). The discovery in 1994 of a different myotonic dystrophy and the mapping of the CCTG mutation in 2001 to the *CNBP* gene (Liquori et al., 2001) also supported the RNA gain-of-function model of toxicity, as the two types of disease are caused by repeat sequences located in different genes and they share similar features.

Generation of the transgenic mouse model HSA^{LR} expressing 250 CTG repeats in the 3'UTR of the human skeletal actin gene strongly supported the RNA gain-of-function model. These mice developed myotonia and had muscle histology similar to DM. These results demonstrated that CUG repeats alone are sufficient to induce a DM phenotype (Mankodi et al., 2000). Later on the length of the repeats was associated with the phenotype, when two transgenic mice were generated, one carrying 20 CTG repeats (DM20) and the other carrying 300 CTG repeats (DM300). DM300 developed a DM-like phenotype including myotonia, progressive muscle weakness and nuclear foci in various tissues, whereas DM20 mice were normal (Seznec et al., 2000). The generation of an inducible mouse model demonstrated the reversible nature of RNA toxicity. Tetracycline-induced expression of 200 CTG repeats produced a DM phenotype, with mice presenting myotonia, cardiac conduction abnormalities, RNA splicing defects and CELF1 increased expression in skeletal muscle. The effects were reversed after silencing of the DMPK 3'UTR containing the CTG repeats. These mice add to the evidence that the mutant RNA plays a role in the disease pathogenesis and point at the repeat expansion as a potential therapeutic target.

The idea that the RNA gain-of-function is the main pathomechanism of myotonic dystrophy is generally accepted nowadays. How the RNA causes toxicity in the cells is a different matter that will be discussed in section 1.5. Interestingly, recent findings demonstrated that the mutant RNA is in fact exported to the cytoplasm and translated into toxic, homopolymeric proteins by Repeat-associated non-ATG (RAN) translation. These proteins should be now considered in pathogenic models of the disease.

1.4.4 RAN translation

Repeat-associated non-ATG (RAN) translation is the expression of proteins in all three frames without the need of an ATG-initiation codon. RAN translation was first discovered by Zu *et al.* while studying the molecular mechanism of spinocerebellar ataxia type 8 (SCA8). A pure polyGln protein transcribed from a CAG expansion transcript with an unusual ATG-initiated ORF was found, and furthermore, elimination of the only ATG codon in the SCA8 minigene didn't prevent the production of the polyGln protein. RAN translation was also observed to happen in all three reading frames, as polyGln, polyAla and polySer proteins were identified and seen to accumulate in the cells. RAN proteins expressed from the CAG repeats increased apoptosis, which suggest a contribution of these proteins to the disease pathogenesis (Zu et al., 2011). RAN translation was seen to occur across hairpin-forming CAG repeats, as well as hairpin-forming CUG repeats, so structured RNAs might be required for RAN translation.

DM1 myoblasts, skeletal muscle and blood from DM1 patients were used to study the presence of polyGln proteins expressed by RAN translation from the CAG antisense repeat tract. Positive immunostaining using an antibody against the predicted protein and its C-terminal region were observed in the three models used, as well as in a DM1 mouse model carrying at least 300 CTG repeats in its human *DMPK* context (Zu et al., 2011). In DM1 cells the polyGln proteins aggregated with caspase-8, which is involved in apoptosis, suggesting that in DM1 RAN proteins may also contribute to the disease. The evidence that microsatel-

lite expansion mutations can be transcribed in both directions (Batra et al., 2010) leads to the idea that up to six RAN proteins could be expressed from the expansion mutations, each of them contributing to the pathogenesis (Cleary and Ranum, 2013).

So far RAN proteins have been reported in a series of microsatellite expansion disorders, besides SCA8 and DM1, which include the fragile X-associated tremor ataxia syndrome (FXTAS) (Todd et al., 2013) and C9orf72 amyotrophic lateral sclerosis/frontotemporal dementia (ALS/FTD) (Ash et al., 2013; Mori et al., 2013). A common link between the four diseases in which RAN proteins have been reported are the neurological features, but it is still not clear how these proteins are involved in the CNS (Cleary and Ranum, 2013).

Recently, sense and antisense poly-Leu-Pro-Ala-Cys(LPAC) and poly-Gln-Ala-Gly-Arg(QAGR) RAN proteins in all three reading frames have been reported in DM2 (Zu et al., 2017). LPAC and QAGR proteins are toxic to neural cells and their expression is regulated by MBNL1. Nuclear sequestration of CCUG repeats by MBNL1 reduces the amount of RAN proteins, whereas release of CCUG repeats by overexpression of MBNL1 increases RAN protein levels. When analysed using antibodies against the RAN proteins, LPAC shows a cytoplasmic staining with small punctuate aggregates, whilst QAGR shows nuclear staining, also with small punctuate aggregates located close or at the nuclear membrane. LPAC proteins accumulate in neurons, astrocytes and glia in gray matter, and macrophages have been found in sites of LPAC accumulation. QAGR proteins accumulate in white matter. The location of the RAN proteins in the brain of DM2 patients supports the idea that these proteins may contribute to the neurological defects observed in the microsatellite expansion disorders in which RAN proteins have been reported (Zu et al., 2017).

1.5 The Molecular and Cell Biology of Myotonic Dystrophy

The RNA-induced toxicity is the most accepted model for the pathogenesis of myotonic dystrophy. However, how the RNA dominance might produce the DM-phenotype is not clear yet, and a combination of different events in the cell, such as aberrant alternative splicing of many pre-mRNAs, altered transcriptional regulation, CUG-hairpin induced stress pathways that inhibit translation, RAN translation of repeat expansions (producing toxic peptides), or dysregulation of miRNA processing or function by the repeats, is though to contribute to the disease pathogenesis.

The protein sequestration model was proposed to explain how the toxic RNA causes DM. This model suggests that the repeat expansion transcripts sequester RNA-binding proteins and depletion of these proteins cause their abnormal function. Different RNA-binding proteins can be sequestered. CUG-Binding Protein 1 (or also CELF1) was one of the first RNA-binding proteins reported to bind CUG repeats in DM1 (Timchenko et al., 1996). It was later shown that CELF1 is indeed affected by the toxic RNA, however these proteins are not sequestered in ribonuclear foci, instead they are upregulated by PKC phosphorylation (Michalowski et al., 1999). CELF1 belongs to a family of RNA processing factors and it is implicated in the alternative splicing of many different transcripts (Ladd et al., 2001). Another RNA-binding protein implicated in the alternative splicing of many pre-mRNAs that has been seen to be sequestered by the repeat expansions is Muscleblind-like 1 (MBNL1). MBNL1 interacts with the hairpin structure of the repeats and is sequestered in the ribonuclear foci in both DM1 and DM2, leading to alterations in the alternative splicing of many genes, which contributes to the clinical manifestations of the disease (Miller et al., 2000; Mankodi et al., 2001). The RNA-binding proteins hnRNP H and hnRNP F also colocalize with foci and appear to be sequestered as well (Jiang et al., 2004).

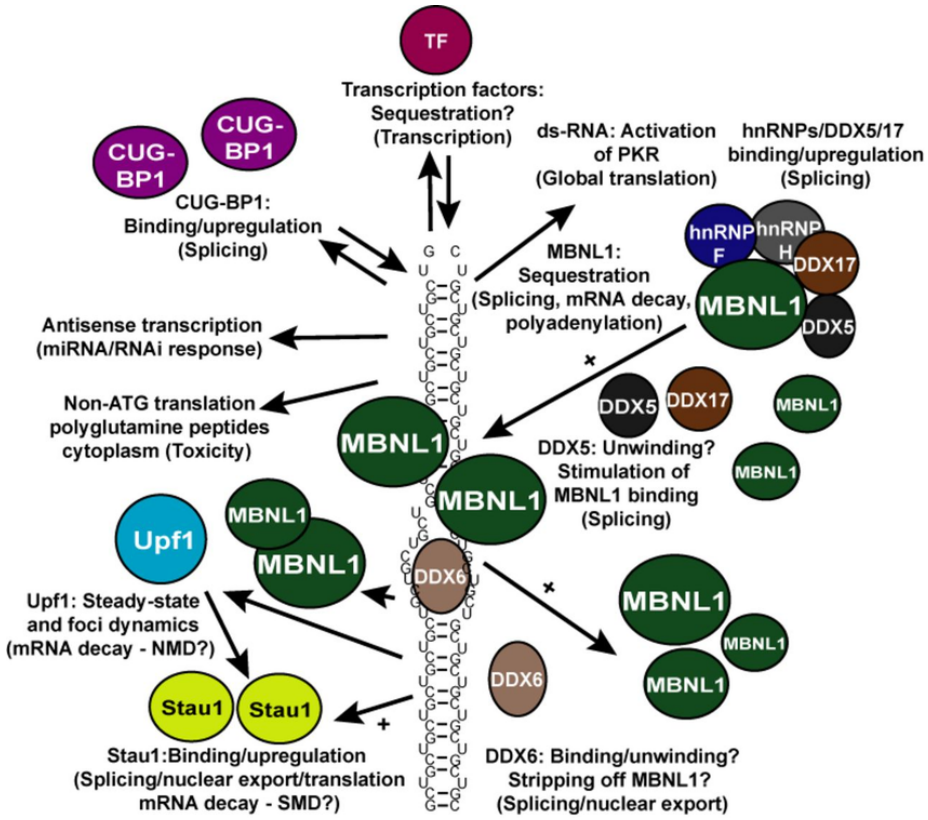


Figure 1.4: **DM1-proposed disease mechanisms of CUG repeat expansions.** CUG repeat expansions form a hairpin structure which sequester several RNA-binding proteins including MBNL1, DDX6 and possibly transcription factors (green, gray and purple circles, respectively). Examples of the molecules and the processes or pathways they affect are presented. Overall, MBNL1 is sequestered by the CUG repeats, enhanced by some of its interacting proteins, the dead box helicases DDX5 or DDX17, promoting foci formation. On the contrary, another dead box helicase DDX6 unwinds the hairpin and potentially releases MBNL1 to allow for nuclear export and translation. Staufen1 (Stau1) stimulates this action performed by the DDX6. Miss-regulation of MBNL1 has an important impact on the alternative splicing of many pre-mRNAs. Another splicing factor is affected by the repeats, but it doesn't interact with the hairpin structure. CELF1 is upregulated by phosphorylation of the protein, also affecting the alternative splicing of pre-mRNAs. Taken from (Pettersson et al., 2015).

1.5.1 CELF1 in DM

CELF1 (CUGBP-Elav Like Family member, or also known as CUGBP1), was identified *in vitro* in HeLa cells, as a protein with ability to bind specifically $(CUG)_8$ oligonucleotides. This protein was localized mainly in the nucleus and was found to be associated with polyadenylated RNAs *in vivo* (Timchenko et al.,

1996). It was later established that CELF1 is an RNA-binding protein of around 50 KDa that binds to GU-rich elements present in the 3'UTR of target transcripts and regulates degradation of these transcripts (Vlasova et al., 2008). CELF1 can also bind to transcripts containing a GC-rich element in their 5'UTR and promote their translation (Timchenko et al., 2005). The ability of CELF1 to bind RNA led to the proposal that it binds to the CUG repeats in DM1 and is sequestered in foci (Timchenko et al., 1996). Nonetheless, it was later clarified that CELF1 does not bind to CUG repeats, especially double-stranded CUGs, and is not sequestered in foci (Michalowski et al., 1999).

In addition to the role of CELF1 in mRNA decay of short-lived transcripts, it has a role in translation and alternative splicing. CELF1 increases the translation of *CDKN1A* (Timchenko et al., 2001), p21 and *Mef2a* (Timchenko et al., 2004). Increase of *Mef2a* is suggested to contribute to delayed myogenesis (Timchenko et al., 2004) and increase of p21 could cause impaired cell cycle withdrawal in DM1 muscle cells (Timchenko et al., 2001).

Its role in alternative splicing is very important in DM pathogenesis. CELF1 regulates the alternative splicing of the insulin receptor (*INSR*), chloride channel 1 (*CLCN1*) and cardiac troponin T (*TNNT2*), among others, which are aberrantly spliced in myotonic dystrophy because of the effect of the toxic transcripts on proteins involved in alternative splicing (Philips et al., 1998; Savkur et al., 2001; Charlet et al., 2002; Ho et al., 2005). CELF1 promotes the embryonic splice form of its target genes. It induces skipping of exon 11 of *INSR* by binding to an exonic splicing enhancer on exon 11 and an intronic splicing enhancer on the upstream intron (Sen et al., 2009). CELF1 also promotes inclusion of exon 7a of *CLCN1* and inclusion of exon 5 of *TNNT2*.

The repeat expansion in DM1 causes an increase of the steady-state level of CELF1 caused by PKC (Protein Kinase C)-mediated phosphorylation. The mutant RNA activates the PKC signalling pathway through an unknown mechanism, and PKC activation increases CELF1 stability and phosphorylates this protein *in vitro*. PKC has been seen to be activated in heart tissues from DM1 patients and a DM1 mouse model (Kuyumcu-Martinez et al., 2007). In the case

of DM2, increase of CELF1 in muscle biopsies has not been detected (Lin et al., 2006), nonetheless, it has been observed to be upregulated in the liver of DM2 transgenic mice (Salisbury et al., 2009).

1.5.2 Muscleblind in DM

Muscleblind-like proteins (MBNL) are a family of RNA-binding proteins encoded by three genes in mammals *MBNL1*, *MBNL2* and *MBNL3* on chromosomes 3, 13 and X, respectively (Fardaei et al., 2002). The three members of the family are differently expressed. MBNL1 and MBNL2 are ubiquitously expressed, and MBNL1 has a role in most tissues, except in the brain, where MBNL2 is predominantly expressed. MBNL3 is much more restricted, with some roles in muscle differentiation and regeneration and a strong expression in placenta (Fardaei et al., 2002; Poulos et al., 2013). All three MBNLs have four zinc-finger (ZnF) domains important for recognizing a common consensus sequence in their pre-mRNA and mRNA targets. These zinc-finger domains play a key role in the most important function of the proteins, the regulation of the alternative splicing (Wang et al., 2012). In addition to the role in alternative splicing, MBNL proteins also mediate cellular mRNA transport and stability and miRNA processing (Adereth et al., 2005; Rau et al., 2011; Wang et al., 2012).

MBNL1 is predominantly expressed in muscle, and thus is the best-studied member of the MBNL family. MBNL1 is encoded in the chromosome 3q25.1. It has 12 exons, 10 of which correspond to the coding sequence and six (exons 3, and 5 to 9) undergo alternative splicing, which can result in the production of at least seven *MBNL1* mRNA variants (Pascual et al., 2006; Tran et al., 2011). Inclusion of exons 5 and 7 enhances sequestration of MBNLs in the nucleus of DM cells (Terenzi and Ladd, 2010). Exon 5 and the conserved amino acid KRAEK motif in exon 6 compose a bipartite nuclear localization signal. Inclusion of exon 7 induces formation of multimeric ring-like structures upon binding to the expanded repeats (Yuan et al., 2007; Tran et al., 2011).

Exons 1, 2 and 4 encode the ZnF domains arranged in two similar structured tandem pairs that constitute ZnF1/2 and ZnF3/4. Each of the motifs has

the amino acid sequence CCCH (Pascual et al., 2006). The ZnF domains recognize the YGCY motif in the RNA targets, and also the UGCU, especially when repeated (Ho et al., 2004; Goers et al., 2010). In the characterization of the ZnF domains, depletion of either ZnF1 or ZnF4 greatly reduced the interaction of MBNL1 with the CUG repeat expansions *in vivo*. Also, depletion of a linker sequence encoded by exon 3 that separates two ZnF domains also reduced the MBNL1 binding ability (Kino et al., 2004). This linker could support the protein flexibility, enabling MBNL1 to interact with its targets. However, MBNL1 isoforms lacking exon 3 can interact with CUG repeat expansions and are sequestered in foci (Cass et al., 2011; Tran et al., 2011). The binding of MBNL1 to its hnRNA targets allows the performance of one of its function, which is alternative splicing of the targets. It has been shown that the ZnF1/2 motifs are predominantly required for an effective splicing regulation of most of the MBNL1 targets, whereas the ZnF3/4 has higher RNA-binding affinity, helping in the binding step of the MBNL1 to its target RNA (Purcell et al., 2012; Hale et al., 2018).

MBNL1 regulates the alternative splicing of many pre-mRNAs, including *INSR*, *TNNT2* and *CLCN1*. It has an antagonistic role to CELF1. Sequestration of MBNL1 in foci causes the missplicing of many mRNAs, and this has been demonstrated to contribute to the DM phenotype. Knockdown of MBNL1 in cells produced the aberrant splicing of *INSR* and *cTnT* (Ho et al., 2004), whereas the aberrant splicing of *INSR* was rescued by overexpressing MBNL1 in DM1 myoblasts (Dansithong et al., 2005). Furthermore, MBNL1 knockout mice exhibited the same splicing defects observed in DM1, supporting the implication of MBNL1 in alternative splicing regulation. However, these mice did not have muscle wasting, suggesting that the aberrant alternative splicing may not be sufficient to induce all DM symptoms (Kanadia et al., 2003).

1.5.3 Aberrant alternative splicing in DM

Alternative splicing is a post-transcriptional process that occurs before the translation of the mRNAs into proteins. After the splicing of the introns, some of the

coding exons can be either retained or spliced out to generate different sets of mRNAs, which results in several protein isoforms from a single gene. Most of the genes undergo alternative splicing, thus misregulation of splicing has been linked to different diseases (Cooper et al., 2009). The CTG and CCTG repeat expansions in DM1 and DM2 have been shown to sequester MBNL1 proteins and aggregate in the nucleus of the cells in the form of foci, which alters the function of these proteins and has an impact in the regulation of the alternative splicing (Miller et al., 2000; Mankodi et al., 2001). Another protein implicated in alternative splicing that is affected in DM1 by the CTG repeats, but not in DM2, is CELF1. Unlike MBNL1, CELF1 is not sequestered in foci, instead it is overexpressed by PKC phosphorylation (Timchenko et al., 2004).

MBNL1 and CELF1 are antagonists in the regulation of alternative splicing. CELF1 promotes the embryonic splice forms of its target mRNAs, whereas MBNL1 promotes the adult splicing form of the same targets. Overexpression of CELF1 and sequestration of MBNL1 alters the imbalance of these proteins and has been proposed to reverse a subset of genes to their embryonic isoform, contributing to the DM phenotype (Lin et al., 2006). CELF1 promotes the skipping of *INSR* exon 11, the inclusion of *cTnT* exon 5 and the inclusion of *CLCN1* exon 7a, whereas MBNL1 does the opposite (Sen et al., 2009; Kino et al., 2009). Exclusion of the *CLCN1* exon 7a by MBNL1 occurs when the protein binds to a hairpin in the junction between intron 6 and short exon 7a, which promotes the skipping of the exon (Kino et al., 2009). Skipping of exon 5 of *cTnT* occurs when U2 snRNP recruitment is inhibited after the splicing factor U2AF65 can not recognize the polypyrimidine tract. This blocking is caused when MBNL1 interacts with a stem-loop structure in a portion of the intron upstream the exon 5 (Warf et al., 2009a).

More than 25 splicing events are misregulated in DM1 and some of them have been associated with some of the DM symptoms (Ranum and Cooper, 2006). The aberrant splicing of *INSR* produces a protein isoform with a lower signalling capacity and a decreased metabolic response to insulin, which has been associated with insulin resistance in DM1 patients (Savkur et al., 2001). The Clc1 encoded by the *CLCN1* gene is a voltage-gate chloride channel that

transports chloride ions across cell membranes. It stabilizes the resting membrane potential and regulates the electrical excitability in skeletal muscle (Dutzler et al., 2002). Aberrant alternative splicing of *CLCN1* which includes retention of the intron 2, and thus inclusion of a premature stop codon, or inclusion of exon 7a has been associated with myotonia in DM1 and DM2 (Charlet et al., 2002; Mankodi et al., 2002). A direct link between missplicing of the *CLCN1* and myotonia was established when reversal of the *CLCN1* using a morpholino in the HSA^{LR} mice rescued myotonia (Wheeler et al., 2007). Myotonia is a hallmark of myotonic dystrophy and is caused by repetitive firing of action potentials that delay skeletal muscle relaxation after muscle contraction (Harper, 2001).

Cardiac Troponin T (*cTnT*) was the first gene found to be misspliced in DM1. The binding of CELF1 to the cTnT pre-mRNA causes the inclusion of the exon 5 and the missplicing of this gene is thought to contribute to the cardiac conduction defects observed in DM (Philips et al., 1998). Muscle weakness is another characteristic symptom of DM that has been recently associated with the missplicing of *BIN1* (Bridging Integrator 1) and *CACNA1S* (calcium channel voltage-dependent L type alpha 1S subunit). BIN1 is required for the biogenesis of muscle T tubules, which are essential structures for excitation-contraction coupling of the skeletal muscle membrane. MBNL1 regulates the alternative splicing of BIN1, but in DM the missplicing of the gene produces an inactive protein that lacks phosphatidylinositol 5-phosphate-binding and membrane-tubulating activities, contributing to muscle weakness (Fugier et al., 2011). Skeletal muscle excitation-contraction coupling is also controlled by CACNA1S. In DM1 and DM2 patients exon 29 of *CACNA1S* is skipped, which correlated with severity of muscle weakness in tibialis anterior muscle of DM1 patients (Tang et al., 2012).

The misplicing of the rest of the genes has not been linked yet to any of the symptoms, but they are thought to contribute to the DM phenotype. The severity of the disease correlates with the amount of MBNL1 sequestered in foci (Miller et al., 2000; Cardani et al., 2014), so aberrant alternative splicing is an important factor in myotonic dystrophy that contributes to the pathogenesis, and a suitable target to develop a therapy.

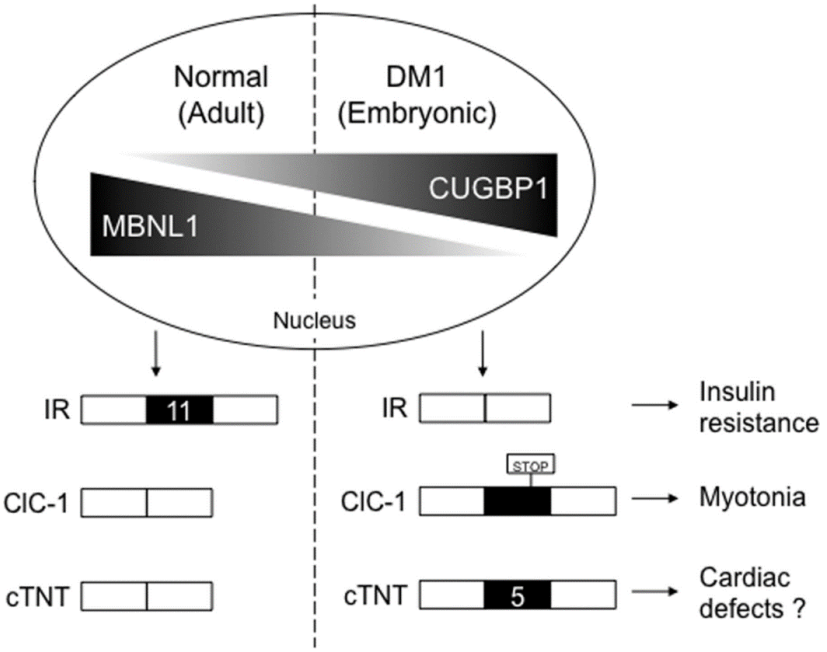


Figure 1.5: **The balance of MBNL1 and CUGBP1 regulate a subset of splicing events that are reversed in DM1.** In a healthy individual the levels of MBNL1 in the nucleus are high whereas the levels of CUGBP1 are low, which causes the production of the adult isoforms of the genes *INSR*, *CLCN1* and *cTNT*. In DM1 MBNL1 is sequestered in the ribonuclear foci and its level is reduced in the nucleus and CUGBP1 is overexpressed by PKC phosphorylation. The imbalance of these proteins leads to the production of embryonic splice isoforms of the genes aforementioned. The aberrant alternative splicing of each of these genes has been linked to a symptom characteristic of DM1. Taken from (Lee and Cooper, 2009).

1.5.4 Deregulation of microRNAs

MicroRNAs (miRNAs) are short non-coding RNA molecules typically of 22 nucleotides that function as post-transcriptional regulators of gene expression by inhibiting translation or degrading target mRNAs. More than 60% of all transcripts are regulated by one or more miRNAs (Bartel, 2004). Several miRNAs have been shown to be deregulated in DM1 and DM2. This deregulation alters the expression of target transcripts, potentially contributing to disease pathogenesis (Perbellini et al., 2011). Some of the miRNAs that are deregulated in DM1 include miR-1, miR-206 and miR-335, which are upregulated in DM1 skeletal muscle (Perbellini et al., 2011; Gambardella et al., 2010); miR-29b, miR-29c, miR-33 and miR-133, which are downregulated in DM1 skeletal muscle

or have an altered subcellular localization in DM1 myofibres (Perbellini et al., 2011). In the case of DM2, deregulation of miRNAs has also been observed. In a study where miRNAs levels were studied in skeletal muscle biopsies of 13 DM2 patients, miR-34a-5p, miR-34b-3p, miR-34c-5p, miR-146b-5p, miR-208a, miR-221-3p and miR-381 were found to be upregulated, whereas miR-125b-5p, miR-193a-3p, miR-193b-3p and miR-378a-3p were downregulated compared to control samples (Greco et al., 2012). All these miRNAs regulate transcripts involved in different cell processes, such as muscle differentiation, myogenesis, cardiogenesis, metabolism, apoptosis, etc (Perbellini et al., 2011; Greco et al., 2012). Interestingly, miRNAs have been seen to play a protective role in polyQ (polyglutamine) induced neurodegeneration (Jiang et al., 2005). Based on this finding, deregulation of miRNAs in DM1 and DM2 could be contributing to the neurodegenerative features of the disease.

1.6 Animal Models for Myotonic Dystrophy

DM-patient cell lines are easy to obtain, and in the absence of these cell lines, different models have been created for myotonic dystrophy research by simply transfecting cells with plasmids containing the repeats or engineering them. Plasmids with the repeats can be prepared in the lab, however due to the instability of the repeats and the difficulty to amplify them by PCR, it is difficult to generate plasmids with long and pure CTG or CCTG repeats. Currently there are a few of these plasmids containing CTG repeats, ready to transfect into the cells, such as the DM480 plasmid created by Yoshihiro *et al.* (Kino et al., 2009), or a plasmid containing 960 interrupted CTG repeats (Philips et al., 1998). They have been transfected into HeLa cells or C2C12 myoblasts many times in order to create a DM1 cell model, and used for different purposes. Ultimately, the cell models are very useful but the information we can get from them is limited. We need animal models that recreate the phenotypic features of DM.

Animal models in science are a valuable source of knowledge with the advantages of an unlimited supply, ease of manipulation, low costs and high genetic similarity to humans. Over the years, in order to investigate the patho-

physiology of myotonic dystrophy, several mouse models have been developed. Currently there are available some important models, however none of them fully recapitulates the DM1 and DM2 phenotype, and some of them come with associated problems at the time of being properly generated due to the instability of the repeat expansion. The mouse models available could be classified into three categories, depending on the manipulations made on them: inactivation of genes of the DM1 or DM2 locus, overexpression of toxic repeats or introduction of these unstable repeats, and inactivation of *MBNL1* or overexpression of *CELF1* in order to study abnormal alternative splicing regulation.

1.6.1 Inactivation of genes of the DM1 or DM2 locus

Dmpk$^{-/-}$, *Six5*$^{-/-}$, *Cnbp*$^{+/-}$ and *Cnbp*$^{-/-}$ mice have been developed. The *Dmpk*$^{-/-}$ and *Six5*$^{-/-}$ mice failed to reproduce the full DM1 phenotype, confirming that haploinsufficiency of these genes in the DM1 locus is not the primary mechanism of the disease, however their loss contributes to the development of some of the DM1 features, such as cataracts or cardiac dysfunction (Reddy et al., 1996; Sarkar et al., 2000). On the contrary, *Cnbp*$^{+/-}$ and *Cnbp*$^{-/-}$ mice showed many of the DM2 features, including myotonia, mild muscle weakness, defective walking, cardiac conduction defects and cataracts, suggesting that *CNBP* haploinsufficiency plays a more significant role in DM2 pathology (Chen et al., 2007).

1.6.2 Transgenic models of toxic RNA gain-of-function

Several mouse models have been generated in order to test the hypothesis of the toxic RNA gain-of-function as the pathomechanism of myotonic dystrophy. Currently the most suitable mouse model to investigate DM1 is the *HSA*$^{\mathrm{LR}}$ mice carrying 250 CTG repeats in the human skeletal actin (HSA) gene (Mankodi et al., 2000). These mice show sequestration of MBNL1 in foci but not CELF1 upregulation, and abnormal splicing of several genes (e.g. *Clcn1*, *Mbnl1*, *Serca*, *Atp2a1*), suggesting that MBNL1 loss of function is enough to alter the alternative splicing. However they do not present myopathy or muscle wasting.

Another mouse model, DM300, carries a larger CTG repeat expansion (300 to 600) in the context of the human DM1 locus. These mice show foci in multiple tissues, myotonia, progressive muscle weakness, muscle histopathology, *Insr* missplicing associated with defects in glucose metabolism, growth retardation and high mortality (Seznec et al., 2000). The intergenerational instability of DM300 led to mice carrying 1000 to 1800 CTG repeats. These mice, DMSXL, present a more severe phenotype and could be used as a model for congenital DM1 (Gomes-Pereira et al., 2007). Despite the fact that DM300 and DMSXL recapitulate multisystemic DM1 features, their phenotype is relatively mild compared to other mouse models.

These mouse models constitutively express the CTG repeats, however a different approach to create the models is the inducible expression of expanded CTG repeats or the tissue-specific expression. The Cre-lox system allows tissue-specific expression of interrupted CTG repeats. EpA960 is an inducible mouse model for heart-specific expression of 960 interrupted CUG RNA repeats in the 3'UTR of the human version *DMPK* exon 15. After tamoxifen-induction of the expression of the expanded repeats in the cardiac muscle these mice developed severe cardiomyopathy and arrhythmias, as well as molecular features of DM1 such as RNA foci, CELF1 upregulation and sequestration of MBNL1 in foci (Wang et al., 2007). Expression of the expanded CUG repeats in skeletal muscle in EpA960/HSA-Cre-ER^{T2} mice resulted in the formation of RNA foci, sequestration of MBNL1 proteins, aberrant alternative splicing of a subset of genes, myotonia and progressive muscle impairment (Orengo et al., 2008).

Inducible overexpression of *DMPK* 3'UTR is a different but also viable option to create a mouse model of myotonic dystrophy. GFP-DMPK-$(CTG)_{200}$ mice carrying 200 CTG repeats had RNA foci and sequestration of MBNL1, however the repeat number was insufficient to generate a noticeable DM1 phenotype (Mahadevan et al., 2006). Another mouse model, DMPK-GFP-$(CTG)_{91}$ displayed defective muscle differentiation, but the phenotype was aggravated by the repeat expansion (Storbeck et al., 2004).

1.6.3 Mouse models for abnormal alternative splicing

The role of MBNL1 and CELF1 in the pathomechanism and the DM phenotype has been studied using mouse models created by knocking out MBNL1 or overexpressing CELF1. Knock out of *Mbnl1* by disruption of the exon 3 of the gene in the *Mbnl1*$^{\Delta 3/\Delta 3}$ mice recapitulates some of the symptoms of DM1, such as myopathy, myotonia associated with missplicing of *Clcn1*, cataracts, apathy and conduction defects. However, muscle weakness was not detected, suggesting that it might result from different mechanisms, such as CELF1 upregulation (Kanadia et al., 2003; Matynia et al., 2010).

MBNL2 also colocates with foci in DM1 and DM2. Two MBNL2 knock out mice were generated. In the first one, reduction of the *Mbnl2* mRNA by more than 90% didn't recapitulate any of the DM1 symptoms (Lin et al., 2006). The second *Mbnl2* knock out mouse model developed myotonia, skeletal muscle pathology and *Clcn1* missplicing, suggesting that MBNL2 might play a role in DM pathogenesis (Hao et al., 2008).

The role of CELF1 in myotonic dystrophy has been clarified after the generation of CELF1-overexpressing mouse models. The first transgenic mice overexpressing CELF1, CUGBP1-TR line, suffered from muscular dystrophy, growth retardation, histological and molecular abnormalities and overxpression of *Mef2a* and *p21* (Timchenko et al., 2004). MCKCUG-BP1, another transgenic mice overexpressing CELF1 in heart and skeletal muscle had missplicing of *Tnnt2*, *Mtmr1* and *Clcn1* (Ho et al., 2005). Nevertheless, these mice had high mortality and were difficult to breed. TRECUGBP1 mice was developed as a conditional CELF1 overexpression model. Induced overexpression in adult skeletal muscle resembles DM1 muscle phenotype and reproduced missplicing characteristic of DM1 (Ward et al., 2010).

1.6.4 Mouse models of DM2

Many things are known about DM2, however it has been less studied than DM1 since it was discovered. Compared to the amount of mouse models available for

DM1, for DM2 there are only very few models. *CNBP* was knocked down in the *Cnbp*$^{+/-}$ and *Cnbp*$^{-/-}$ mice, resulting in the development of myotonia, cardiac conduction defects, cataracts, defective walking and muscle defects, suggesting a pathogenic role for *CNBP* haploinsufficiency (Chen et al., 2007). Similar to HSA^{LR} mice for DM1, DM2-HSAtg mice are transgenic expressing an intronic $(CCTG)_{121}$ expansion. These mice have RNA foci, increased levels of the translational targets of CELF1 and recapitulate part of the DM2 muscle phenotype, although no missplicing was detected (Salisbury et al., 2009). These mouse models are useful to study DM2, however, improved transgenic mice that recapitulate more features of the disease are crucial to gain insight into specific aspects of DM2 and develop a therapeutic strategy.

1.7 Therapies for Myotonic Dystrophy

There is currently no 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 DNA repeats, the toxic mRNA transcripts, RNA-binding proteins, splicing regulators and misspliced transcripts. There are many levels at which the therapy can be directed, and different strategies have been developed and tested.

1.7.1 Targeting the DNA repeats

The repeat expansion is highly unstable in the germline and in somatic cells. To date, targeting the DNA repeats has been difficult because the expansion 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. One of the first reports of a small molecule that binds to repeat expansions came from Nakatani *et al.*. In this study, the napthyridine-azaquinolone (NA) ligand was found to bind CAG repeats with high affinity, increasing the stability of the NA-CAG complex (Nakatani et al., 2005).

A different way 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 disrupted nuclear foci and released MBNL1, correcting the missplicing of *cTnT*, *CLCN1*, *INSR* and *SERCA1* in a HeLa cell model transfected with 960 interrupted CTG repeats and in a HSA^{LR} mice model (Warf et al., 2009b). Pentamidine is able to bind DNA and it was originally thought to bind the CUG repeats. Nevertheless, it was later demonstrated that it acts by inhibiting the transcription of the CTG repeats in myotonic dystrophy (Coonrod et al., 2013). Actinomycin D is another small molecule with high affinity for double-stranded DNA, specially GC-rich sequences, but not for RNA. It inhibited transcription of the CTG repeats, producing a reduction in the number of foci and a partial rescue of the genes *CLCN1*, *ATP2A1*, *MBNL1*, *VPS39*, *NFIX* and *LDB3* (Siboni et al., 2015). The discovery of these small molecules suggested that inhibition of the transcription of the CTG repeat expansion could be a suitable therapy to prevent expression of toxic RNA.

A recently discovered genome-editing technology, the CRISPR system, has also proved to be effective to target the DNA repeats of microsatellite expansion disorders. The CRISPR system (explained in detail later) 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, that can then be repaired by NHEJ, introducing random mutations, or by the introduction of precise mutations using a repair template. The use of two gRNAs, one upstream of the CTG repeats and the other downstream, allowed for the removal of the CTG tract in DM1 myoblasts from a mouse model carrying around 500 repeats and in DM1 myoblasts from a DM1 patient carrying 2600 CTG repeats (van Agtmaal et al., 2017). Interestingly, cleavage upstream or downstream of the repeat expansion tract with a single gRNA also produced elimination of the repeats by triggering uncontrollable deletion of the repeats. Elimination of the repeats did not affect the transcription of *DMPK* or its neighbour gene *SIX5*. The alternative splicing of exons 12 to 15 of the *DMPK* transcripts was not altered either by the loss of repeats. *DMPK* transcripts lacking

the CTG repeats were found mainly in the cytoplasm and therefore no foci were observed in the nuclei of the cells and MBNL1 was released from the ribonuclear complex (van Agtmaal et al., 2017). Provenzano *et al.* adopted a similar approach to eliminate the repeats. They used 4 different gRNAs, two upstream the CTG tract and the other two downstream, to excise the repeat expansion from myogenic cells derived from DM1-patients fibroblasts. Clones without the repeats didn't have foci in their nucleus and presented a normal alternative splicing of *INSR* and *SERCA1* (Provenzano et al., 2017). Off-target effects could not be detected in neither of the CRISPR works, suggesting that it is a suitable technology to eliminate the repeat expansions in microsatellite expansion disorders.

Inhibition of the transcription of the CTG repeats can also be achieved using the CRISPR system. By using a deactivated version of the Cas9 (dCas9), Pinto *et al.* impeded the transcription of the CTG repeats in a length-, PAM- and strand-dependent manner. The dCas9 bound the CTG repeats blocking the action of the RNA polymerase II, reducing the levels of toxic transcripts. It also rescued several missplicing events, and specifically, correction of the *Clcn1* splicing reduced myotonia in a DM1 mouse model. The rate of RAN translation was also decreased, as the levels of RAN proteins were drastically reduced. Inhibition of the transcription of the toxic repeats is another way of using the CRISPR system that does not involve the cleavage of the genome, and thus may be more desirable (Pinto et al., 2017).

1.7.2 Targeting the toxic RNA

The toxic RNA has been the main target to develop a therapy. Ribozymes and antisense oligonucleotides have been widely used to either degrade or inhibit transcription of the toxic RNA. Ribozymes are RNA molecules that can cleave other RNAs leading to their degradation or to splicing defects that replace the target RNAs with embedded sequences (Sioud, 2004). The first applications of ribozymes to treat DM1 were performed *in vitro*, when a tetrahymena group 1 intron ribozyme cleaved a *DMPK* RNA sequence with 12 CTG repeats and replaced them with 5 repeats contained in the engineered ribozyme. A similar

replacement of repeats was also achieved when ribozymes were applied to DM fibroblasts (Phylactou et al., 1998; Phylactou, 2004). A hammerhead ribozyme was used with the aim to target the mutant allele over the wild-type. Nonetheless, both transcripts were greatly reduced, producing a reduction in the number of foci and a partial restoration of alternative splicing of the *INSR* gene (Langlois et al., 2003). Ribozymes still need to be optimized before they can be effectively used for gene therapy.

Antisense oligonucleotides (AONs) have also been important in the development of a therapy for myotonic dystrophy. AONs are short, single-stranded sequences of nucleic acids that hybridize by Watson-Crick base-pairing to complementary sense mRNAs, inhibiting the expression of the target mRNA and in some cases, the mRNAs can be degraded. The term "antisense oligonucleotide" comprises four chemically different molecule types, but only two of them are being employed to develop a therapy for DM, the 2'-O-methyl modified AONs (MOE) and the phosphorodiamidate morpholino antisense molecule (PO or Morpholino). An advantage of morpholinos is their increased stability, improved permeability and high specificity over other AONs (White et al., 2009). Despite the restrictions to the application to cell culture, one of the first studies that applied AONs in DM was quite promising. Using an adenoviral vector expressing an antisense RNA to target the CUG repeats, the level of the mutant *DMPK* mRNA was reduced and CELF1 was downregulated. The missplicing of *INSR* was partially rescued, improving the ability to respond to insulin (Furling et al., 2003). A few years later, the use of a $(CAG)_7$ MOE selectively targeted the mutant transcript. Delivery of the AON by local injection and electroporation into two different DM1 mouse models produced the degradation of the mutant transcript, reduced foci and partially rescued some misspliced genes. However, the proper administration remains a limitation for these AONs (Mulders et al., 2009).

Triggering of RNase H-mediated degradation of the toxic RNA was achieved by engineering a chimeric AON, gapmer, containing modified nucleic acid residues. The RNase H is a non-sequence-specific enzyme that recognizes DNA-RNA heteroduplexes and cleaves the RNA strand. To induce the RNase H-mediated degradation of an RNA, the AONs used need to have a "gap" region

with 7-10 nucleotides containing RNase H-compatible phosphorothioate (PS) DNA (Stein and Hausen, 1969; Lima et al., 2004). Administration of the gapmer to cultured cells and DM1 mouse models preferentially reduced the mutant transcript and RNA foci by RNase H degradation of the toxic RNA. However, optimization of the delivery methods is required to apply this strategy as a therapy for myotonic dystrophy (Lee et al., 2012).

The CUG and CCUG expanded repeats form a hairpin structure which is a substrate for Dicer, a ribonuclease that recognizes double-stranded RNA duplexes and induces cleavage of these structures into shorter 21-nucleotide duplexes. The short duplexes are then incorporated into the RISC complex (RNA-induced silencing complex) and guide substrate selection for degradation. This mechanism is active in cells containing CUG repeats (Soifer et al., 2007; Krol et al., 2007). This suggested that siRNAs (small interfering RNAs) could serve as a therapy to target the mutant RNA in DM1, which would produce the degradation of the toxic RNA by the Dicer pathway. Indeed, a shRNA (short hairpin RNA) which is then processed into a 21-nt siRNA by RISC, delivered into DM1 myoblasts produced a preferential reduction of the mutant transcript by triggering its degradation. Nonetheless, the RNA interference approach needs optimization before its use *in vivo*, specially regarding the delivery and specificity.

The CRISPR system can 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 the alternative splicing of several genes such as *INSR*, *ATP2A1*, *SERCA1*, *MBNL1* and *MBNL2* (Batra et al., 2017). Thus, the CRISPR system can be used as a therapy for myotonic dystrophy and can be directed to different levels of targets involved in the pathogenesis of the disease.

1.7.3 Targeting protein-RNA binding

The toxic RNA sequesters MBNL1 proteins, altering its function in alternative splicing and causing the missplicing of many genes, which contributes to the disease pathogenesis. Thus, the protein-RNA complex is another target to develop a therapy for myotonic dystrophy, and different technologies have been used to prevent the binding of MBNL1 to the toxic RNA, including small molecules and engineered oligonucleotides. A series of studies reported the design of several compounds improved over the time by modifying their chemical structure that target the CUG repeats and displace MBNL1 from foci. One of these first compounds tested in a DM model was the Ligand 1, which contains the DNA intercalator acridine conjugated with a triaminotriazine unit that recognizes U-U. Ligand 1 destabilized the complexes formed between MBNL1 and $(CUG)_4$ and $(CUG)_{12}$ (Arambula et al., 2009). Despite the little potential of Ligand 1 to become an anti-DM drug because of its poor solubility and failure to penetrate the cellular membrane, its scaffold has served as starting point to develop other molecules that target the CUG repeats. Some of the molecules that emerged by modifying the structure of Ligand 1 are the New Ligand 1, which inhibited the formation of the complex between MBNL1 and $(CUG)_{12}$ and corrected the *INSR* missplicing in a HeLa cell DM1 model (Jahromi et al., 2013). Ligand 3 was designed by linking two triaminotriazine units with a bisamidinium component. It was then combined into a new bivalent ligand, Ligand 2a. Both ligands reduced foci, partially rescued the missplicing of *cTnT* and *INSR* and rescued the degenerative phenotypes of DM1 in a *Drosophila* model (Wong et al., 2014; Luu et al., 2016).

Similarly to the ligands reported to inhibit the complex MBNL1-$(CUG)_{exp}$ in DM1, in DM2 a series of ligands with the same function have been reported recently. All of them contain two main units: a triaminopyrimidine unit for targeting the CU mismatch and a bisamidinium unit which provides a general A-form RNA targeting ability and confers greater water-solubility and cellular uptake. Additional chemical groups have been added to this scaffold to create different ligands. Some of them effectively inhibited the formation of the

complex MBNL1-$(CCUG)_8$, dispersed foci and had very low toxicity in a DM2 cell model (Nguyen et al., 2014).

Antisense oligonucleotides can be used to prevent the formation of the MBNL1-CUG complex. A CAG_{25} morpholino, a 25-mer antisense oligonucleotide composed of CAG-repeats, was used *in vitro* and *in vivo* in the HSA^{LR} mouse model and in both models the morpholino reduced nuclear foci by binding the CUG repeats and preventing it from binding with MBNL1, and disrupted already formed complexes, redistributing MBNL1. Release of MBNL1 partially rescued the missplicing of *Clcn1*, *Serca1*, *m-Titin* and *Zasp*. Furthermore, rescue of the *Clcn1* markedly reduced myotonia in the DM1 mouse model (Wheeler et al., 2009). These results demonstrate that antisense oligonucleotides are another technology that can be used to target the MBNL1-CUG complexes, producing beneficial effects in DM1 models.

Childs-Disney *et al.* reported two small molecules that inhibit the binding between MBNL1 and CUG repeats. Compound 1 is a thiophene-containing small molecule that binds MBNL1 in the ZnF domain preventing its binding to the CUG repeat expansion. However, targeting of the ZnF domain caused dysregulation of alternative splicing, so this small molecule is not suitable as a therapy for DM. Compound 2 is a substituted naphtyridine that binds CUG repeats and displaces MBNL1, reducing foci and reversing some of the misspliced events in DM1 (Childs-Disney et al., 2013).

In addition to chemical ligands, antisense oligonucleotides and small molecules, some antibiotics have proven to be anti-DM1 agents. Lomofungin, and its dimer form, dilomofungin, two antifungal antibiotics, and erythromycin, an antibiotic to treat chest infections, are able to inhibit the binding of MBNL1 to CUG repeats. These antibiotics have high affinity for the CUG repeats, which releases MBNL1 from foci and restores the alternative splicing. Lomofungin and dilomofungin were able to partially rescue the missplicing of *Serca1*, whereas erythromycin was able to rescue the missplicing of *Clcn1*, *Atp2a1*, *Bin1*, *Nfix*, *Ryr1*, *Cacna1s*, *Ldb3* and *Camk2b* in the HSA^{LR} mouse model (Hoskins et al., 2014; Nakamori et al., 2016). Antibiotics can also serve as a starting point to

develop a therapy for myotonic dystrophy. Here have been described the most relevant compounds that inhibit the formation of the MBNL1-CUG_{exp} complex, but many more molecules have been developed and tested, and some of them have proven to be effective inhibitors of these complexes.

1.7.4 Altering levels of RNA binding proteins

The aberrant alternative splicing of the genes misspliced in myotonic dystrophy is caused by the reduction of the levels of MBNL1 and the increase of CELF1. To restore the levels of these antagonist proteins could have beneficial effects for the disease. Originally, the approach used for this purpose was to overexpress MBNL1. Overexpression of a skeletal muscle isoform of MBNL1 in the HSA^{LR} mouse led to the rescue of the missplicing of a subset of genes, including the *Clcn1* gene and a reduction of the myotonia (Kanadia et al., 2006).

Overexpression of MBNL1 can also be achieved using small molecules. A plant-derived alkaloide, Harmine, was found to increase the levels of *MBNL1* transcript. This caused a reduction of the nuclear foci and a partial rescue of the missplicing of *INSR* and *TNNT2* in human DM1 myoblasts and ameliorated the missplicing of *Clcn1* in the HSA^{LR} mouse model (Herrendorff et al., 2016). Non-steroidal anti-inflammatory drugs (NSAIDs) can also increase the levels of MBNL1. Phenylbutazone promoted by 1.3-fold the expression of *Mbnl1* mRNA in myogenic C2C12 cells in a dose-dependent manner and also in the HSA^{LR} mice. Increased expression of MBNL1 corrected the misplicing of *CLCN1* and increased the levels of CLCN1. Phenylbutazone increases expression of MBNL1 by suppressing methylation of CpG dinucleotides in the intron 1 of *Mbnl1* (Chen et al., 2016). MBNL1 protein levels were also increased in DM1 fibroblasts by two histone deacetylase inhibitors (HDAC), ISOX and Vorinostat. Nonetheless their mechanism of action was suggested to be epigenetic and on *MBNL1*, as their effects were not DM-specific (Zhang et al., 2017).

On the other hand, downregulation of CELF1 has been attempted using a PKC inhibitor, Ro-31-8220, with the aim to prevent phosphorylation of CELF1 by PKC and thus overexpression of the protein. Treatment of the EpA960 mice

with Ro-31-8220 improved mortality, cardiac conduction defects and rescued some of the genes misspliced in DM which are regulated by CELF1 (Wang et al., 2009).

1.7.5 Targeting the mispliced genes

Another level of disease pathogenesis that can be targeted are the misspliced genes, in order to prevent the inclusion or exclusion of the exons/introns that produce the spliceopathy. Only the aberrant splicing of a few genes has been linked to DM1 symptoms, but reversal of the missplicing of these genes could have potential to modify the disease. Very few papers have focused on this approach because modification of the alternative splicing can be achieved by other means, as discussed previously, but one of the more significant studies is the one from Wheeler *et al.*. They used a morpholino to target the *CLCN1* and induce suppression of exon 7a during the splicing process. Delivery of the morpholino in two DM mouse models rescued the missplicing of *Clcn1*, improving the chloride channel function and ameliorating myotonia (Wheeler et al., 2007). This study provides proof that the misspliced events can constitute another target to which direct potential therapies for myotonic dystrophy.

1.7.6 Additional targets

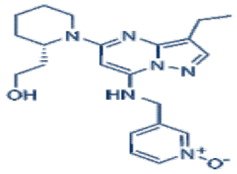
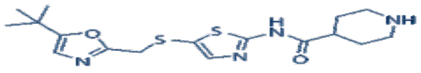
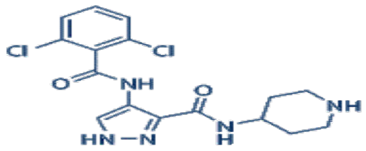
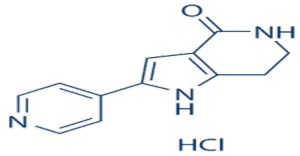
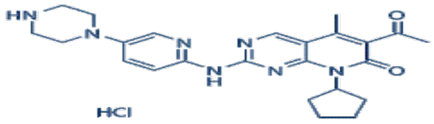
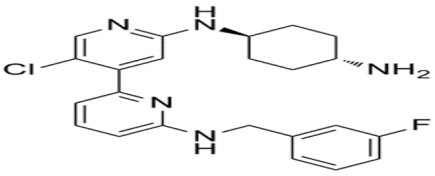
Different proteins are involved in many cellular processes, and their modulation can be beneficial for a certain disease. A relevant protein for this work is the kinase CDK12. In a recent study from Ketley *et al.* (unpublished), CDK12 was identified as the main target of a subset of kinase inhibitors that successfully reduce foci in DM1 cells. Inhibition of CDK12 with shRNAs also eliminated foci from the nucleus of DM1 cells. Hence, CDK12 appears to have a role in foci generation/maintenance and could potentially provide a novel target for therapeutic development.

1.7.7 Kinase Inhibitors

Protein phosphorylation plays a key role in several important cellular processes such as cell division, metabolism, survival and apoptosis. Any disruption to phosphorylation can cause disease or even be lethal (Cohen, 2002). The phosphorylation events are catalysed by enzymes called protein kinases, which transfer phosphoryl groups onto target proteins to direct their activity, localization and overall function. They are key regulators of cell function and particularly prominent in signal transduction and cell cycle regulation and co-ordination. To date, 518 human protein kinases have been characterized and they can be classified in different groups depending on the residue they phosphorylate (Manning et al., 2002). Most protein kinases add the phosphate group to a serine, threonine or tyrosine amino acid on the target protein, although other residues such as histidine can be phosphorylated.

The implication of the protein kinases in so many cellular processes make them a good target for drug development. The first protein kinase inhibitors were developed in the 1980s by Hiroyishi Hidaka and since then many others have been developed and optimized (Cohen, 2002). Kinase inhibitors are a type of enzyme inhibitor that block the action of the protein kinases and prevent the phosphorylation of the target proteins. They can be used to treat diseases such as cancer and inflammation or to modulate cell functions. Table 1.3 shows the chemical structure of the 6 kinase inhibitors used in this study, and their selectivity towards different protein kinases. It is worth mentioning that CDK12 is not included in Table 1.3 as a target of the kinase inhibitors as it has been recently identified as the main target for some of them.

Table 1.3: **Summary table of the six kinase inhibitors used in this study.** The name, chemical structure, proteins that inhibit and the *in vivo* function are listed in the table.

Name	Chemical structure	Target CDKs	Function *in vivo*
Dinaciclib (SCH727965)		CDK1 CDK2 CDK5 CDK9 CDK12	Tumor inhibition
SNS-032 (BMS-387032)		CDK2 CDK7 CDK9	Prevents tumor cell-induced VEGF secretion in a tumor coculture model
AT7519		CDK1 CDK2 CDK4 CDK6 CDK9	Inhibits tumor growth
PHA-767491		CDK9	Inhibits tumor growth
Palbociclib (PD-0332991)		CDK4 CDK6	Antitumor activity
CDK9 inhibitor II		CDK4 CDK9	Antitumor and antimetastatic agent

1.8 CDK12 in Myotonic Dystrophy

The Cyclin-dependent kinase (CDK) family comprises 21 different protein kinases involved in different cellular processes, such as regulation of the cell cycle (CDK1, CDK2, CDK4 and CDK6), regulation of transcription, post-transcriptional processing (CDK7 and CDK9), neuronal development, etc. These proteins bind to specific cyclin subunits to form a functional cyclin/CDK complex to perform their functions (Malumbres et al., 2009).

CDK12 is a serine/threonine kinase of 1490 amino acids and a molecular weight of 164 KDa encoded in chromosome 17 (17q21). It contains argi-

nine/serine (RS) rich domains in the N- and C-terminal regions, and a conserved central C-terminal domain (CTD) kinase domain (Figure 1.6) (Ko et al., 2001). CDK12 associates with cyclin K and has a role in transcription elongation and RNA processing, in maintaining genomic stability and in tumorogenesis (Blazek et al., 2011).



Figure 1.6: **Schematic of the structure of the human CDK12 protein.** CK12 contains large N- and C-terminal regions, with many Ser/Thr kinase domains (R/S), and a central kinase domain between the amino acids 737 and 1020. Taken from (Bartkowiak and Greenleaf, 2015).

RNA polymerase II (RNAPII) is required for the transcription of protein coding genes and many non-coding RNA genes. Phosphorylation of the C-terminal domain (CTD) by CDKs is important for transcription initiation and regulation. For many years CDK9 was thought to be the only CDK that phosphorylates the RNAPII, however it was recently discovered that CycK/CDK12 also phosphorylates RNAPII CTD at Serine 2 of the YSPTSPS heptad repeat, *in vitro* and *in vivo* (Ko et al., 2001). Nonetheless, unlike CDK9, CDK12 is not required for transcription initiation, instead it is important during the elongation phase of transcription (Bartkowiak et al., 2010).

CDK12 colocalizes to SC35 nuclear speckles, which act as a store for many proteins involved in pre-mRNA processing, thus CDK12 is thought to have a role in RNA processing (Ko et al., 2001) and in alternative splicing regulation (Chen et al., 2006). The complex CycK/CDK12 also maintains genomic stability by regulating the expression of a subset of DNA damage response genes (BRCA1, ATR, FANCJ and FANC2), and cells without CycK/CDK12 are sensitive to DNA damage agents (Blazek et al., 2011). Some of the genes regulated by CDK12 are cancer-related genes, such as BRCA1, so dysregulation of CDK12 has been observed in several cancers. Its implication has been well characterized in ovarian cancer, in the development of breast cancer and in gastric cancer, amongst others (Kohoutek and Blazek, 2012; Chila et al., 2016).

CDK12 is the main target of a subset of kinase inhibitors that were found to reduce foci in DM1 cells by others in the laboratory (unpublished). Nevertheless, these kinase inhibitors are not 100% selective against CDK12. Thus, in this work CDK12 was selected as a potential target to develop a therapy for myotonic dystrophy, and in order to study its role in the disease, DM1 cells were engineered using the CRISPR/Cas9 system to knock out *CDK12*.

1.9 Genome Engineering using CRISPR/Cas9 System

Since the discovery of the DNA double helix, technologies to manipulate DNA have enabled many of the discoveries in biology in the past years. Manipulating the genome for functional studies, cancer modelling or therapeutic purposes require efficient systems able to make site-specific modifications in the DNA. Among the technologies used to engineer the genome, nucleases of bacterial origin represent a powerful tool and have been widely used in the past few years. These include zinc-finger nucleases (ZFNs), transcription activator-like nucleases (TALENs) and the CRISPR/Cas9 system. I will focus on the CRISPR/Cas9 system as it was used in this work to engineer DM1 and DM2 fibroblasts.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) are a series of short direct repeats interspaced with short sequences first reported by Japanese researchers in 1987 in the genome of *Escherichia coli* (Ishino et al., 1987). In 2000, further research showed these sequences to be present in all Archea and Bacteria, in some cyanobacterias and in proteobacterias. The CRISPR cluster was described as short palindromic sequences of 24-40 bp interspaced with repeat sequences of up to 11 bp (Mojica et al., 2000). The repeat sequences are very conserved, suggesting a common function in the different organisms. The spacers can derive from bacteriophages, conjugative plasmids and chromosomal sequences (Mojica et al., 2005). An additional feature of the spacer sequences was discovered by Bolotin *et al.*. They realised that all spacers have a similar sequence at the end, which is required for target recognition.

This sequence was named Protospacer Adjacent Motif (PAM). Furthermore, the CRISPR locus is located near genes that encode proteins with nuclease and helicase domains, the *Cas* proteins (Cas-associated) (Jansen et al., 2002; Bolotin et al., 2005; Haft et al., 2005). The CRISPR loci was found to be transcribed, and the spacer sequences are transcribed into short RNAs, named crRNAs (CRISPR RNAs), which guide the Cas proteins to the target DNA for its cleavage (Tang et al., 2002; Brouns et al., 2008). The function of the crRNAs and the cas proteins became more clear after finding that the CRISPR system acts as an adaptive immune system in bacteria, using the small RNAs transcribed from the spacer sequences as memory signature of past invasions (Makarova et al., 2006; Barrangou et al., 2007).

The adaptive immunity begins after a virus or bacteriophage infects the bacteria, and then the bacteria integrates new spacers derived from the pathogen's genome. The adaptive immunity occurs in three stages (Figure 1.7): (1) the acquisition stage, in which a short sequence of the invading agent is added as a spacer in the CRISPR locus, (2) the expression stage, in which the spacers are transcribed into precursors crRNA (pre-cRNAs) that then undergo maturation to generate crRNAs, and (3) the interference stage, when the crRNA directs the Cas proteins to the target site in the invading agent's genome for its cleavage.

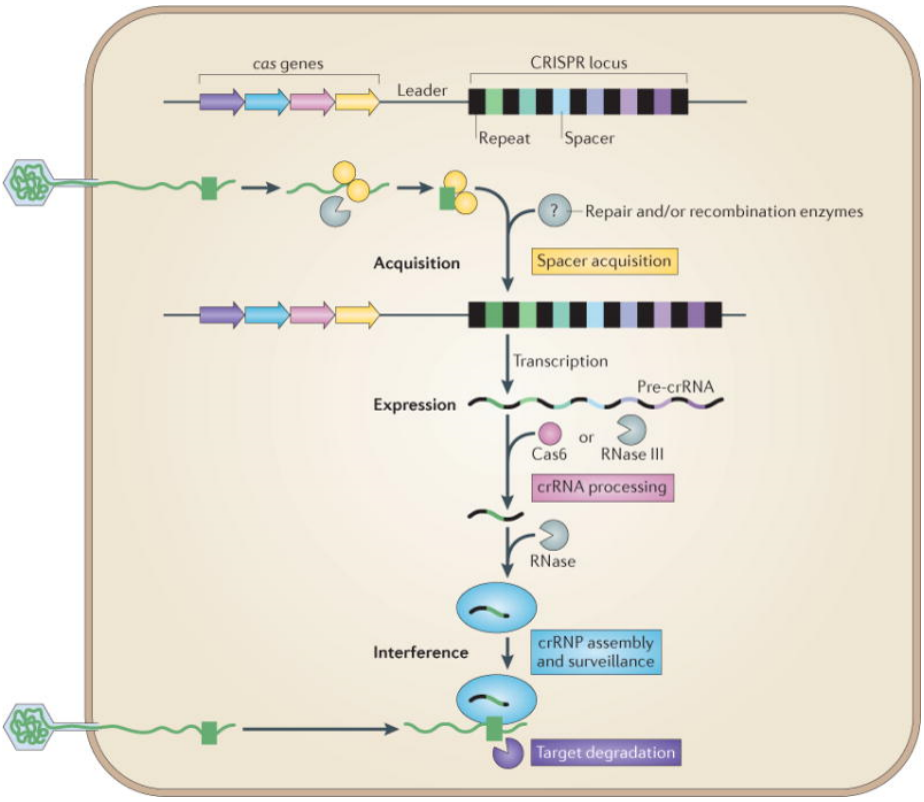


Figure 1.7: **Overview of the CRISPR adaptive immunity.** The CRISPR locus, in close proximity to the cas genes (coloured arrows), contains short spacer sequences interspaced by short repeats. After the infection of the bacteria by a pathogen, a short sequence from the pathogen's DNA is incorporated into the CRISPR locus as a new spacer. The CRISPR locus is then transcribed and Cas6 and the RNase III are involved in the maturation of short single crRNA, each of them containing a spacer sequence and a repeat sequence. When the same pathogen infects again the bacteria, the crRNA associates with Cas proteins to produce the cleavage of the DNA of the foreign invader. Taken from (van der Oost et al., 2014).

To integrate new spacers in the CRISPR locus, the CRISPR system first detects suitable PAM sequences, and then processes the adjacent DNA into spacer precursors that are then integrated in a PAM-dependent orientation in the bacteria's CRISPR locus (Swarts et al., 2012). Expression of both Cas1 and Cas2 is required for the spacer acquisition, however their precise functions remain elusive (Yosef et al., 2012). Transcription of the CRISPR locus produces long precursors known as pre-crRNAs. Processing of the pre-crRNAs and maturation into single crRNAs involves either the Cas6 or the RNase III. There are 3 types of CRISPR system and a few subtypes (Figure 1.8), type I (type I-A and type I-E), type II and type III (type III-A and type III-B), and each of them has different

requirements to work. In the types I and III, Cas6 is responsible for the processing of the pre-crRNAs. Cas6 enzymes cleave the pre-crRNA by hydrolysing a single phosphodiester bond in the repeat sequences of the transcript, usually resulting in a short crRNA that contains a 5' repeat of 8 nucleotides, the complete spacer sequence and a 3' repeat of variable length that forms a hairpin structure (Reeks et al., 2013). In the type II systems, processing of the pre-crRNA is mediated by the RNase III. Furthermore, the CRISPR locus includes a gene that encodes a transactivating crRNA (tracrRNA), which is a 25 nucleotide sequence complementary to the repeat region of the pre-crRNA. Watson-Crick base pairing between the pre-crRNA and the tracrRNA results in a double-strand region that is recognized and cleaved by the RNase III. Cas9 is also required for the processing of the transcript, and it remains associated with the crRNA-tracrRNA hybrid. The mature crRNA contains a 20 nucleotide spacer sequence and is required for target interference (Deltcheva et al., 2011; Jinek et al., 2012). Finally, mature crRNA associated with Cas proteins form stable crRNP complexes.

Each system requires a different Cas protein or set of Cas proteins. The targeting of the invading DNA requires the scanning of the DNA looking for a protospacer sequence complementary to the crRNA. Binding of the crRNA to its target sequence produces a R-loop structure which triggers a conformational change in the crRNP (Westra et al., 2012; Sashital et al., 2012). Ultimately, the Cas protein cleaves the DNA. In the type II system, the Cas9 protein catalyses the cleavage of the foreign DNA. It has two domains, the HNH-like domain, which cleaves the DNA strand that pairs with the crRNA, and the RuvC-like domain that cleaves the displaced DNA strand. This creates a double-strand break typically 3 bases upstream the 3' end of the protospacer, causing the death of the invading pathogen (Garneau et al., 2010; Jinek et al., 2014). The domains of the Cas9 protein can be mutated so a Cas9 with different functions can be generated (Sapranauskas et al., 2011). Mutations in the RuvC catalytic domain produce a Cas9 nickase mutant that nicks the DNA and produces single-strand breaks (Jinek et al., 2012). Inactivated Cas9 can be also generated by inducing other mutations in its sequence.

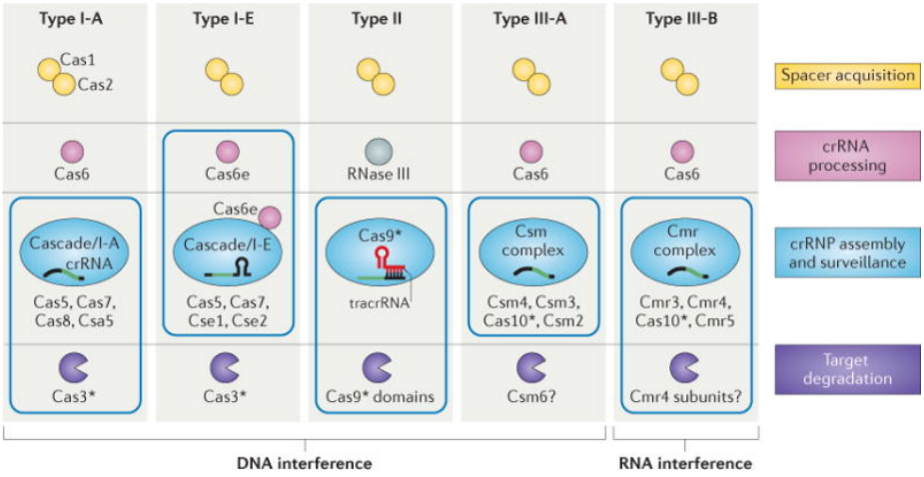


Figure 1.8: **Types of CRISPR systems.** There are different Cas proteins. Each type of CRISPR system relies on a Cas protein or a set of Cas proteins. The types I and III require multiple Cas proteins for the cleavage of the target DNA, whereas the type II requires only a single protein, the Cas9. Boxes indicate components of the crRNP complexes. All types of CRISPR system target DNA except for the type III-B which targets RNA. Taken from (van der Oost et al., 2014).

The type I and type III CRISPR systems require a large complex of Cas proteins to cleave the target DNA (Brouns et al., 2008). Nonetheless, the type II system requires only a single Cas protein, the Cas9, to recognise the target site and cleave it (Jinek et al., 2012). Because of its simplicity compared to the other types of CRISPR systems, the type II has been the preferred one for genome engineering applications. A few years ago the dual tracrRNA:crRNA was engineered as a single guide RNA (sgRNA). It retains the 20-nucleotide sequence at the 5' end which determines the target DNA, and the double-stranded structure at the 3' end that binds to the Cas9. This was the first step to develop a simple two-component system that relies on a 20-nucleotide guide RNA (gRNA) and the Cas9 nuclease to target any DNA sequence as long as it is adjacent to a PAM sequence (Jinek et al., 2012). Shortly after the production of the gRNA, three independent studies reported the application of the *Streptococcus pyogenes* type II CRISPR system to eukaryotic cells. Human codon-optimization of the *S. pyogenes* Cas9 and attachment of nuclear localization signals (NLS) ensured compartmentalization in mammalian cells. Efficient modification of all the cell models was achieved with the gRNA and the Cas9 (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013). These studies opened the door to a world of almost

endless applications of the CRISPR system, due to the possibility to modify virtually any sequence in the genome of an organism.

The modification of the genome is produced after the Cas9 cleaves the target DNA. The double-strand breaks, or single-strand breaks produced by a mutated version of Cas9, can be repaired either by non-homologous end joining (NHEJ) or by homology-directed repair (HDR) (Figure 1.9) (Symington and Gautier, 2011; Chapman et al., 2012). The NHEJ occurs in the absence of a repair template, and the DSBs are simply re-ligated, which produces scars in the form of insertion/deletion mutations. The NHEJ repair can mediate the generation of knockouts, as indels in coding exons can produce frameshift mutations and premature stop codons (Perez et al., 2008). The NHEJ pathway is thought to occur throughout the cell cycle, but it predominantly occurs during the G1 phase (Mjelle et al., 2015). On the contrary, the HDR repair occurs in the presence of a repair template which can be in the form of conventional double-stranded DNA constructs with homology arms flanking the insertion sequence, or single-stranded DNA oligonucleotides. The use of a repair template allows the precise modification of the genome. HDR is usually active only on dividing cells, and is most active during the S and G2 phases of the cell cycle, although its efficiency depends on the cell type. Furthermore, frequency of the HDR is lower compared to NHEJ (van Gent et al., 2001; Saleh-Gohari and Helleday, 2004; Chen et al., 2011).

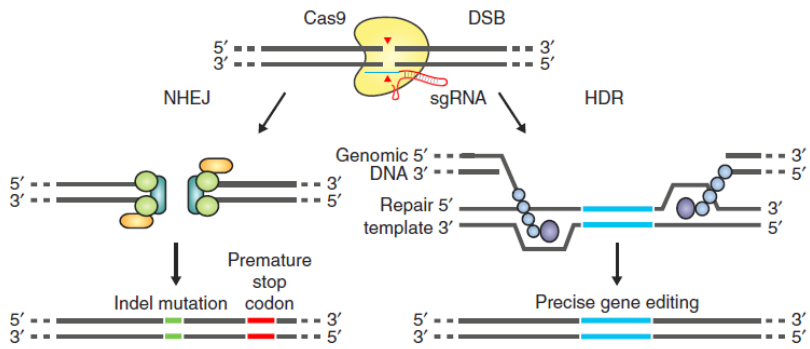


Figure 1.9: **DNA repair pathways after the introduction of DSBs.** In the NHEJ pathway the ends of the DNA are re-ligated, which can result in the introduction of mutations. Indel mutations in the coding regions of a gene can result in the gene inactivation. In the HDR pathway, the homologous recombination of a repair template in the DSBs allows for precise modification of the genome. Taken from (Ran et al., 2013).

In human cells, efficient knock in of foreign DNA has been long desired. So far, most studies have focused on HDR-based strategies. Nonetheless, HDR is inefficient in human cells and targeted integration has always been reported to be low, whereas NHEJ-mediated repair is prevalent (Mao et al., 2008). The addition of overhangs to the repair template or the linearisation of a donor DNA plasmid promotes targeted integration mediated by NHEJ. A few studies have reported efficient NHEJ-mediated knock in in cultured cells (Orlando et al., 2010; Cristea et al., 2013), in zebrafish (Auer et al., 2014), and recently a 4.6Kb fragment was integrated by NHEJ into the *GAPDH* locus of LO_2 cells, yielding up to 20% GFP-expressing cells, and in human embryonic stem cells (ESCs), yielding up to 1.7% GFP+ cells (He et al., 2016). A combination of HDR and NHEJ repair pathways has been adopted as a novel knock-in strategy, the microhomology-mediated end-joining (MMEJ). MMEJ is a DSBs repair mechanism that uses microhomologous arms (5-25 bp) for error-prone end-joining (McVey and Lee, 2008). MMEJ is active during the G1/early S phases of the cell cycle (Taleei and Nikjoo, 2013). A MMEJ-based knock-in strategy, the PITCh system (Precise Integration into Target Chromosome) enables efficient integration of large DNA fragments into a wide range of cells and organisms. The PITCh system requires a donor repair template with a specific design. The fragment to be integrated needs to contain microhomologous arms (5-25 bp), flanked by the sequence of

the gRNA including the PAM sequence. Co-expression of this donor template with the gRNA and the Cas9 will produce the cleavage of both the genomic DNA and the donor template at both ends of the exogenous fragment to be integrated. Linearisation of the donor template by the Cas9 and the presence of the micro-homologous arms allow for an efficient integration of the foreign DNA into the target site with reliable junctions. PITCh systems have been successfully applied to cells and zebrafish (Nakade et al., 2014; Hisano et al., 2015; Sakuma et al., 2016).

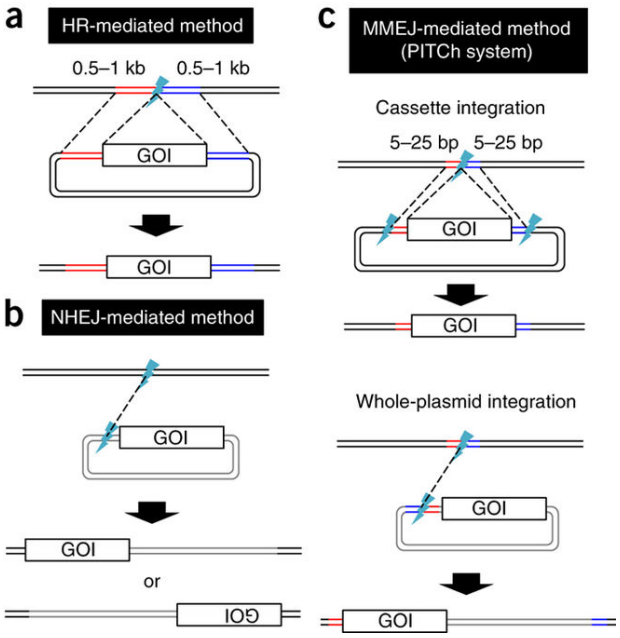


Figure 1.10: **Outcomes of HDR-, NHEJ- and MMEJ-mediated knock in. (a)** The HR (Homologous Recombination)-mediated knock in requires long homology arms and has a very low efficiency. **(b)** The NHEJ-mediated knock in does not require homology arms and the efficiency is higher than HR. Nonetheless, the exogenous DNA can be randomly integrated and indels can be created. **(c)** The MMEJ-mediated knock in requires short homology arms that enable a precise integration in the target site. A short cassette or a whole linearised plasmid can be integrated into the target site. GOI: gene of interest. Taken from (Sakuma et al., 2016).

The CRISPR system is a promising technology in the field of genome engineering and synthetic biology. Efficient genome modification has been achieved in many different cell types and organisms, with a wide range of different applications in biotechnology, agriculture or biomedicine (Figure 1.11). The availability of different types of CRISPR systems allows for the exploration of

still unknown functions and applications of the system, and every month lots of new papers about CRISPR are published.

Biology		Biotechnology		Biomedicine
Model organisms	**Cell lines**	**Crop plants**	**Fungi**	Organoids
Mice	HEK293	Rice	Kluyveromyces	hESCs
Rat	U2OS	Wheat	Chlamydomonas	iPSCs
Fruit fly	K562	Sorghum		
Nematode		Tobacco		
Arabidopsis				
Salamander				
Frog				
Monkey				

Figure 1.11: **Examples of cell types and organisms that have been successfully engineered using the CRISPR system.**

In the work presented here the CRISPR system was used to engineer the DNA of DM1 and DM2 fibroblasts. In DM1 cells the aim was to knock out the *CDK12* gene. When this experiment was started, knock in of an exogenous DNA fragment was being achieved mostly by HDR, and not much had been reported about NHEJ-mediated knock in. So, originally a donor repair template was constructed, in order to integrate an antibiotic resistance gene into *CDK12* by HDR, to disrupt it and be able to select the engineered cells. In DM2 cells the aim was to knock in by HDR as well, an inducible promoter into intron 1 of the *CNBP* gene, upstream of the CCTG repeats that cause DM2, to control the expression of the gene. In this case, a donor repair template was constructed as well, containing the inducible promoter flanked by 1 Kb homology arms. Both repair templates were expected to integrate the exogenous DNA in the target site by homologous recombination. Nonetheless, outcome of the experiment with DM2 cells and very recent publications about NHEJ-mediated knock in made me redesign both experiments in order to get a NHEJ-mediated knock in integration in both genes and cell lines. Outcome of the NHEJ-mediated knock in in DM2 cells and newly reports about MMEJ-based strategies made me re-design again the CRISPR experiment for DM2 cells and a MMEJ-based approach to knock in the inducible promoter was adopted.

1.10 Inducible Gene Expression Systems

Genome modification has proved to be a very valuable tool for organismal biology, but sometimes controlling gene expression is also very useful to study gene and protein function *in vivo*. Engineering the genome to create irreversible changes in essential genes can be lethal. To overcome this problem new techniques for the temporal control of gene expression have been developed. The Cre/loxP recombination system and the Tet-controlled transcription activation system are the most widely applied and characterized (Schonig et al., 2012).

In the Cre/loxP system the gene of interest is flanked by loxP sequences of 34-bp and when the Cre recombinase from bacteriophage P1 is expressed, it recognises the loxP sites leading to an inversion or deletion of the loxP-flanked DNA sequence. This system can be used to delete genes, to activate transcription by removing transcriptional termination sequences or even to integrate exogenous DNA in the genome (Sauer and Henderson, 1988, 1990).

Tetracycline-inducible systems can modulate gene transcription in a tissue- and time-dependent manner. They exist both as tet-ON or tet-OFF systems in which the administration of tetracycline or an analogue such as doxycycline, either allows or prevents gene expression, respectively (Gossen and Bujard, 1992; Gossen et al., 1993). In this work the Tet-ON system was used to control the expression of the mutant transcript in DM2 cells.

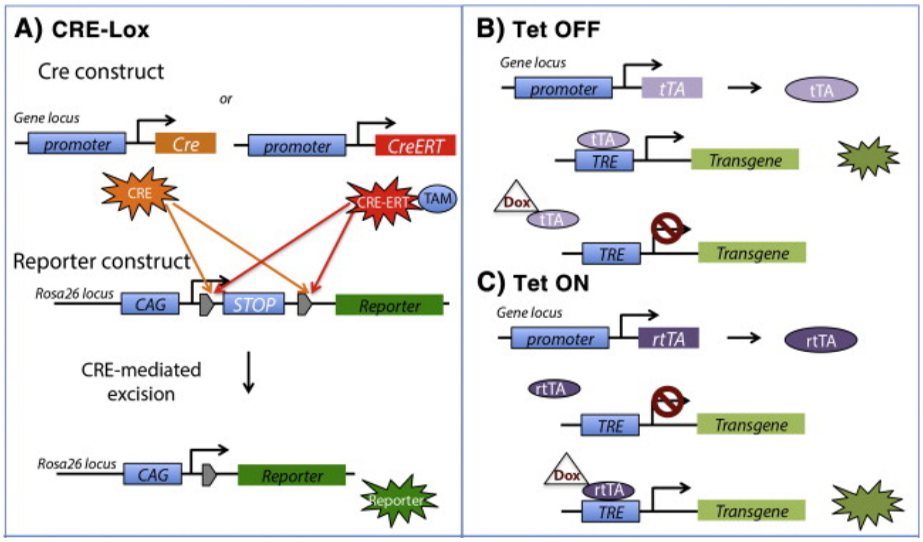


Figure 1.12: **Schematic representation of inducible gene expression systems. A.** CRE-Lox system, which consists of two elements: a construct with the gene CRE or a modified tamoxifen-dependent form (CreERT), and a reporter construct, with a "stop" cassette that inhibits expression of the reporter. In the presence of CRE the "stop" cassette is excised and the marker protein is expressed. **B.** Tet-Off system. tTA (tetracycline-sensitive transcriptional activator) activates the expression of the transgene downstream the TRE (Tetracycline Response cis-regulatory Element). Addition of doxycycline inhibits the expression of the transgene. **C.** Tet-On system. The expression of the transgene is produced only after the addition of doxycycline, which allows the binding of the rtTA to the TRE. Taken from (Chong et al., 2014).

pTet-One System

The Tet-One system is an inducible gene expression system for mammalian cells that contains all the necessary components in a single plasmid and allows gene expression after the addition of tetracycline or an analogue. The pTetOne vector is a 4207 bp plasmid containing two main elements: the **Tet-On 3G Transactivator Protein**, which is expressed constitutively and is a modified form of the Tet-One Advanced transactivator protein which was evolved to display higher sensitivity to tetracyclin or doxycycline (Zhou et al., 2006). In the presence of doxycycline the Tet-On 3G protein binds to the second element of the system, the **pTRE3GS Inducible Promoter**, and activates transcription of the downstream gene. The pTRE3GS inducible promoter consists of 7 repeats of a 19 bp *tet* operator sequence located upstream of a minimal CMV promoter, and it lacks

binding sites for endogenous mammalian transcription factors, so is silent in the absence of induction (Loew et al., 2010).

Conventionally the gene of interest (GOI) is cloned into the multiple cloning site (MCS) of the plasmid, downstream of the inducible promoter. In the work described here no GOI was cloned, instead the aim was to integrate a fragment of about 2.1Kb containing the Tet-On 3G protein and the inducible promoter from the pTetOne vector into intron 1 of the *CNBP* gene in DM2 cells, upstream of the CCTG repeats, to control its expression. Figure 1.13 shows the mode of action of the Tet-One systems.

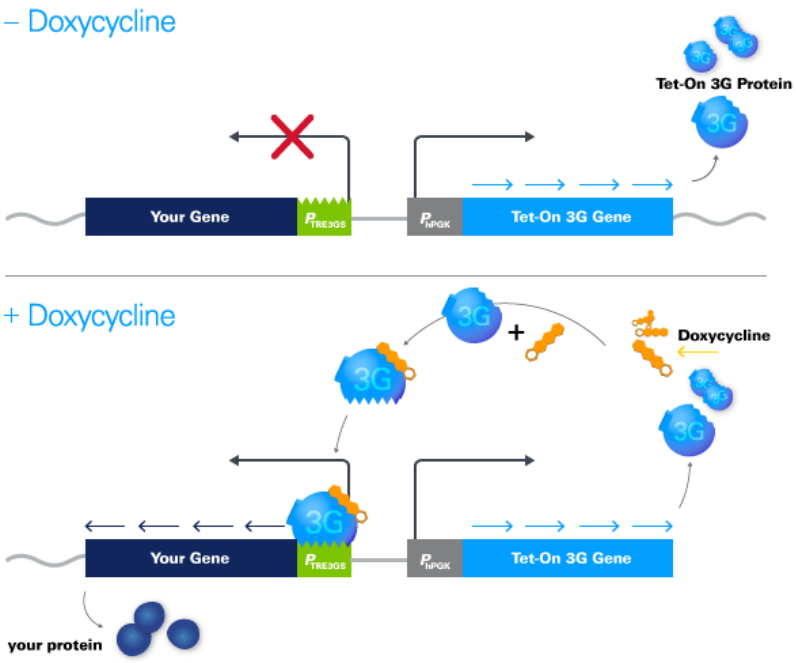


Figure 1.13: **Tet-One design**. The Tet-On 3G transactivator protein is expressed constitutively from the hPGK promoter, but it does not bind to the TRE3GS promoter in the absence of doxycycline (top panel). In the presence of doxycycline, the Tet-On 3G protein undergoes a conformational change, binds to the TRE3GS promoter and activates transcription of the gene downstream (bottom panel).

1.11 Aims of Study

My initial and main hypothesis was that CDK12 is involved in the production of the mutant transcript in both DM1 and DM2, probably by controlling transcription of the mutant allele. A secondary hypothesis, which was that the CRISPR/Cas9 system can be used to create an inducible cell model for DM2, led me to conduct further experiments. The main objectives established for this thesis are listed next, and they were oriented to test the hypotheses mentioned.

- To study the relationship of expansion nuclear foci with other components of the cell, such as CDK12 protein or SC35 speckles.
- To test a series of kinase inhibitors in DM2 cells that have already proven to be effective for foci reduction in DM1 cells, and to study additional effects of the compounds in both DM1 and DM2 cells.
- To unravel the mechanism of action of the kinase inhibitors tested by which they eliminate foci in DM1 and DM2 cells.
- To generate a CDK12 knockdown DM1 cell line using the CRISPR/Cas9 system in order to study the role of CDK12 in myotonic dystrophy and in foci maintenance.
- To generate a DM2 cell line with stable integration of an inducible promoter to control the expression of *CNBP* using the CRISPR/Cas9 system. The cell lines generated could be useful to understand the foci formation process and to study the molecular events downstream of mutant transcript expression.

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.

2.1.1 Cell lines

Cell lines used in this study include human fibroblasts from a non-DM (SB Telo MyoD) person, a DM1 (KB Telo - 400 CTG repeats) patient and a DM2 (KAGO Telo - 3000-4000 CCTG repeats) patient. All the cell lines were previously telomerized by Denis Furling to allow indefinite growth of the cells (Chaouch et al., 2009).

2.1.2 Media, solutions and antibiotics

The media 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; Penicillin/Streptomycin from Gibco (10,000 units of Penicillin G and 10,000 μg/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; and Hygromycin B (50

mg/ml) from Thermo Fisher Scientific.

2.1.3 Cell culture

All cells were grown and maintained in a humidified incubator at 37°C with 5% CO_2. The media used to grow them was Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) and 2.5% of the antibiotics penicillin (10000 units of Penicillin G)/streptomycin (10000 μg/ml).

2.1.4 Cell passaging

Cells were visualized regularly under the microscope and sub-cultured after reaching 80 to 100% confluency. For passaging, media was removed from the T75 flask using serological pipettes and cells were briefly washed with 6 ml of pre-warmed 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°C 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 was 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 20 ml.

2.1.5 Cell count

Cell number was measured using a haemocytometer, which contains 2 sets of 16 squares, each of which is divided into 16 smaller squares. 300 μl of cell suspension was placed into the haemocytometer. Number of cells within three 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 5000, as the depth of the haemocytometer was 0.2 mm.

2.1.6 Cryopreservation of cells

DMEM media with 10% 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°C for about 5 minutes and complete dissociation of the cell monolayer was confirmed by visualization 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 2500 g for 10 min. The cell pellet was resuspended in 1 ml of media containing 10% DMSO and the solution containing the cells was transferred to a cryotube and stored at -80°C for 24 hours. The next day the cells were moved into liquid nitrogen for long term storage.

2.1.7 Defrosting cells from liquid nitrogen

Cryotubes containing the frozen cells were removed from liquid nitrogen and thawed by immediate incubation in a water bath at 37°C. 1 ml of fresh media was added and cells were transferred to a 15 ml 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°C and 5% CO_2.

2.1.8 Compound treatment of DM cells

KAGO Telo cells (DM2) were seeded onto 96-well plates at a confluency of 8000 cells/well and grown for 24 hours in the incubator at 37°C. Each plate was used to screen in triplicate the kinase inhibitors. Six kinase inhibitors were tested: Dinaciclib, SNS-032, AT7519, PHA-767491, PD0332991 and CDK9 inhibitor II. Cells treated with DMSO served as a negative control. Each compound was added to the cells in a 12-point dilution, at a range of concentration from 500 nM to 242.5 pM for Dinaciclib, from 1 μM to 485 pM for SNS-032 and from 20 μM

to 9.75 nM for the rest of inhibitors. The media containing the kinase inhibitors or the DMSO was removed 24 or 48 hours later, and the protocol for *in situ* hybridization or ICC was started. The same treatment protocol was followed to treat DM1 cells when necessary.

Four additional compounds, ADS160181, ADS160919, ADS158784 and ADS158081, were used to treat DM2 cells at a range of concentration from 1.25 μM to 0.6 nM for ADS160919 and from 40 μM to 20 nM for the other three compounds. After a 48-hours treatment, the compounds were removed and the protocol for *in situ* hybridization was started.

2.1.9 MTT assay

The MTT assay was used to assess the toxicity of the kinase inhibitors on the cells. The MTT solution was prepared by dissolving 0.1 g of MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) in 20 mL of RPMI-1640 medium and added to the DM2 cells after each compound treatment, in a volume of 10% of the total volume in each well of the 96-well plate. After incubation at 37°C for 4 hours, formazan crystals were solubilized with 0.1 N isopropanol. Toxicity of compounds was assessed after analyzing the plate in a spectrophotometer and the mean number of living cells in treated wells was compared against the mean number of living cells in untreated wells.

2.1.10 RNA Fluorescence *In Situ* Hybridization

RNA FISH was performed in all cell lines in order to visualize in the nucleus of the cells bright foci which contain mutant RNA transcripts with expanded CUG or CCUG repeats. Media was removed from the cells and they were washed thrice with PBS. The fixation step was performed by incubating the cells in 4% PFA for 30 minutes at room temperature. Cells were washed again thrice with PBS and then permeabilised in 80% ethanol for 15 minutes at room temperature. After another 3 washes with PBS cells were incubated in the pre-hybridization solution for 15 minutes at room temperature. The pre-hybridization solution

contained: 40% formamide, 10% 20x SSC and 50% DEPC water. The pre-hybridization solution was then replaced with the hybridization solution, which contained: 40% formamide, 10% 20x SSC, 50% DEPC water, 10 μl of VAC (RNase inhibitor, 200 mM, Sigma-Aldrich), 100 μl of ssDNA (sperm salmon DNA, 10 mg/ml, Sigma-Aldrich), 20 μl of BSA (20 mg/ml, NEB) and the Cy3-labelled or Cy5-labelled $(CAG)_{10}$ or/$(CAGG)_8$ probe. Cells were incubated overnight at 37°C in the incubator. The next day the cells were washed thrice with PBS, stained for 5 minutes at room temperature with 4',6-diamidino-2-phenylindole, dihydrochloride (DAPI, 10 mg/ml) and finally left in PBS. Cells were then ready to image using either the confocal microscope or the Widefield plate reader.

2.1.11 Immunocytochemistry (ICC)

Unless stated otherwise, the protocol for ICC was as described below. The cells were seeded onto 96-well plates (Widefield plate reader) or in 35 mm petri dishes (suitable for confocal microscopy) and allowed to grow overnight.

Fixation

Media was removed and cells were washed thrice with PBS (5 minutes washes) and fixed with ice cold Methanol:Acetone (50:50) for 30 minutes at room temperature. The cells were washed again with PBS before proceeding with the blocking step.

Blocking

The cells were blocked in 5% BSA + 5% sheep serum in PBS for 1 hour at room temperature. The blocking step decreases the non specific interactions and reduces background.

Antibody staining

The blocking solution was removed and the cells were incubated with 1% BSA in PBS with the antibody at the appropriate dilution (mouse MBNL1 1:2000, mouse CDK12 1:1000, mouse SC35 1:1000, rabbit CDK12 1:50, rabbit MBNL1 1:5000) overnight at 4°C. The next day the cells were washed thrice in PBS 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 555 goat anti-mouse and AlexaFluor 488 or 555 goat anti-rabbit. The cells were rinsed again in PBS and stained with DAPI for 5 minutes at room temperature, to stain the nuclei of the cells. Finally DAPI was replaced with PBS and the cells were imaged using either the Widefield plate reader or the confocal microscope.

For a double ICC, two antibodies raised in different species were added to the cells, followed by incubation with two appropriate secondary antibodies labelled with a different dye.

2.1.12 Co-localisation experiments (ICC + FISH)

To study the co-localisation of foci with CDK12 or SC35 speckles, cells were seeded on 35 mm petri dishes suitable for confocal microscopy and grown for 24 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-CDK12 (Abcam ab57311) or anti-SC35 (Abcam ab11826) antibody was added to the cells at a dilution 1:1000 and incubated overnight at 4°C, followed by staining with the secondary antibody AlexaFluor 647 goat anti-mouse (1:500) for 30 minutes. Cells were fixed again in 4% PFA for 5 minutes and incubated with the pre-hybridization solution (50% DEPC water, 40% formamide and 10% 20xSSC) for 15 minutes. Finally cells were incubated with a Cy3-labelled $(CAG)_{10}$ or/$(CAGG)_8$ probe overnight at 37°C and the next day stained with DAPI for 5 minutes and imaged using the Zeiss 710 confocal microscope.

Images were analysed using LSM image browser and ImageJ.

2.2 Microbiology and Cloning Techniques

Escherichia coli (E. coli) strain DH5α was used for all microbiological procedures, unless stated otherwise.

2.2.1 Media

Different types of media were used for microbiological work, including Luria Broth (LB), SOB and SOC. LB media consists of 5 g/l yeast extract, 10 g/l bacto-tryptone, 10 g/l NaCl and 2 mM NaOH (pH 7.0). The SOC media was prepared fresh every time before use, by mixing 20 ml of SOB media with 400 μl of 20% glucose and 100 μl of 1 M $MgCl_2$. The SOB media consists of 5 g/l bacto-yeast extract, 10 g/l bacto-tryptone, 0.548 g/l NaCl, 0.186 g/l KCl and 2 mM NaOH (pH 7.0).

2.2.2 Antibiotics

The antibiotics used for selection were Ampicillin and Kanamycin. Final concentration of both was 100 μg/ml, diluted from the stock solution at a concentration of 100 mg/ml.

2.2.3 Restriction digestion

Restriction digestion of DNA was performed using restriction enzymes and reagents from New England Biolabs (NEB) following manufacturers' instructions. Reactions were composed of 1 to 4 μg of DNA, 1x restriction buffer, 1-2 μl of 10 U/μl restriction enzyme and distilled water. The mixture was incubated in the water bath at 37°C for 3-4 hours, or overnight in some cases, followed by the inactivation of the enzyme when it was required. Enzymes were usually inactivated by heating the mixture at 65°C or 80°C for 20 minutes. Finally agarose

gel electrophoresis was used to separate the fragments depending on their size. Many different enzymes were used, and each of them requires a specific restriction buffer. The buffers used were:

- Buffer 3: consisting of 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM dithiothreitol (DTT).
- Buffer 4: consisting of 50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate and 1 mM DTT.
- CutSmart: consisting of 50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate and 100 μg/ml BSA

2.2.4 Vector-Insert Ligation

DNA ligation was carried out using the T4 DNA Ligase (NEB). 70 ng of plasmid were mixed with the insert at a ratio 1:3 or 1:5 in a 20 μl reaction containing 1x T4 DNA Ligase buffer (50 mM Tris-HCl, 10 mM MgCl2, 1 mM ATP, 10 mM DTT), distilled water and 400 U of T4 DNA Ligase. The mixture was kept at 16°C overnight. The next day the ligase was inactivated by heating the mixture at 65°C for 10 minutes and the ligated product was used to transform bacteria.

2.2.5 pGEM-T-easy Vector

The pGEM-T easy vector system (Promega) was used to clone PCR products following PCR amplification with Taq polymerase. The Taq polymerase leaves 3'-A overhangs on the PCR product, which are ligated with the 3'-T overhangs present in the pGEM-T easy vector. Successful cloning interrupts the *lacZ* gene, stopping the production of β-galactosidase. Recombinant clones can be then identified by white/blue screening on X-gal plates.
Briefly, inserts were mixed with 50 ng of the T-easy vector at a ratio 3:1 or 5:1, in a reaction containing 2x Rapid Ligation Buffer, 400 U of T4 DNA ligase and deionized water to a final volume of 10 μl. The reaction was incubated overnight at 4°C and the next day it was used to transform electrocompe-

tent cells (described below). 100 μl of each transformation culture were plated onto LB/ampicillin/IPTG/X-gal plates and incubated at 37°C overnight. White colonies were selected and miniprepped.

2.2.6 Transformation

For cloning purposes and plasmid propagation, bacteria were transformed by heat shock or electroporation, depending on the availability of electrocompetent or $CaCl_2$ competent cells stocks at the start of the experiment. Aliquots of both type of cells are routinely prepared by the technicians in the laboratory from DH5α glycerol stocks.

Heat shock

A 100 μl aliquot of $CaCl_2$ competent cells, which contains 1 x 10^6 cells, per transformation was thawed on ice and 1 - 2 μl (usually 10 pg - 100 ng) of plasmid DNA added. The mixture of cells with DNA was incubated on ice for 30 minutes, followed by a heat shock at 42°C for 1.5 minutes and immediately placed back on ice for 2 minutes. 900 μl of LB medium was added and the mixture was incubated for 1 hour in a shaking incubator at 37°C. Cells were then pelleted by centrifugation at high speed for 1 minute. The pellet was resuspended in 100 μl of LB medium and plated at a ratio of 30/70 onto LB agar plates containing antibiotic. Plates were incubated overnight at 37°C to allow the bacterial growth.

Bacterial electroporation

SOC media was prepared fresh before each electroporation round as described in the Media section. For each transformation, a 40 μl aliquot of electrocompetent cells, which contains 1 x 10^{10} cells, was thawed on ice and mixed with 1 or 2 μl (usually 10 pg - 100 ng) of plasmid DNA. This mix was transferred into a pre-chilled 2 mm electroporation cuvette (Flowgen Bioscience) and tapped to ensure the mix was at the bottom without bubbles. It was placed in a GenePulser (BioRad), set up as follows: voltage 2.5 kV and capacitance 25 μF. Following

transformation, 200 μl of SOC media was added to the cuvette, mixed with the cells and transferred into a 15 ml falcon tube containing 760 μl of SOC media. The falcon tube was incubated at 37°C in a shaking incubator for 1 hour. Cells were then centrifuged at low speed, 1000 g for 10 minutes, and the pellet resuspended in 100 μl of SOC media. Culture was plated at a ratio of 30/70 onto LB agar plates containing antibiotic and incubated overnight at 37°C to allow the bacterial growth.

2.2.7 Preparation of Glycerol Stocks

Glycerol stocks of bacterial cultures were prepared for long term storage. Cells were centrifuged after overnight culture in 5 ml of LB medium and then resuspended in 875 μl of LB medium and 125 μl of 80% glycerol. The mixture was then stored at -80°C.

2.2.8 Miniprep Plasmid DNA Isolation

Single colonies of bacteria from a LB agar plate were transferred into tubes containing 5 ml of LB medium and antibiotic, and incubated at least for 8 hours, but usually overnight at 37°C in a shaking incubator. 3 ml of bacterial culture were pelleted by centrifugation at high speed and plasmid DNA was isolated using the Plasmid Miniprep kit (SIGMA) according to the manufacturers' protocol. The purified plasmid DNA was eluted in 80 to 100 μl of ultra-pure water and stored at -20°C.

2.2.9 Maxiprep and Endotoxin-free Maxiprep DNA Isolation

Single colonies of bacteria from a LB agar plate were transferred into tubes containing 5 ml of LB medium and antibiotic, and incubated at least for 8 hours, but usually overnight at 37°C in a shaking incubator. 400 to 800 μl of each culture were inoculated into 150 ml of LB medium containing antibiotic and incubated overnight at 37°C in a shaking incubator. Every culture was pelleted by centrifugation at high speed for 40 minutes and the plasmid DNA was purified using

either the Plasmid Maxiprep kit (SIGMA) or the Endo-free Plasmid Maxiprep kit (QIAGEN). The purified plasmid DNA was eluted in 140 to 200 μl of ultra-pure water and stored at -20°C. The endo-free maxiprep was performed only with the plasmids that were going to be transfected into the fibroblast cell lines.

2.2.10 Analysis of Plasmid DNA Integrity and Concentration

Plasmid DNA integrity was analysed by agarose gel electrophoresis. Digestion with restriction enzymes was used in some cases after ligation of an insert in order to confirm the presence of the insert and the structure of the vector.
Plasmid DNA concentration and purity was assessed using a Nanodrop 2000 Spectrophotometer.

2.3 Molecular Biology Techniques

2.3.1 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 100 to 200 ng of cell DNA or 10 to 20 ng 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 μM forward and reverse primers, 8.5 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°C, 30-32 cycles of 30 seconds at 95°C, 30 seconds at appropriate annealing temperature (53-70°C) and 1 minute at 72°C, with a final cycle at 72°C for 5 to 10 minutes.

High Fidelity PCR

For sequences to be cloned or to screen CRISPR-generated clones, High Fidelity Phusion polymerase (2000 units/ml, NEB) was used in a reaction containing the DNA, 1.5 μl of 10 μM forward and reverse primers, 1 μl of 10 mM dNTPs mix (NEB), 0.5 μl of $MgCl_2$ (50 mM), 5 μl of HF buffer, 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°C for 30 seconds, 30 cycles of 10 seconds at 98°C, 30 seconds at 50-65°C and 30 seconds at 72°C, followed by a final cycle at 72°C 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 μl of 10 μM forward and reverse primers, 12.5 μl of Long Amp Taq Mix and water up to a final volume of 20 μl. PCR conditions were: 1 cycle at 94° C for 30 seconds, 30-32 cycles at 94°C for 30 seconds, 59-61°C for 30 seconds, 65°C for 4-5 minutes, and a final cycle at 65°C for 10 minutes.

2.3.2 Gel Electrophoresis

Agarose gels were used to visualize 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 were 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 1 Kb ladder) was always loaded along with the samples and gels electrophoresed at 80 to 110V for 0.5 to 1.5 hours in 1x TAE buffer. Gels were resolved using a BioRad GelDoc system where DNA was visualized under UV light.

2.3.3 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 always measured using a Nanodrop spectrophotometer.

2.3.4 DNA Sequencing

70 to 150 ng/μl of plasmid DNA or 5 to 80 ng/μl of PCR product were sent, along with an appropriate set of primers, to a Sanger sequencing facility in order to sequence the DNA. Sequences were then analysed using ChromasLite and/or BLAST searches.

2.3.5 DNA Extraction from Cells

DNA from fibroblast cells was extracted using the QIAGEN DNeasy Blood & Tissue kit. Briefly, 1 - 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 proteinase K (20 mg/ml, Thermo Fisher Scientific) and 4 μl of RNase A (50 mg/ml, Geneaid) 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°C. 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 8000 *g* and two washes with AW1 and AW2 washing buffers, also followed by centrifugation steps, the DNA was finally eluted in 140 to 200 μl of sterile distilled water and stored at -20°C.

2.3.6 RNA Extraction from Cells

TRI-Reagent (Applied Biosystems) was used for the RNA extraction because it combines phenol and guanidine thiocyanate to rapidly inhibit RNase activ-

ity in order to prevent RNA degradation. All centrifugation steps were carried out at 4°C. Total RNA was extracted from fibroblast cells grown in T75 flasks. When the flasks were very confluent, 1 ml of TRI-Reagent was added and they were incubated at room temperature for 5 or 6 minutes, until the cell monolayer was fully dissociated from the bottom and the cells were homogenized. The homogenate was centrifuged at 12000 *g* for 10 minutes, in order to pellet residual insoluble material. After the addition of 100 μl of bromochloropropane and the centrifugation at 12000 *g* for 15 minutes, the mixture was separated into 3 phases: the aqueous phase (contains RNA), the interphase (contains DNA) and the organic phase (contains proteins). The aqueous phase was then transferred to a new eppendorf tube and 500 μl of isopropanol was added to precipitate the RNA. The solution was centrifuged at 12000 *g* for 8 minutes and the pellet was washed with 75% ethanol. Ethanol was removed by centrifuging at 7500 *g* for 5 minutes and air drying the pellet. Finally, the pellet containing the RNA was dissolved in 100 to 300 μl of ultra-pure water. The concentration and quality of the RNA were measured using the Nanodrop and the RNA was stored at -80°C for future use.

2.3.7 Quantification of DNA and RNA

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

2.3.8 Reverse Transcriptase PCR

RNA was transcribed into cDNA using the iScript cDNA synthesis kit (Bio-Rad). 1 μg of RNA was mixed with 4 μl of the 5x iScript reaction mix, which contains an oligo(dT) and random hexamer primers, 1 μl of the iScript reverse transcriptase and nuclease-free water up to a final volume of 20 μl. Negative reactions for each sample were prepared by mixing 1 μg of RNA with all the reagents mentioned above but without the iScript reverse transcriptase. The reaction mix was incubated in the PCR machine as follows: 5 minutes at 25°C, 30 minutes at

42°C and 5 minutes at 85°C.

Following every new cDNA synthesis, *GAPDH* was amplified from the positive and negative samples using specific primers for GAPDH cDNA, in order to check for any contaminants in the samples. 1 μl of the cDNA was mixed with MegaMix 2x (Microzone) and the primers in a 20 μl reaction. PCR conditions for *GAPDH* were: 1 cycle of 5 minutes at 95°C, 26 cycles of 20 seconds at 95°C, 20 seconds at 58°C and 20 seconds at 72°C, and a final cycle of 5 minutes at 72°C.

For amplification of CDK12 cDNA in Chapter 4, the primers used were *CDK12-cDNA-F'*: 5'-TGGACTTGCTCGGCTCTATAACTC-3' and *CDK12-cDNA-R'*: 5'-CCCAAGAATACATCCACAGCTCCA-3', which produce a 148-bp product. PCR conditions for both CDK12 and GAPDH cDNA primers were as follows: 1 cycle of 5 minutes at 95°C, 26 cycles of 30 seconds at 95°C, 30 seconds at 62°C and 1 minute at 72°C, and a final cycle of 7 minutes at 72°C.

The PCR products were loaded on a 1-1.5% agarose gel to visualize the bands in all cases.

2.3.9 Alternative Splicing Assays

Total RNA was extracted from DM2 cells treated with 5 μM of Dinaciclib or DMSO and 1 μg was reverse transcribed into cDNA using the iScript cDNA synthesis kit (Bio-Rad) according to manufacturers' protocol. 2 μl of cDNA were used as a template for the PCR amplification using MegaMix (Microzone) and primers listed in Table 2.1.

Table 2.1: **Sequences of the primers used to analyse the alternative splicing after treatment of DM2 cells with 5 μM of Dinaciclib.**

Gene	Forward sequence 5'→ 3'	Reverse sequence 5' → 3'
INSR	CCAAAGACAGACTCTCAGAT	AACATCGCCAAGGGACCTGC
PHKA1	TGCACACACTTGAGCTTCATGGA	AAAGTCCACCTCCCCAGACTGGTC
SERCA1	GCTCATGGTCCTTCAAGATCTCAC	AGCTCTGCCTGAAGATGTGTCAC
NCOR	ACACCCACAACCGGAATGAGCCTG	GGACTTGGCTTTTCGGCTGCTG
NFIX	GAGCCCTGTTGATGACGTGTTCTA	CTGCACAAACTCCTTCAGTGAGTC
ITGA6	TCGGGAGTACCTTGGTGGAT	ATCGGGGAACACTGTCATCG
MBNL1	GCTGCCCAATACCAGGTCAAC	TGGTGGGAGAAATGCTGTATGC

Conditions for the PCR amplification were as follow: 1 cycle of 5 minutes at 95°C, 30-35 cycles at 95°C for 30 seconds, 55-60°C for 30 seconds, 72°C for 1 minute and a final cycle at 72°C for 5 minutes. PCR products were analysed on a 2% agarose gel and the intensity of the bands was quantified using the ImageJ software.

2.4 Genome Engineering by CRISPR/Cas9 System

2.4.1 Target Genes

The two genes chosen for the CRISPR projects were *CNBP* and *CDK12*. All the sequence information for them was taken from NCBI or Ensembl.

2.4.2 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. The next tables contain the gRNA sequences that were selected for *CDK12* and *CNBP*.

Table 2.2: **gRNA sequences targeting *CDK12***. The score and number of off-target effects were taken from the CRISPR Design software.

gRNA	Sequence 5' – 3'	Score	# Off-targets
gRNA1	GAGCTGCCGCCTCCCGATGA	86	22
gRNA2	GCTTGGTATCGAAGCACAAG	83	6
gRNA3	AAGTTCAAAGCGTTCGAATG	91	9
gRNA4	CTTCATCGATCCCGACTTGC	97	4
gRNA5	ACGAACGTCGTGGATCAGAT	96	2
gRNA6	CTCCCTTCGGTCTAGTTTGA	87	9
gRNA7	AAACAAGATGATAGCCCCTC	80	12
gRNA8	GGCTCCTGGATCTCCGTTTT	87	26

Table 2.3: **gRNA sequences targeting *CNBP***. The score and number of off-target effects were taken from the CRISPR Design software.

gRNA	Sequence 5' – 3'	Score	# Off-targets
gRNA1	TAACCCTATCGCAGTGAATC	86	6
gRNA2	GATTGCATACATAGTACCTA	83	5

2.4.3 *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 (Figure 2.1 **A**)). An additional oligo containing the gRNA scaffold sequence was also synthesized (Figure 2.1 **B**)).

Each of the oligos was annealed to the scaffold oligo (Figure 2.1 **C**)). For the annealing reaction, 1 μl of each oligo was mixed with 1 μl of the scaffold oligo and 8 μl of water to a final volume of 10 μl. The annealing reaction was carried out in a PCR thermocycler set up as follows: 5 minutes at 95°C, 5 cycles of 2 seconds each, starting at 95°C and decreasing 2°C per cycle, then 580 cycles of 1 second, starting at 83°C and decreasing 0.1°C per cycle, and a final incubation at 10°C for 30 seconds.

Once the PCR run was completed, the next reagents were added to the mixture in order to obtain a double-stranded oligonucleotide: 2.5 μl of dNTPs (10 mM), 2 μl of 10x NEB buffer 2.1, 0.5 μl of T4 NEB DNA polymerase (3000 units/ml) and water up to a final volume of 20 μl. This new mixture was incubated at 12°C for 20 minutes and then PCR purified using the PCR Clean-up kit (Sigma) following manufacturers' protocol (Figure 2.1 **D**)). 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).

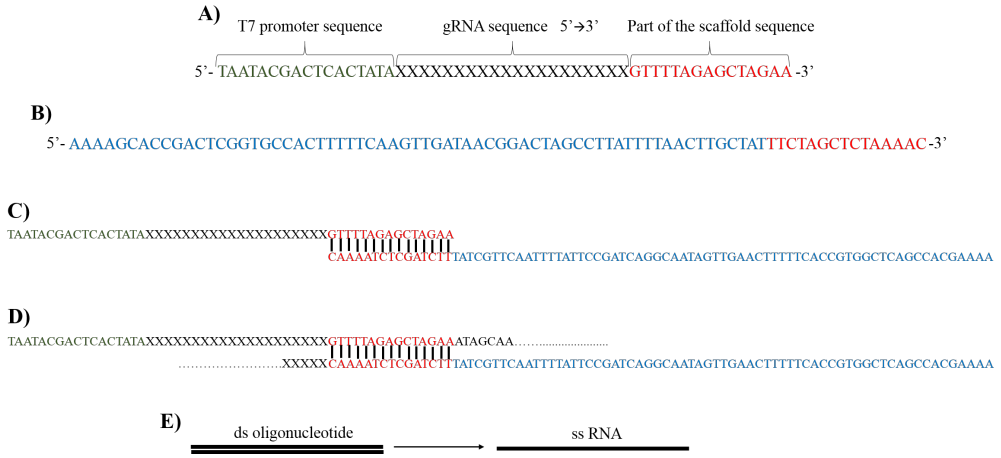


Figure 2.1: **Details of the oligonucleotides used for the *in vitro* assay and the steps followed in order to transcribe the gRNAs *in vitro*. A)** Sequence of the oligos synthesized to be used as gRNAs later on. All of them contain both the T7 promoter sequence at the 5' end and part of the scaffold sequence at the 3' end, whereas they differ in the 20 nt sequence in the middle, corresponding to the gRNA sequence. **B)** Sequence of the scaffold oligonucleotide synthesized. **C)** Annealing of the gRNA oligonucleotides with the scaffold oligonucleotide. **D)** Synthesis of a double-strand oligonucleotide. **E)** *In vitro* transcription of the double-strand oligonucleotide into single-strand RNA, to serve as gRNA.

In vitro transcription of the gRNAs

The double-strand oligos were transcribed into RNA with the MAXIscript T7 kit (Ambion) following manufacturers' instructions (Figure 2.1 **E)**). After an incubation of 4 hours at 37°C with the T7 polymerase, 2 U of TURBO DNase was added to the reaction and incubated for another 15 minutes at 37°C. To clean up the reaction, 2 μl of 5 M ammonium acetate and 60 μl of ice-cold 100% ethanol were added to the transcription reaction and it was incubated overnight at -20°C. The next day the reaction was centrifuged for 30 minutes at maximum speed at 4°C 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

DNA was extracted from DM1 and DM2 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 *CDK12-F'*: 5'-AGGTCAGGGGAAAGGGAATG-3' and *CDK12-R'*: 5'-GCCACTTCTTTCGTAGCTGG-3' for *CDK12*, and *CNBP-gRNA-F'*: 5'-GAGCAGTCACAGGCAATGAG-3' and *CNBP-gRNA-R'*: 5'-CCCAGCCTGTTTTGCCATTT-3; for *CNBP*. 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 Cas9 protein 1 μM (NEB) and water up to 20 μl. This reaction was incubated for 10 minutes at 25°C 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°C. Finally the samples were run on a 2% agarose gel.

2.4.4 Construction of the Cas9 Plasmid

A plasmid co-expressing the gRNA under the control of a U6 promoter and the Cas9 protein under the control of a CBh promoter was purchased from VectorBuilder. An endotox-free maxiprep was performed and the plasmid was electroporated into the cells.

pCas9_gRNA_CDK12 (Figure 2.2) contains the gRNA to target the exon 1 of *CDK12* and the Cas9 sequence. pCas9_gRNA_ZNF9 (Figure 2.3) contains the gRNA to target intron 1 of *CNBP* and the Cas9 sequence. Both plasmids have the same sequence except for the 20-nt gRNA sequence (sequence shown in Appendix 1).

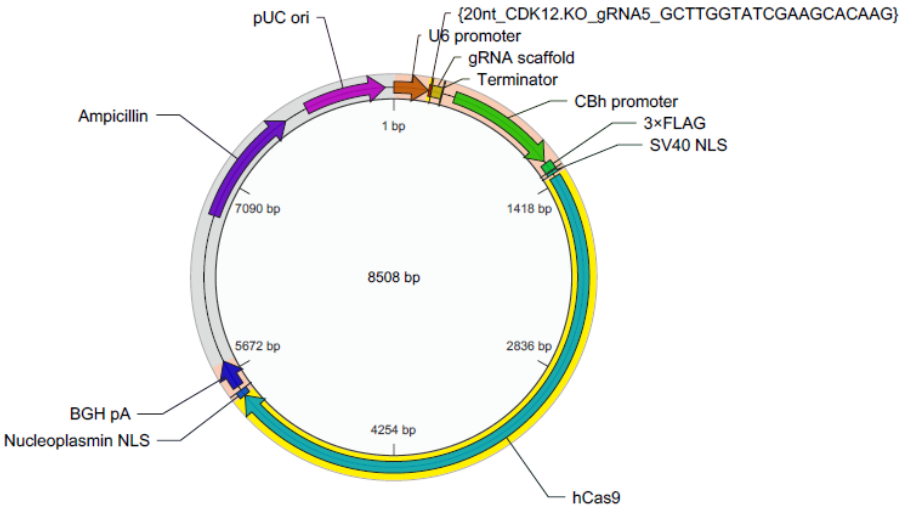


Figure 2.2: **Map of the pCas9_gRNA_CDK12 plasmid.**

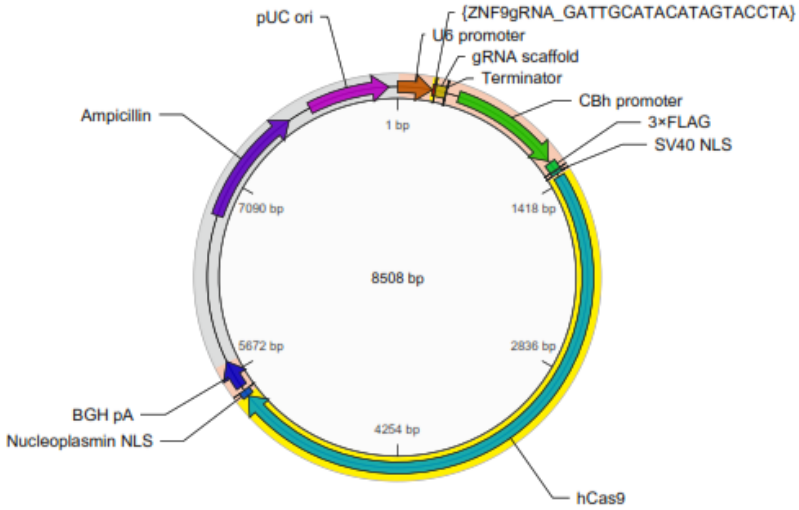


Figure 2.3: **Map of the pCas9_gRNA_ZNF9 plasmid.**

2.4.5 pScramble_Cas9 Plasmid

A scramble gRNA was used as a lethality control for the knock out experiments. The pScramble_Cas9 plasmid was purchased from VectorBuilder. This plasmid contains the *Streptococcus pyogenes*, human codon optimized, Cas9 sequence and a scramble gRNA sequence driven by the U6 promoter. The target sequence for the scramble gRNA is GCACTACCAGAGCTAACTCA. The plasmid was

Endotox-free maxiprepped and confirmed by sequencing before transfected into the cells. The whole sequence of the plasmid is shown in Appendix 1.

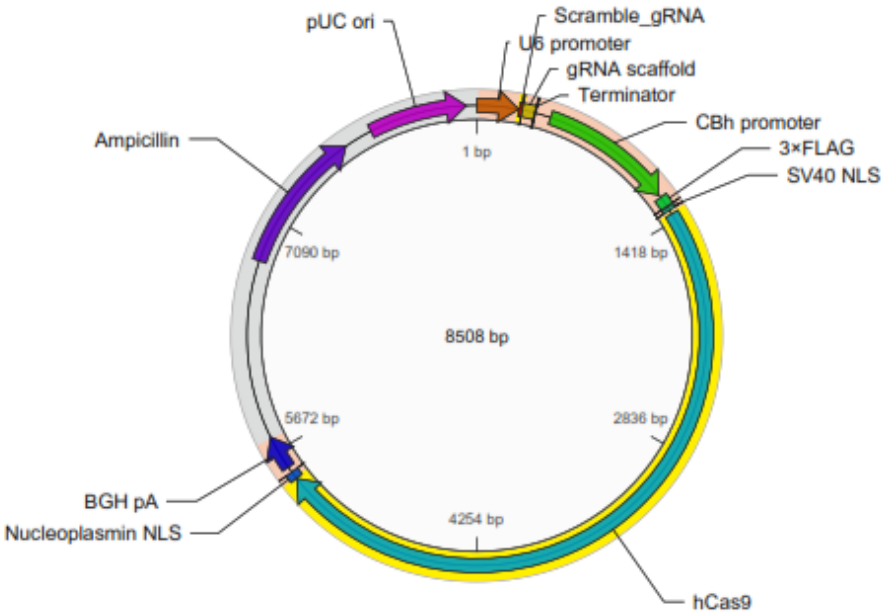


Figure 2.4: **Map of the pScramble_Cas9 plasmid.**

2.4.6 Construction of the Repair Template for Homologous Recombination

The plasmids to serve as a repair template in the CRISPR experiments targeting *CDK12* and *CNBP* were constructed in the laboratory using cloning techniques. The aim was to clone homology arms of around 1 Kb in length flanking the exogenous sequence to be introduced into the cell's genome.

pCDK12_HDR plasmid

To construct the donor plasmid that would serve as a repair template in order to introduce the Neomycin resistance gene into exon 1 of *CDK12*, 1 Kb left and right homology arms were synthesized from GeneScript, containing several restriction sites at both ends of the fragment. Each arm came in a separate pUC57 vector. The Neomycin resistance gene was amplified from the pCl-Neo Mammalian expression vector using primers containing MfeI

restriction sites compatibles with the ones in the left homology arm and cloned next to it. The orientation of Neomycin gene was confirmed by sequencing and the right arm was digested from the other pUC57 vector with NotI and cloned in the pUC57-Neo vector previously linearized with the same restriction enzyme. The plasmid was Endotox-free maxiprepped and confirmed by sequencing. The map of the plasmid used as donor template is shown in Figure 2.5.

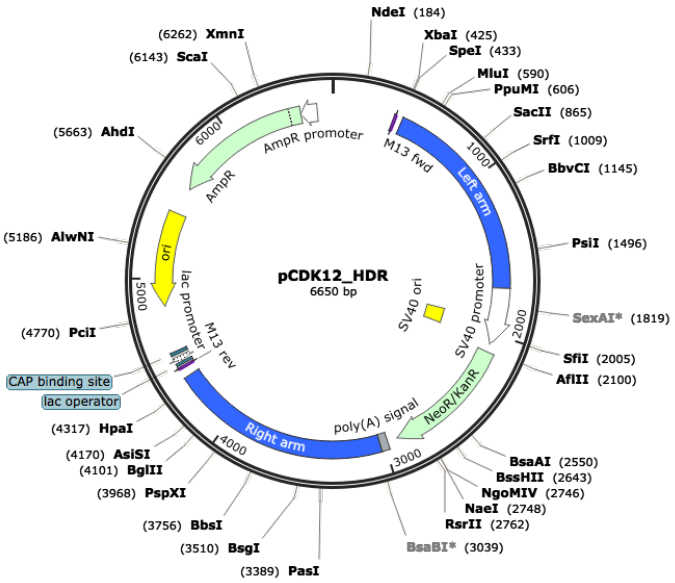


Figure 2.5: **Map of the pCDK12_HDR plasmid.** Fragments in blue correspond to the left and right arms cloned into the pUC57 vector.

pZNF9_HDR Plasmid

In order to construct the donor plasmid containing the inducible promoter and homology arms, 1 Kb homology arms were synthesized from GenScript due to difficulties to amplify the intronic sequence by PCR. The right arm was designed with an EcoRI and an AgeI sites at both ends of the fragment, and the left arm contained PciI sites at both ends. These fragments came in separate pUC57 vectors, which were digested with the aforementioned restriction enzymes and cloned into the pTet-One vector previously digested with the same enzymes. Orientation of the left arm and final structure of the plasmid were confirmed by sequencing. The map of the final construct is shown in Figure 2.6.

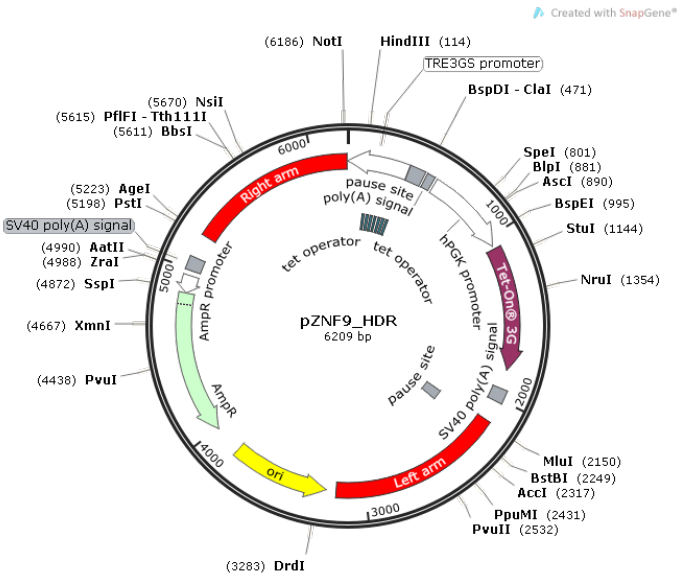


Figure 2.6: **Map of the pZNF9_HDR plasmid.** Fragments in red correspond to the right and left arms cloned into the pTet-One vector.

pZNF9_MMEJ Plasmid

The ZNF9_MMEJ plasmid was synthesized by GenScript. To design the plasmid, the inducible promoter sequence was taken from the pTet-One vector. Then 23-bp homology arms were incorporated at both sides of the promoter, followed by the sequence of the gRNA including the PAM site. This cassette was cloned into the pUC57 vector by GenScript and I performed a maxiprep and sequenced it to confirm the presence of the cassette. A schematic of the cassette designed and the recombination event is shown in Figure 2.7.

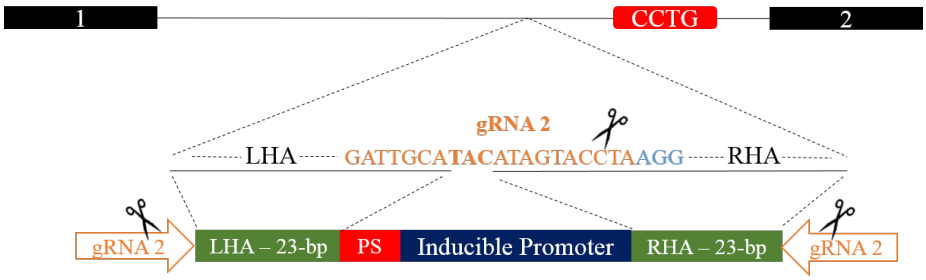


Figure 2.7: **Schematic of MMEJ-mediated gene targeting strategy to knock in the inducible promoter in DM2 cells.** The gRNA target sequence is shown in orange and the PAM sequence in blue. Black scissors indicate Cas9 cleavage sites. LHA: Left Homology Arm, RHA: Right Homology Arm, PS: Pause Site.

2.5 Transgene Generation

2.5.1 Solutions, Media and Antibiotics

The Amaxa Basic Nucleofector Kit for Mammalian Fibroblasts (#VPI-1002) was used to electroporate the cells. The kit contains a Basic nucleofector solution and a supplement that are mixed prior to the electroporation. Bullet media (FGM-2 Bulletkit, Lonza) was used to grow the cells for 2 days in the incubator after the electroporation. Hygromycin-B (Thermo Fisher Scientific #10687010) was the antibiotic used in the selection process.

2.5.2 Titration of a Kill Curve for Hygromycin

To determine the best concentration of antibiotic required to select the transfected cells after the transfection with the CRISPR plasmids, a kill curve was performed. Hygromycin-B was used for cell selection.

Cells were aliquoted on a 24-well plate 24 hours before treatment, with 50000 cells per well. Hygromycin-B was used to treat the cells in 12 different concentrations and in duplicate: 0, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400 and 500 μg/mL. The media was replaced every 3 days with fresh media and antibiotic, and the cell viability was checked under the microscope daily. After 14 days of

treatment, the minimum concentration of Hygromycin-B that killed all the cells was assessed.

2.5.3 Transfection of the Cells

DM1 and DM2 human fibroblasts were transfected by electroporation using Amaxa nucleofection (Lonza). Briefly, for each transfection 1 to 2 x 10^6 cells were washed with PBS, trypsinized and pelleted by centrifugation at 100 g for 10 minutes. Then each pellet was resuspended in 100 μl of nucleofector mixture (82 μl Solution-1 and 18 μl Solution-B); the cell suspension was mixed with the DNA, placed in the cuvette provided with the kit and electroporated using the program U-23. Next, cells were transferred to a T75 flask containing 20 ml of pre-warmed Bullet media and kept in the incubator for 48 hours.

2.5.4 Antibiotic Selection of Positive Clones

Each T75 flask containing the electroporated cells was washed, trypsinized and split into 6-8 petri dishes 48 hours post-transfection. After 3 days in the petri dishes, DMEM media was replaced with fresh DMEM containing 50 μg/ml of Hygromycin-B. Petri dishes were placed back in the incubator and media containing Hygromycin-B was replaced every 3-4 days until the formation of colonies.

2.5.5 Colony Pick Up

After two weeks of antibiotic selection colonies started to appear in the petri dishes. To pick up the colonies, the medium from the petri dish was removed and it was washed with PBS. Pyrex cloning cylinders of 10 mm of diameter were dipped in vaseline and placed in the dish surrounding single colonies, in order to isolate each one of them from the rest. 150 μl of Trypsin-EDTA was added inside the cloning cylinders and the dish was incubated at 37°C for 5 minutes. Each colony was visualized under the microscope to confirm that all the cells detached completely from the surface of the dish and 500 μl of DMEM medium

with 50 μg/ml of Hygromycin-B was added to the cloning cylinder and mixed with the cells. The colonies were then transferred to a 24-well plate and grown in the incubator at 37°C.

2.5.6 Colony PCR of Human Fibroblasts

Colonies resistant to Hygromycin-B were seeded onto 24-well plates. Once they were confluent in each well they were washed with 1 ml of PBS and detached from the surface of the plate with 250 μl of Trypsin-EDTA. 750 μl of media was added to the well, mixed, transferred to an eppendorf tube and centrifuged for 5 minutes at 10500 rpm to pellet the cells. The supernatant was removed and the pellets were washed with PBS and re-centrifuged. The pellets were then suspended in 30 μl of Tris-HCl pH 8.5 and boiled for 10 minutes. To digest any proteins present in the samples, 1.5 μl of Proteinase K (20 mg/ml) was added and samples were incubated at 55°C for 10 minutes. Proteinase K was inactivated by boiling again for 10 more minutes. Finally, 8 μl of this solution was mixed with 12.5 μl of Long Amp Taq 2x Mix, 1 μl of 10 μM forward primer, 1 μl of 10 μM reverse primer and water up to a final volume of 25 μl, to perform PCR amplification.

2.5.7 Identification of Mutations in the Target Site

To detect possible mutations in the target site, it was amplified using the Long Amp Taq DNA polymerase in a reaction described previously. PCR program for ***CDK12*** was: 1 cycle of 30 seconds at 94°C, followed by 29 cycles of 30 seconds at 94°C, 30 seconds at 62°C, 4:20 minutes at 65°C and a final cycle of 10 minutes at 65°C. The primers used for *CDK12* were *CDK12-F*: 5'-AGGTCAGGGGAAAGGGAATG-3' and *CDK12-R*: 5'-GCCACTTCTTTCGTAGCTGG-3'. PCR program for ***CNBP*** was: 1 cycle of 30 seconds at 94°C, followed by 31 cycles of 30 seconds at 94°C, 30 seconds at 61°C, 4:30 minutes at 65°C and a final cycle of 10 minutes at 65°C. The primers used for *CNBP* were *CNBP-gRNA-F'*: 5'-GAGCAGTCACAGGCAATGAG-3'

and *CNBP-gRNA-R'*: 5'-CCCAGCCTGTTTTGCCATTT-3.
Samples were run on a 1.5% agarose gel, the DNA was extracted from the gel as previously described and sent for sequencing. The sequence from each sample was analysed using the software TIDE (Tracking of Indels by DEcomposition) and blasted against the sequence from a control sample.

2.6 Protein Analysis

2.6.1 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:

- **RIPA buffer** (lysis buffer): NaCl 150 mM, Tris 50 mM (pH 8.0), NP-40 1%, SDS 0.1%, Deoxycholic acid 0.5%, Protease inhibitor cocktail (AEBSF, aprotinin, bestatin, E-64, leupeptin and pepstatin A), Phosphatase inhibitor cocktail 2 (sodium orthovanadate, sodium molybdate, sodium tartrate and imidazole), Phosphatase inhibitor cocktail 3 (cantharidin, (-)-p-bromolevamisole oxalate, calyculin A) and dH_2O. The Protease and Phosphatase inhibitor cocktails are commercially available and were acquired from Sigma-Aldrich.

- **NuPAGE mix**: 2.5 μl of 4x NuPAGE LDS sample buffer and 1 μl of 10x NuPAGE sample reducing agent. The volumes listed are for 1 sample. For more samples the volume of each solution needs to be increased.

- **Running buffer**: 1x NuPAGE MOPS SDS running buffer (20x) and dH_2O.

- **Transfer buffer**: 1x NuPAGE transfer buffer, 10% Methanol and 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% and dH_2O. The pH of the TBS-Tween was adjusted to 7.6 using HCl.

- **ECL Developing solution**: Clarity western peroxide reagent and Clarity western luminol/enhancer reagent. 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.

- **Antibodies**: Primary antibodies: anti-CrkRS Mouse mAb (Abcam ab57311) and anti-GAPDH Mouse mAb (Merck Millipore 6C5). Secondary antibody: HRP-conjugated anti-mouse antibody.

2.6.2 Protein Extraction from Cells

Confluent T75 flasks were used to extract proteins. 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°C. 10 ml of DMEM media were added to neutralize 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 3000 rpm. The supernatant was decanted and the pellet was resuspended in 200 to 300 μl of RIPA buffer. The eppendorf tubes were left on ice for 30 minutes followed by two freeze-thaw cycles at -80°C. Finally the protein lysate was collected by centrifuging at top speed at 4°C for 30 minutes. The supernatant was collected and stored at -80°C until further processing.

2.6.3 Protein Quantification

Protein concentration was determined using the BioRad DC Protein assay. BSA at a concentration of 1 mg/ml was used to produce a standard curve. 10 μl of the protein sample were mixed with the reagents from the BioRad 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 μg/ml to 50 μg/ml and the standard curve was used to estimate the

concentration of the protein samples.

2.6.4 Detection of *CDK12* on Western Blot

Western Blot was performed using a commercial NuPage system (Invitrogen) according to the manufacturer's instructions.

Sample preparation

A volume reflecting 10 to 30 μg of protein was mixed with 3.5 μl of NuPAGE Mix. Deionized water was added in some cases to adjust the volume to 14 μl in total. Samples were heated at 70°C 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 μl of protein sample was 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 200V.

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 20V.

Protein detection

The membrane was rinsed twice in a tray with distilled water (10 minutes each wash) and blocked overnight at 4°C 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 1:10000, 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 μl 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 visualized using a LAS-3000 Imaging System.

2.7 Microscopy

The imaging of cells in this work was performed with Brightfield, Micro Widefield Plate Reader and Confocal microscopes.

2.7.1 Brightfield Microscopy

A brightfield microscope fitted with 4x, 10x and 40x objectives was used for cell imaging in tissue culture.

2.7.2 Micro Widefield Plate Reader

After performing ICC and FISH, cells were visualized 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.

2.7.3 Confocal Microscopy

Zeiss LSM710 Confocal Microscope was used to visualize the cells after ICC or FISH when seeded on 35mm petri dishes. The 63x oil immersion objective was selected to visualize the cells. Fluorescent images were acquired with the Zen browser and analysed using LSM Image Browser or ImageJ software.

2.8 Electronic Databases

Information about nucleotide, transcripts and protein sequences was obtained using National Centre for Biotechnology Information (NCBI) and Emsembl.
Primers were designed using Primer3 Input version 4.0.0.
gRNAs were designed using either ChopChop version 2 or the CRISPR Design software from the Zhang lab (crispr.mit.edu).

Chapter 3

Characterization of the effect of kinase inhibitors on DM cells

3.1 Aim

The aim of this chapter was to assess the effect of different kinase inhibitors on repeat expansion foci in DM2 cells in order to see if they produce the same effect as in DM1 cells. The effect of the compounds on SC35 speckles, MBNL1 localization and alternative splicing was also studied in both DM1 and DM2 cells. Furthermore, co-localization experiments were performed in both cell types to examine the relationship of SC35 speckles, CDK12 and foci that could explain some of the differences in the pathogenesis and symptoms between the two types of the disease.

3.2 Introduction

Myotonic dystrophy type 1 and type 2 are forms of muscular dystrophy caused by microsatellite repeat expansion mutations. For DM1 it is a CTG expansion located in the 3'UTR of *DMPK* (Brook et al., 1992; Mahadevan et al., 1992). In the case of DM2 the repeat expansion is CCTG, located within the first intron of *CNBP* (Liquori et al., 2001). Despite sharing a gain-of function pathome-

chanism, there are differences between the two disorders, the origins of which have yet to be resolved. DM1 has been more extensively studied, whereas less research has been performed for DM2. Holt *et al.* observed that in DM1, foci aggregate at the periphery of SC35 speckles, whereas in DM2 foci are not associated with these structures (Holt et al., 2007). In this chapter I examined the effect on DM2 cells of kinase inhibitors which target CDK12, a protein that co-locates with SC35 speckles (Ko et al., 2001), and that were previously demonstrated to eliminate foci in DM1 cells by others in the laboratory (unpublished). I also studied other potential beneficial effects of the compounds in both DM1 and DM2 cells. Before testing the kinase inhibitors, the localization of foci in the nucleus and their relationship with SC35 speckles and CDK12 was studied in the DM1 and DM2 cell lines used. The relationship of foci with SC35 speckles and CDK12 was studied by combining *in situ* hybridization (FISH) with immunocytochemistry (ICC) experiments.

Strategies to develop a therapy for myotonic dystrophy are mainly focused on the elimination of the toxic RNAs that form foci in the nucleus of the cells. Elimination of foci could eliminate the toxicity caused by the repeat expansion and release MBNL1, restoring the alternative splicing and alleviating some of the symptoms of the disease. The technologies used to eliminate foci include rybozymes (Langlois et al., 2003), antisense oligonucleotides (Mulders et al., 2009; Wheeler et al., 2012), small molecules (Siboni et al., 2015; Rzuczek et al., 2015, 2017) and chemical inhibitors (Warf et al., 2009b; Childs-Disney et al., 2012; Ketley et al., 2014; Wong et al., 2014).

Some of the technologies mentioned above have important toxic effects or delivery problems that need to be solved before they can become a potential therapy. Small molecules are desirable because of their potential to penetrate the cell membrane, compared to other bigger molecules. For this work, 6 different small molecules, kinase inhibitors, (from Ketley *et al.*, unpublished) that were proven to eliminate foci in DM1 cells, were tested here in DM2 cells to assess foci reduction and additional effects on other components of the cell that play a role in myotonic dystrophy pathogenesis. A mechanism of action for the kinase inhibitors has been proposed based on the results obtained. Four additional ki-

nase inhibitors, generated on an ongoing project in the laboratory by modifying the original kinase inhibitors and with the aim to increase their selectivity towards CDK12, were also tested in DM2 cells. They were already tested in DM1 cells and proved to be effective to reduce foci in the nucleus of the cells.

3.3 Investigation of the localization of foci in the nucleus of DM cells

3.3.1 Co-localization of foci with SC35 speckles

In 2007 Holt *et al.* reported that DM1 foci associate with nuclear splicing structures, but this does not apply to DM2 foci. When using antibodies against SC35, Y12 or 9G8, DM1 foci appeared in close proximity to these structures, whereas DM2 foci were independent and free in the nucleus (Holt et al., 2007). In this work I've tried to establish the degree of co-localization between foci and SC35 speckles in the DM1 and DM2 fibroblast cell lines used, to see whether foci in these cell lines behave as stated by Holt regarding SC35 speckles association. SC35 speckles were stained with an anti-SC35 monoclonal antibody raised in mouse, and foci were detected using a 647-labelled $(CAG)_{10}$ probe for DM1 or a 647-labelled $(CAGG)_8$ probe for DM2. Images were acquired using the 710 Confocal microscope and processed with ImageJ.

As can be seen in Figure 3.1, in DM1 foci strongly co-localize with SC35 speckles. Independent foci can occasionally be found in the nucleus probably due to the fact that foci are very dynamic structures that are constantly changing. The results in the DM2 cell line used for this work match with those reported by Ian Holt. Even though some DM2 foci co-localize with SC35 speckles, the majority of foci are independent in the nucleus. As these images only show a few foci, up to 35 foci were analysed from additional cell images and the ratio of independent versus associated foci was established for each cell line. Figure 3.2 shows the results obtained after the quantification, and the degree of significance obtained by performing a chi-square test. These results further confirm those

observed in the images from Figure 3.1 and already reported in the literature, that in the DM1 cell line used here, foci associate with SC35 speckles, but this is not the case in the DM2 cell line used.

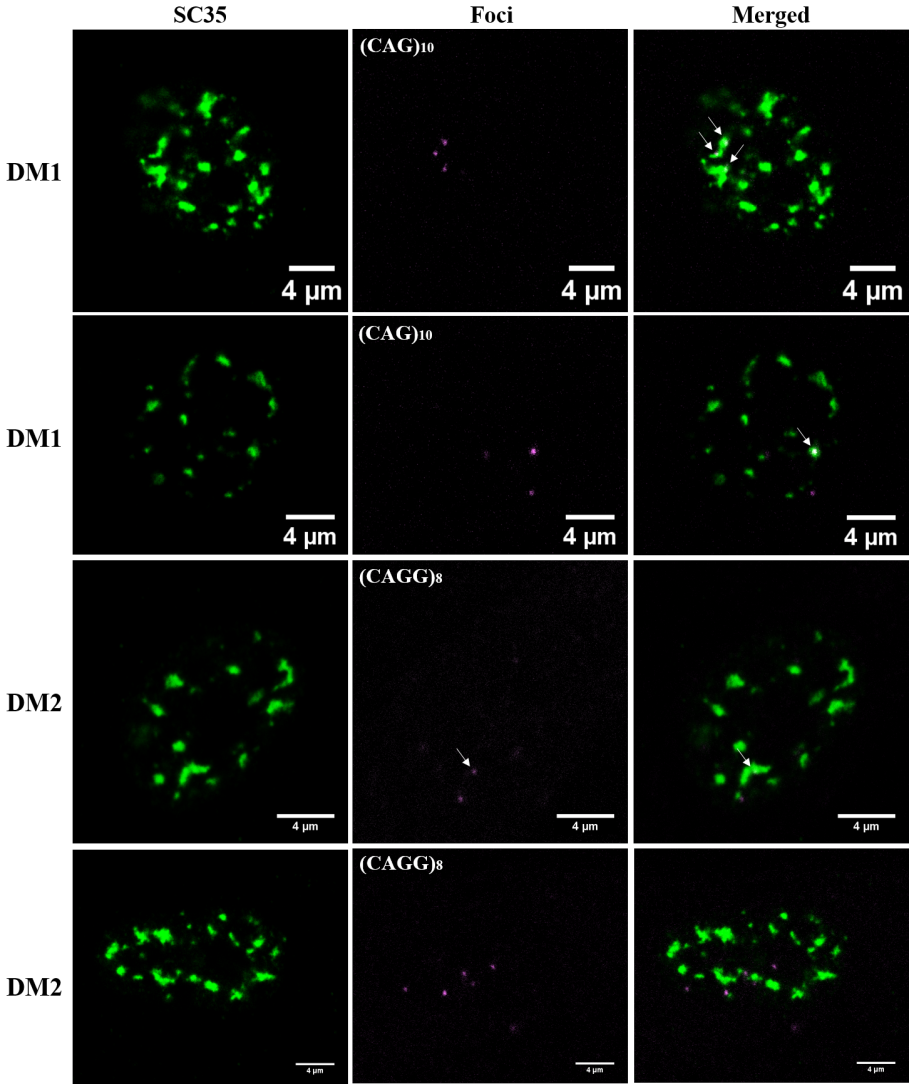


Figure 3.1: **Co-localization or association of nuclear foci with SC35 speckles in DM1 and DM2 cells.** Most nuclear foci labelled with $(CAG)_{10}$ (middle panel) co-localize with SC35 speckles (right panel) in DM1. In DM2 most nuclear foci labelled with $(CAGG)_8$ (middle panel) do not co-localize with SC35 speckles (right panel).

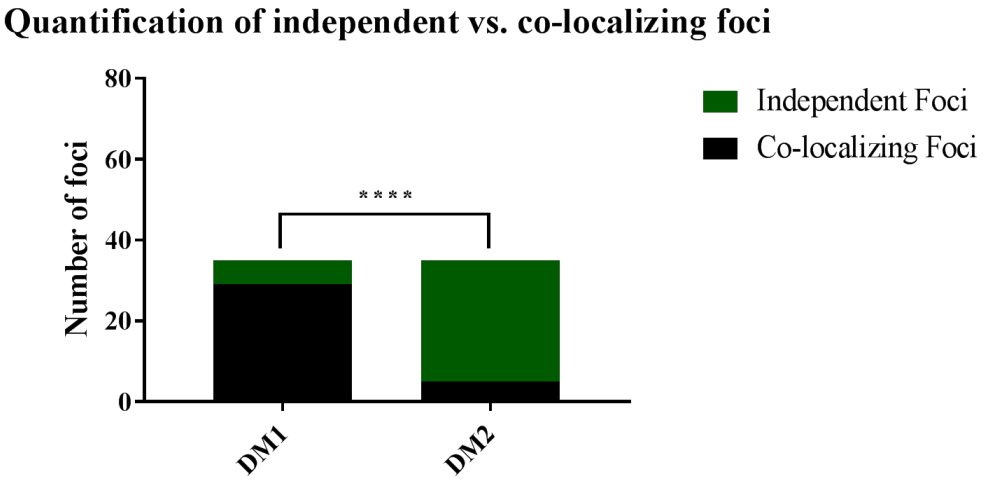


Figure 3.2: **Quantification of the number of foci independent or co-localizing with SC35 speckles in DM1 and DM2 cells.** 35 foci were counted per cell line and the ratio of independent and co-localizing foci with SC35 speckles was established. Out of the 35 foci analysed, DM1 cells presented 29 foci that co-located with SC35 speckles and 6 foci independent in the nucleus. On the contrary, DM2 cells presented 5 foci that co-located with SC35 speckles and 30 foci independent in the nucleus. Chi-square test revealed a statistically significant difference between the two cell lines. $^{****}p<0.0001$.

3.3.2 Co-localization of foci with CDK12

CDK12 is involved in the elongation phase of transcription. Furthermore it plays a role in the processing of mRNAs and it is found in SC35 speckles. If foci co-localize with SC35 speckles, they should also co-localize with CDK12 spots. In order to test this hypothesis the co-localization experiments were repeated in DM1 and DM2 cell lines, but this time an anti-CDK12 monoclonal antibody raised in mouse was used to stain the CDK12 protein. $(CAG)_{10}$ probes were used to detect DM1 foci and $(CAGG)_8$ were used for DM2. Clear-well defined results were very difficult to obtain from this experiment, as the combination of FISH and ICC appeared to be incompatible, especially in DM2 cells.

In order to optimize the co-localization experiment, different methods were used to fix the cells in a series of different experiments. Originally, for FISH cells are fixed with 4% PFA, but CDK12 spots can not be detected with this fixative, as PFA appears to affect the epitope or the CDK12 antibody. The best fix-

ation method to detect CDK12 was found to be 50:50 ice cold MetOH:Acetone, however foci seem to be degraded in this case, especially in DM2 cells. Thus, the fixatives used in the optimization process were:

- 50:50 ice cold Methanol:Acetone for 30 minutes at -20°C
- 4% PFA for 30 minutes at room temperature followed by permeabilization with 50% Acetone in PBS for 10 minutes at room temperature
- 4% PFA for 30 minutes at room temperature followed by permeabilization with 20% Acetone in PBS for 10 minutes at room temperature
- 4% PFA for 30 minutes at room temperature followed by permeabilization with 0.03% Triton X-100 in PBS for 10 minutes at room temperature
- 4% PFA for 30 minutes at room temperature followed by permeabilization with 6% Acetic Acid for 2 hours at room temperature
- 100% ice cold Methanol for 30 minutes at room temperature

Initially FISH was performed first, followed by ICC. Before starting the ICC protocol cells were imaged in order to visualise foci. When the ICC was finished cells were imaged again and the quality of the foci signals were compared to the images from the previous day before performing ICC.

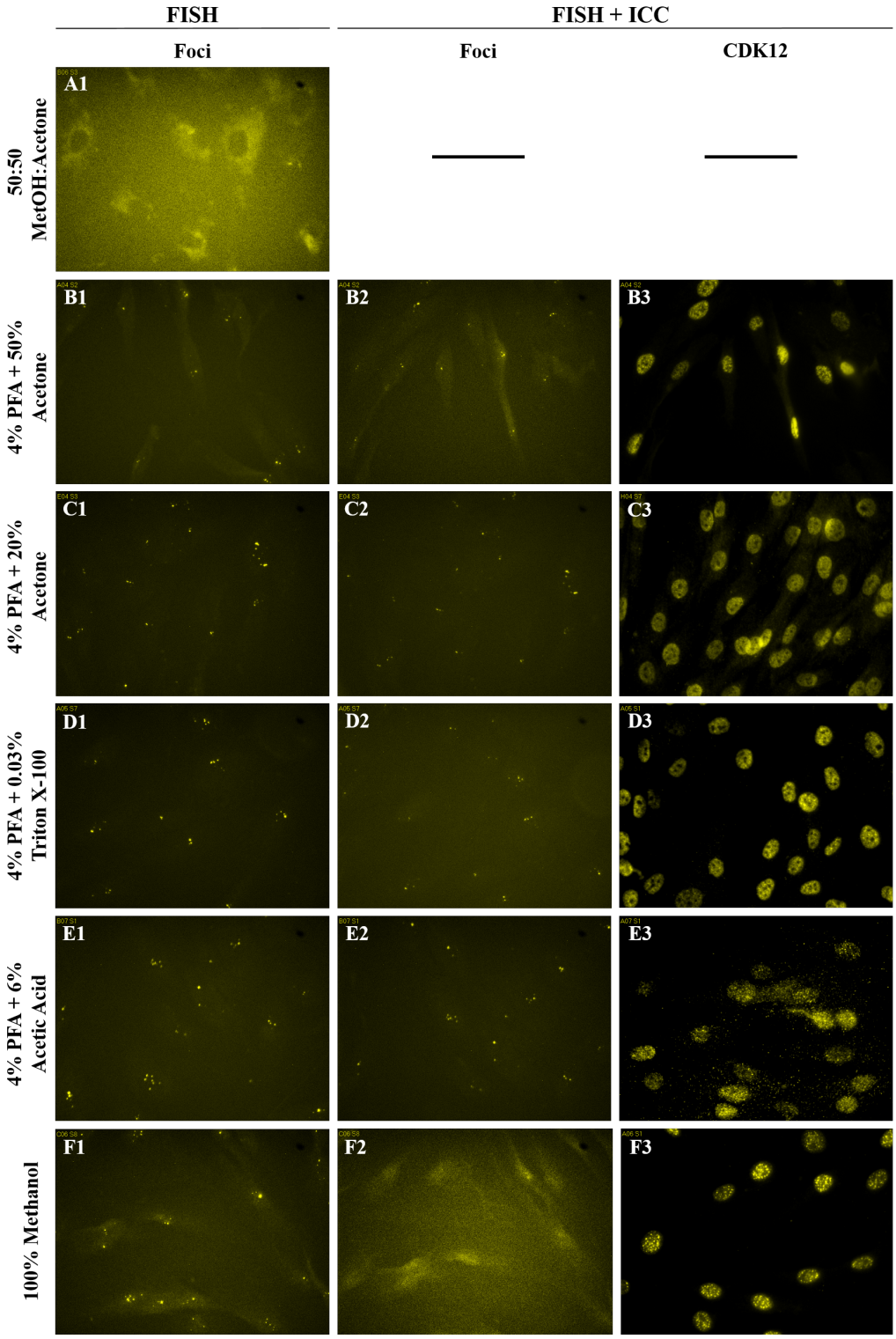


Figure 3.3: **Optimization of different fixative methods for the combination of FISH with ICC in order to study the co-localization of foci with CDK12 protein.** FISH was performed using different fixative methods and cells were imaged (**column A1 - F1**). ICC with an anti-CDK12 antibody was started straight after the FISH. Cells were imaged at the end of the experiment and foci signals were compared before and after the ICC (**column B2 - F2**) and the effect of the different fixatives on the anti-CDK12 antibody signal was observed (**column B3 - F3**).

Fixation of the cells with 50:50 ice cold Methanol:Acetone (Figure 3.3 **A1**) affected foci as none could be detected in the nucleus of the cells after fixation. ICC was not performed and Methanol:Acetone was concluded to not be a suitable fixation method for FISH. Foci could be detected when cells were fixed and permeabilized with 4% PFA and 50% Acetone (Figure 3.3 **B1 - B3**). Performance of the ICC didn't seem to affect foci, and CDK12 spots could also be visualized. Thus this could be a suitable method to perform the co-localization, however, the CDK12 signal could be improved, as CDK12 spots could be detected but the quality of the signal was not very good. Fixation and permeabilization with 4% PFA and 20% Acetone (Figure 3.3 **C1 - C3**) was suitable to detect foci. However, 20% Acetone didn't seem to be sufficient to permeabilize the cells so the antibody could not properly penetrate into the nucleus and detect clear spots of CDK12.

4% PFA and 0.03% Triton X-100 (Figure 3.3 **D1 - D3**) as a fixative and permeabilization method allowed detection of foci. The signal of foci was dimmer after performing ICC but still foci could be detected. However, the CDK12 staining was not clear, probably because permeabilization with 0.03% Triton X-100 is not enough for the antibody to completely enter the nucleus of the cells. Permeabilization of the cells with 6% Acetic acid after fixation with 4% PFA (Figure 3.3 **E1 - E3**), was suitable for foci detection, and the signals were maintained even after the ICC. However, the acetic acid seemed to degrade the antibody, as high background was obtained and CDK12 spots could not be clearly detected. Finally, fixation of the cells with 100% Methanol (Figure 3.3 **F1 - F3**) allowed foci detection when FISH was performed on its own. However, after performing ICC, foci could not be detected any more. This may be due to the combination of the methanol with some of the reagents used for the ICC protocol, that somehow affect the RNA or the probe. Nevertheless, fixation with methanol allowed for the best CDK12 staining, as very good quality spots could be detected.

To further optimize the experiment, ICC was performed first, followed by FISH. In this case, different fixative methods were also tried (images not shown), and it was concluded that the best one was 50:50 ice cold Methanol:Acetone for 30 minutes at the beginning of the ICC, followed by an additional fixation of

5 minutes with 4% PFA the next day, at the start of the FISH protocol. Still Methanol:Acetone is not the best fixative for foci, as it appears to affect signal as fewer foci than usual were found in the cells. Figure 3.4 shows the confocal images obtained after performing the experiment in this way. From the images it can be seen very clearly that in DM1 cells foci co-localize with CDK12 spots, as expected. It is worth mentioning though, that this does not necessarily mean that foci and CDK12 interact physically. Further experiments would be required to analyse this. On the contrary, in DM2 cells foci could not be properly identified despite several attempts. This could be due to a higher sensitivity of the DM2 cells to the chemistry of the procedure or an effect on the CAGG probe. The co-localization experiment was repeated many times trying to optimize the protocol to successfully detect both CDK12 spots and foci in DM2, however it wasn't possible and thus the degree of co-localization of foci with CDK12 in DM2 hasn't been directly established.

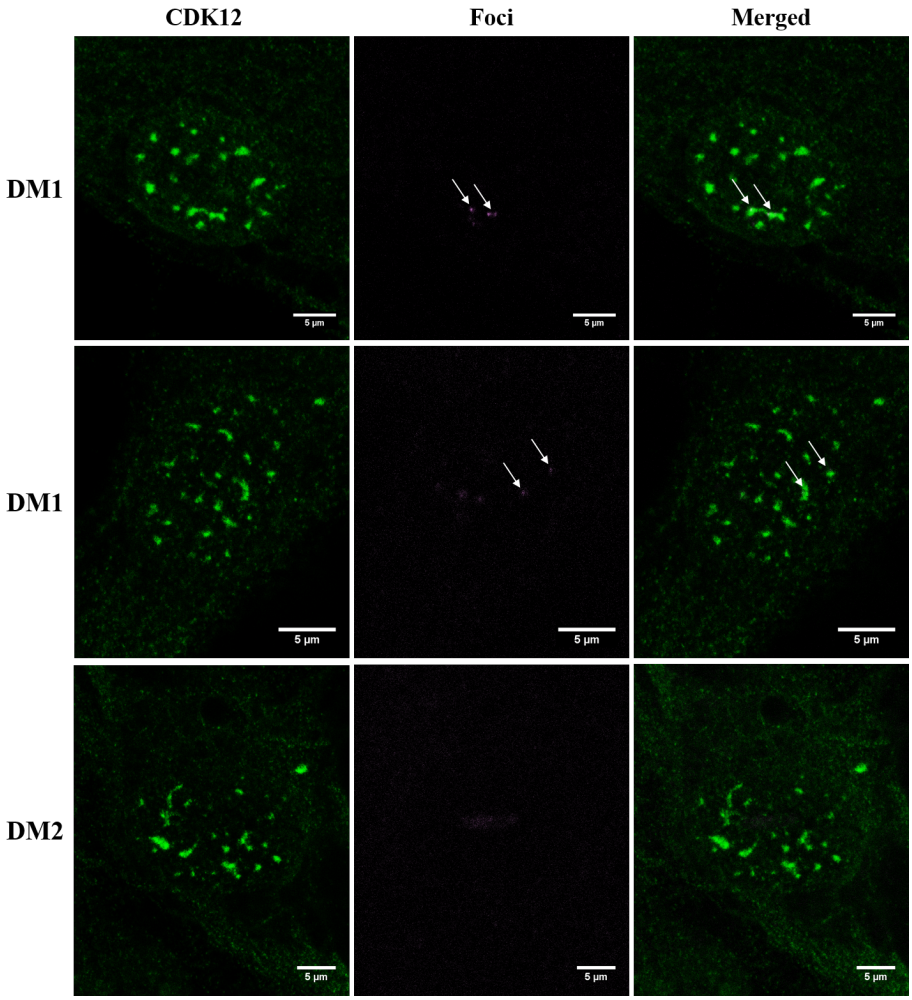


Figure 3.4: **Co-localization of nuclear foci with CDK12 in DM1 but not in DM2 cells.** Nuclear foci labelled with $(CAG)_{10}$ (middle panel) co-localize with CDK12 spots (right panel) in DM1 cells. Combination of ICC and FISH appears to affect foci in DM2 cells as no signal from the $(CAGG)_8$ probe (middle panel) could be detected. Degree of co-localization of foci with CDK12 in DM2 cells could not be established using this method.

3.3.3 Co-localization of CDK12 with SC35 speckles

CDK12 is a protein with a very high molecular weight (164 KDa). The fixative methods required to perform a good CDK12 ICC appear to make it difficult to visualise foci, thus is difficult to combine CDK12 ICC with FISH. From the literature it is known that CDK12 co-localizes with SC35 speckles, where it plays a role in pre-mRNA processing. However, I set out to study the degree of co-localization of these two proteins in the non-DM, DM1 and DM2 cell lines used

for this work, to look for any differences in co-localization of the two proteins. Figure 3.5 shows the confocal images of the CDK12-SC35 co-localization in non-DM, DM1 and DM2 cells. The signal from the anti-SC35 antibody overlaps with the signal from the anti-CDK12 antibody in all three cell lines. However, a difference in the degree of overlap can be appreciated, with DM cells showing a higher degree of co-localization compared to non-DM cells.

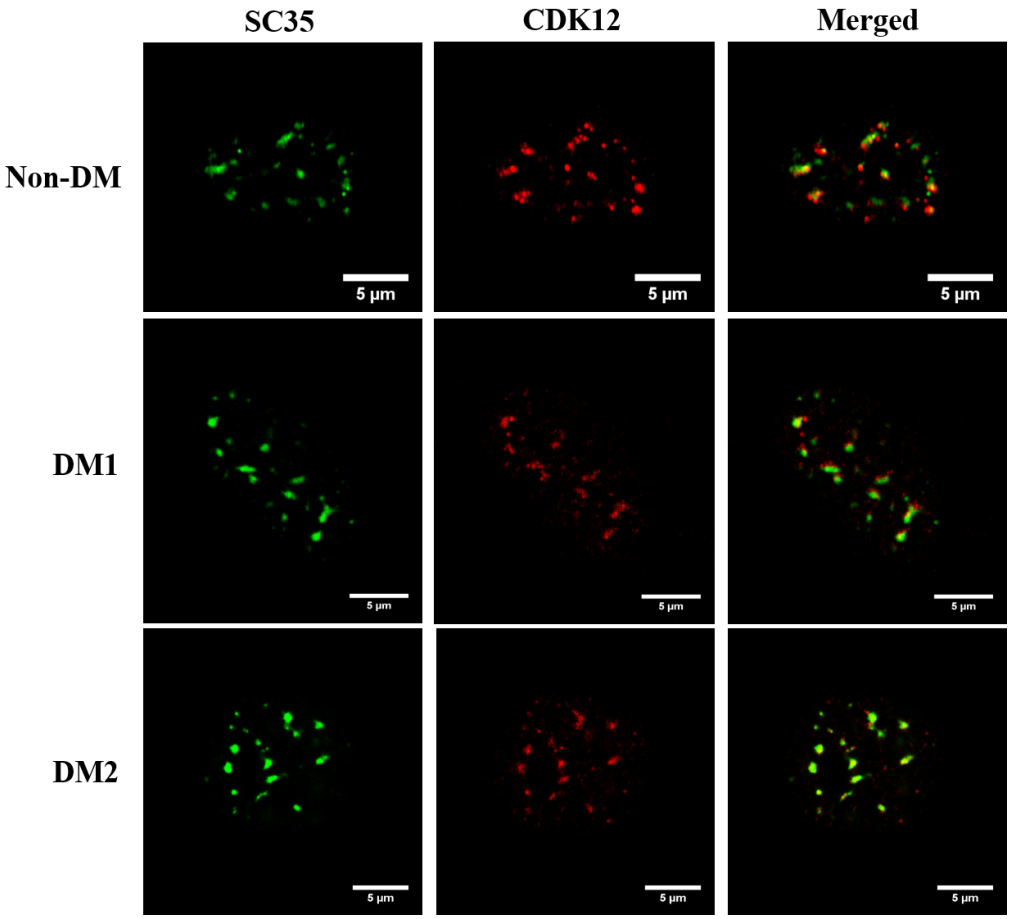


Figure 3.5: **Co-localization of SC35 speckles with CDK12 spots in non-DM, DM1 and DM2 cells.** CDK12 spots co-locate with SC35 speckles in the 3 cell lines. However, a difference in the degree of co-localization can be seen across the different cell lines. Green spots correspond to SC35, red spots correspond to CDK12.

In order to statistically analyse the degree of co-localization of the two proteins, which however, does not necessarily mean a physical interaction, the linear correlation between the signal from the anti-CDK12 antibody and the signal from the anti-SC35 antibody was measured by quantifying the degree of overlap of the signals and the Pearson's Correlation Coefficient (PCC) was es-

tablished. PCC measures the pixel-by-pixel covariance in the signal levels of two images and is used to quantify co-localization. The formula for PCC for an image consisting of two channels (usually red and green) is shown in Figure 3.6, where R_i and G_i refer to the intensity values of the red and green channels, respectively, of pixel *i*, and $\overline{R}$ and $\overline{G}$ refer to the mean intensities of the red and green channels, respectively, across the entire image. PCC values close to 1 mean that fluorescence intensities are linearly related, therefore indicating there is a considerably degree of co-localization. Values of PCC close to zero, reflect distribution of probes that are uncorrelated with one another (Dunn et al., 2011).

$$\mathrm{PCC} = \frac{\sum_i (R_i - \overline{R}) \times \left(G_i - \overline{G}\right)}{\sqrt{\sum_i (R_i - \overline{R})^2 \times \sum_i \left(G_i - \overline{G}\right)^2}}$$

Figure 3.6: **Pearson's Correlation Coefficient formula.** R_i and G_i refer to the intensity values of the red and green channels, respectively, of pixel *i*, and $\overline{R}$ and $\overline{G}$ refer to the mean intensities of the red and green channels, respectively, across the entire image. The closer the PCC value is to 1, the highest co-localization exists between the two probes, or in this case, the two proteins.

PCC values were calculated for each of the cell lines and results were plotted in the graph shown in Figure 3.7, which shows the PCC values for non-DM, DM1 and DM2 cells based on the analysis of 20 cells per cell line. The graph reflects what can be appreciated at simple sight from the confocal images, that in non-DM cells CDK12 and SC35 speckles co-localize but at a lower degree compared to DM cells, with DM2 cells showing the highest degree of co-localization of CDK12 with SC35 speckles. Furthermore, an ANOVA test showed that the difference is statistically significant between the three cell lines.

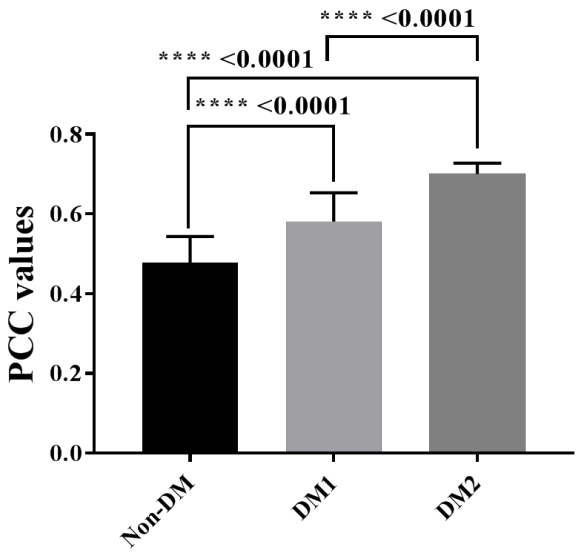


Figure 3.7: **Pearson's Correlation Coefficient for SC35 and CDK12 co-localization.** Analysis of 20 cells per cell line showed greater co-localization in DM2 than in DM1, and in DM1 than in non-DM cells. Lines above each bar represent the standard deviation. The statistical significance was obtained performing an ANOVA test. $^{****}p<0.0001$.

The degree of co-localization of CDK12 with SC35 speckles in the non-DM and the DM cell lines appears to be significantly different. The number of CDK12 spots and SC35 speckles and their area were quantified and Figure 3.8 shows the results obtained. DM cells have slightly higher number of SC35 speckles, nonetheless the total area covered by these speckles is very similar in both non-DM and DM cells. In the case of CDK12 spots, DM cell lines have slightly fewer CDK12 spots compared to non-DM cells, however the area of the spots in the DM cell lines appears to be bigger than in non-DM cells. The trend of a higher area of CDK12 spots in DM1 and in DM2 could account for the higher degree of co-localization compared to non-DM cells. Nevertheless, the differences observed are not statistically significant, as demonstrated by an ANOVA test.

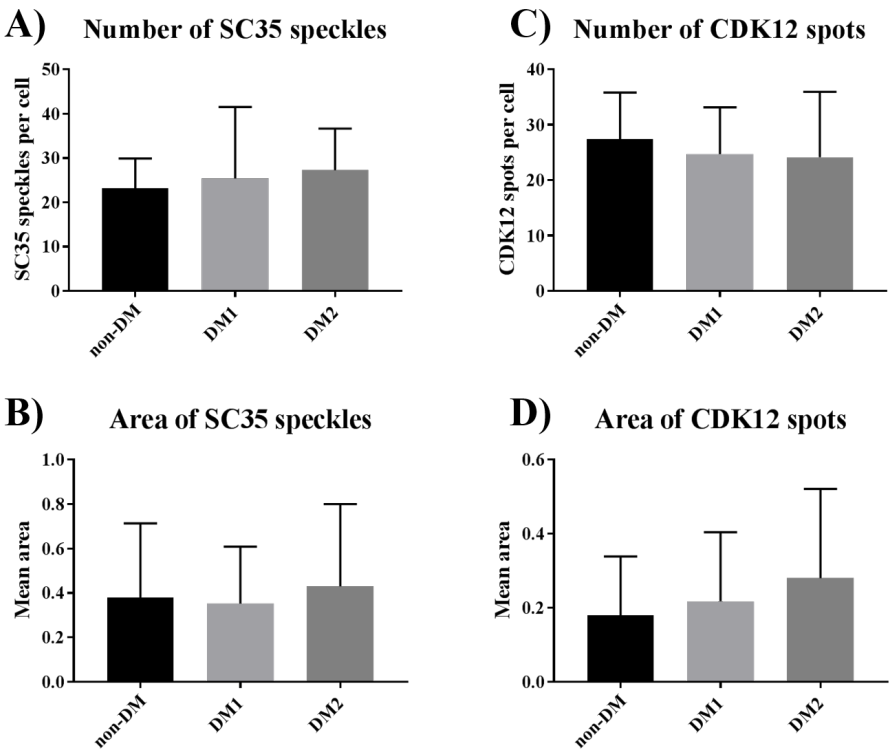


Figure 3.8: **Comparison of the number and area of SC35 speckles and CDK12 spots between non-DM, DM1 and DM2 cells.** An anti-SC35 antibody and an anti-CDK12 antibody were used to detect SC35 speckles and CDK12 spots. The number of spots per nucleus and their area were quantified from the same 20 cells per cell line previously analysed to obtain the PCC's values. Lines above each bar represent the standard deviation. ANOVA tests revealed no statistically significant difference between the different cell lines in any case: **A)** p=0.695; **B)** p=0.8058; **C)** p=0.0.6777; **D)** p=0.4537.

3.4 Effect of the kinase inhibitors on repeat expansion foci

In previous work from Ketley *et al.* (unpublished), treatment of DM1 cells with six kinase inhibitors produced a reduction in the number of foci with some of the compounds. CDK12 was identified to be the main target for this activity. CDK12 has been shown to co-locate with SC35 speckles, and DM1 foci also associate with these speckles. However, as previously shown, DM2 foci do not associate with SC35 speckles. This difference between the two cell lines is particularly im-

portant, as the kinase inhibitors target CDK12, and hence, to compare the effects of the inhibitors in both cell lines could give an insight of the role of CDK12 in myotonic dystrophy. So if CDK12 chemical inhibition reduces foci because CDK12 is involved in the maintenance of foci, the compounds shouldn't eliminate foci in DM2, as they don't co-localize with CDK12. If, on the contrary, the inhibitors eliminate foci in the DM2 cell line as well, it may mean that inhibition of CDK12 is causing transcriptional inhibition of the mutant transcript.

To test the previous hypotheses, DM2 cells were seeded onto 96-well plates and treated with six kinase inhibitors (Table 1.3) for 24 hours. To analyse whether the compounds reduce foci in the nucleus of the cells, an *in situ* hybridization was performed using a $(CAGG)_8$ probe to detect foci. Figure 3.9 shows the effect of each of the kinase inhibitors on foci. Four of the kinase inhibitors, Dinaciclib, SNS-032, AT7519 and PHA767491, significantly reduced the number of foci in the nucleus of DM2 cells at the highest concentrations (Figure 3.9 **A) - D)**). The other two compounds, PD0332991 and CDK9 inhibitor II had no effect on foci (Figure 3.9 **E) - F)**). The apparent increase in the number of foci caused by the highest concentrations of PD0332991 (Figure 3.9 **E)**) was due to high toxicity of the compound, that causes cell death and high background. In this case the software used to quantify foci can pick up as foci other intense signals coming from cell debris or background.

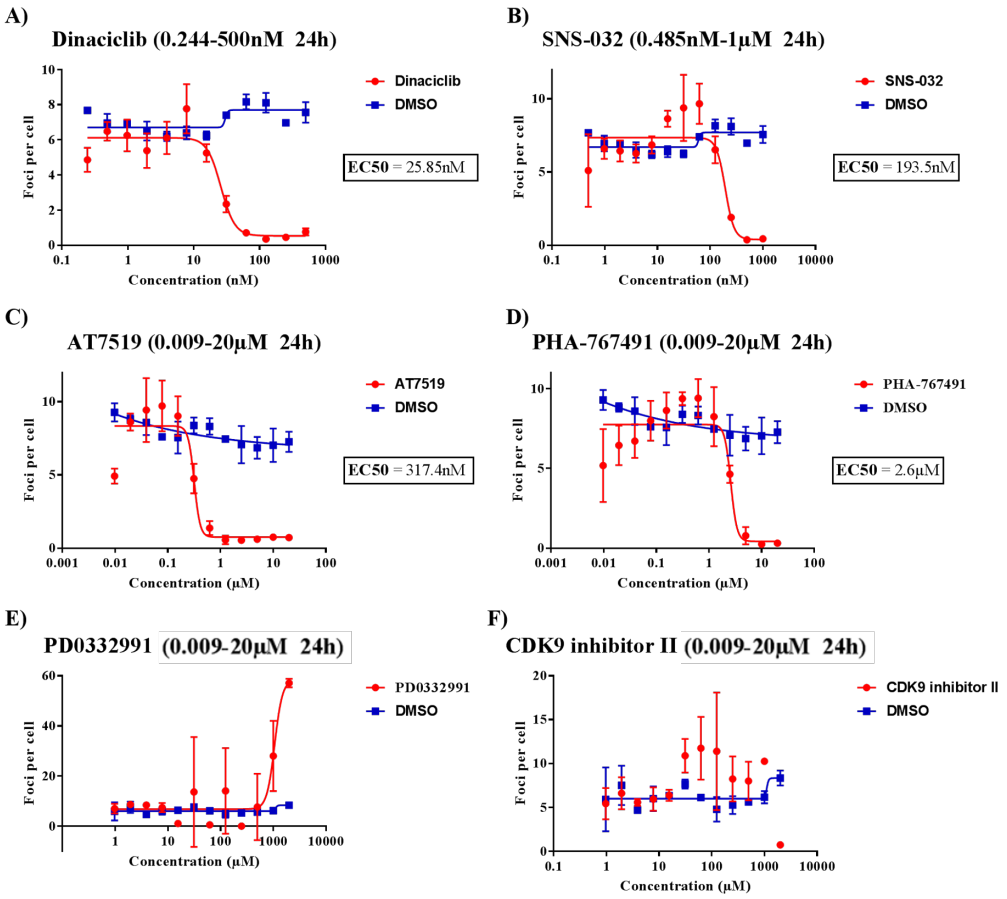


Figure 3.9: **Foci reduction in the DM2 cell line (KAGO Telo) following treatment with 6 kinase inhibitors.** On the X axis are plotted the concentrations in a log scale of compound used to treat the cells. On the Y axis is plotted the number of foci in the cell's nucleus after treatment. Graphs represent a dose-response curve fitted with non-linear regression. Red lines correspond to results from cells treated with the compounds, whereas blue lines correspond to results from cells treated with DMSO as a control. **A)** Cells treated with Dinaciclib for 24 hours, at a range of concentrations from 500 nM to 242.5 pM. **B)** Cells treated with SNS-032 for 24 hours, at a range of concentrations from 1 μM to 485 pM. Cells treated with **C)** AT7519, **D)** PHA-767491, **E)** PD0332991 and **F)** CDK9 inhibitor II for 24 hours, at a range of concentrations from 20 μM to 9.75 nM.

Figure 3.10 shows microscope images of DM2 cells treated with Dinaciclib **(A)** and SNS-032 **(B)** at different concentrations. The highest concentration **(A1, B1)** eliminated almost all foci. The images **(A3, B3)** in Figure 3.10 correspond to a low concentration at which the compounds do not eliminate foci, and the images **(A2, B2)** in Figure 3.10 correspond to concentrations close to the EC_{50} value. In the images of cells treated with concentrations above the EC_{50}

value, cells show virtually no foci in the nucleus.

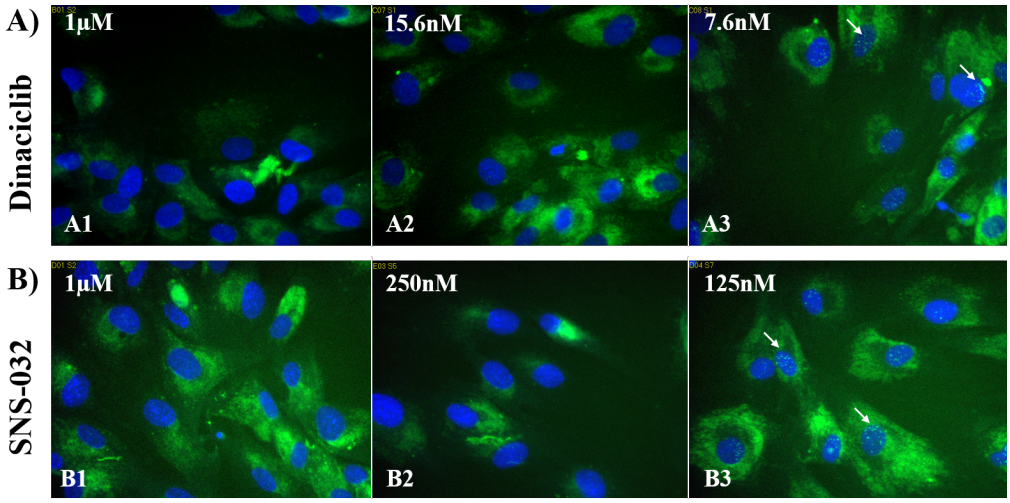


Figure 3.10: ***In situ* hybridization images of DM2 cells (KAGO Telo) treated with Dinaciclib and SNS-032.** Fluorescence images of DM2 cells treated with Dinaciclib **A)** and SNS-032 **B)** for 24 hours. The nucleus of the cell is stained in blue and foci can be seen as tiny green spots in the nucleus. Arrows pointing to foci. Concentrations shown correspond to a high concentration that completely eliminates foci (**A1** and **B1**), the threshold of concentration at which the chemical compounds start being effective (**A2** and **B2**) and a low concentration at which compounds are not able to eliminate foci (**A3** and **B3**).

Dinaciclib and SNS-032 were the most potent inhibitors, as their EC_{50} values were the lowest ones, and eliminated virtually all foci at the highest concentrations (Figure 3.10 **(A1, B1)**). Hence, they were selected to perform the downstream experiments.

3.5 Assessment of the toxicity of the compounds

One of the concerns with the chemical compounds is their toxicity for the cells. If a compound is very potent and selective, but causes cell death, it won't be a suitable drug. Furthermore, in order to see if the foci reduction was due to the effect of the compounds or to a reduction in the number of cells, the toxicity of Dinaciclib and SNS-032 was analysed performing an MTT assay.

DM2 cells were treated for 45 hours with Dinaciclib and SNS-032 and the

MTT assay was performed as described in the Methods section. The absorbance readings were obtained with a spectrophotometer and the mean number of living cells treated with each compound were compared against the mean number of living cells in untreated wells. Results were plotted in a graph (Figure 3.11), representing the percentage of living cells compared to untreated cells, at different compound concentrations.

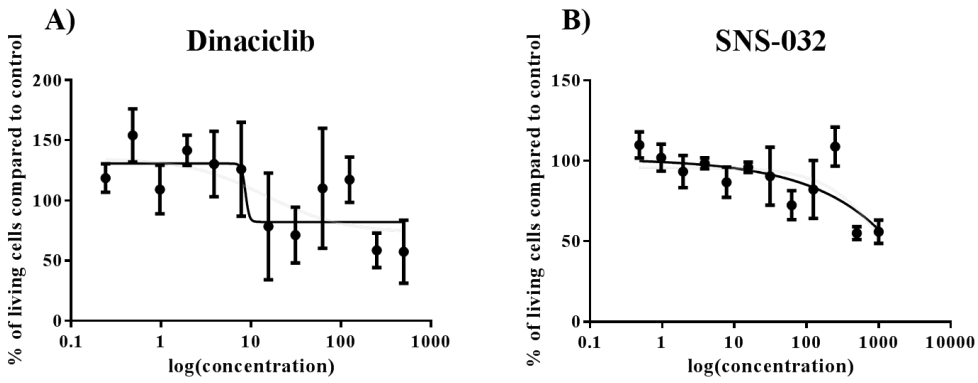


Figure 3.11: **MTT assay for Dinaciclib and SNS-032.** Performed after treating DM2 cells with Dinaciclib **A)** and SNS-032 **B)** for 45 hours. On the X axis are plotted the concentrations in a log scale of compound used to treat the cells. On the Y axis is plotted the percentage of living cells after treatment, compared to untreated cells. The curves were fitted with a non-linear regression. The range of concentrations of Dinaciclib was from 500 nM to 242.5 pM. The range of concentrations of SNS-032 was from 1 μM to 485 pM.

As can be seen in Figure 3.11, the toxicity of Dinaciclib and SNS-032 after 45 hours of treatment is low, being minimal at the lowest concentrations. These results make these inhibitors appropriate for cell culture treatment and for further drug development.

3.6 Effect of the kinase inhibitors on SC35 speckles

The kinase inhibitors used in this work target CDK12, which co-locates with SC35 speckles. These are so-called splicing speckles involved in the processing of many pre- and mRNAs. Thus it is possible that treatment of the cells with CDK12 inhibitors could affect SC35 speckles and thus the processing of other

mRNAs, causing negative effects on the cells. As the kinase inhibitors are as effective in DM2 cells as previously shown in DM1 cells by others in the laboratory (data not shown), both cell lines were treated for 24 hours with Dinaciclib and SNS-032 and the number of SC35 speckles per cell was assessed using an anti-SC35 antibody. Non-DM cells were also treated with the kinase inhibitors. An additional longer treatment of 48 hours was also performed in DM1 cells to rule out any change produced due to a longer exposure to the compounds.

Figure 3.12 shows the reduction produced in the number of SC35 speckles per cell by Dinaciclib and SNS-032 in DM1 cells (**A) and B)**), DM2 cells (**C) and D)**) and non-DM cells (**E) and F)**). The effect produced by both kinase inhibitors is the same in the two DM cell lines, indeed affecting SC35 speckles (Figure 3.12 **A)-D)**). Nevertheless, the compounds did not completely eliminate all SC35 speckles from the cells, the highest concentrations of the compounds caused a reduction of the number of SC35 speckles of around 50%. The inclusion of non-DM cells in this experiment allowed the observation that the effect of the kinase inhibitors appears to be DM-specific, as treatment of non-DM cells didn't produce a reduction in the number of SC35 speckles (Figure 3.12 **E)-F)**). It is worth mentioning that SC35 speckles are dynamic structures, thus a big variability can be seen in the graphs, and also in the error bars, which represent the standard deviation.

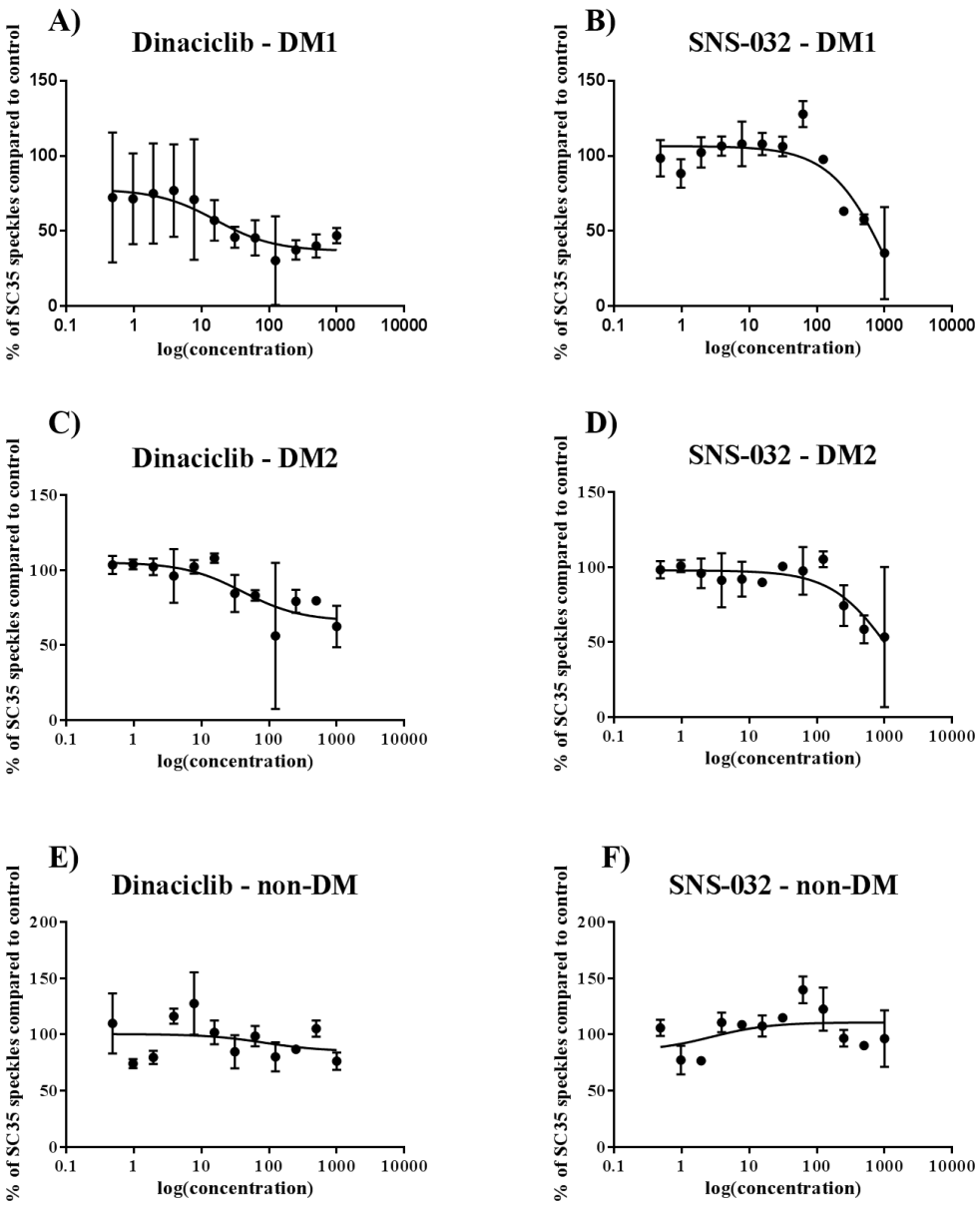


Figure 3.12: **Quantification of SC35 speckles in DM1, DM2 and non-DM cells after treatment with Dinaciclib and SNS-032 for 24 hours.** Plotted on the graph is the percentage of SC35 spots per cell in DM1 (**A) and B)**), DM2 (**C) and D)**) and non-DM cells (**E) and F)**) after treatment with Dinaciclib and SNS-032 for 24 hours, compared to untreated cells. Curves were fitted with a non-linear regression. The range of concentrations for both compounds was from 1 μM to 485 pM.

A longer treatment of DM1 cells with Dinaciclib and SNS-032 didn't change the way the kinase inhibitors affect SC35 speckles, as the same pattern of reduction in the numbers of SC35 speckles was observed after a treatment of 48 hours (Figure 3.13), compared to a treatment of 24 hours.

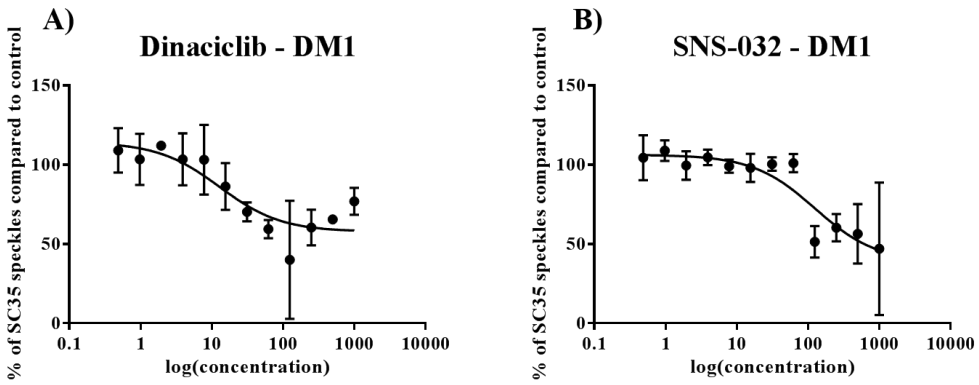


Figure 3.13: **Quantification of SC35 speckles in DM1 cells after treatment with Dinaciclib and SNS-032 for 48 hours.** Plotted on the graph is the percentage of SC35 speckles per cell in DM1 cells after Dinaciclib **A)** and SNS-032 **B)** treatment for 48 hours, compared to untreated cells. Curves were fitted with a non-linear regression. The range of concentrations for both compounds was from 1 μM to 485 pM.

3.7 Effect of the kinase inhibitors on MBNL1

In myotonic dystrophy, MBNL1 proteins are sequestered in the nucleus by the repeat expansion transcripts, where they form inclusions called foci. One recent idea is that MBNL1 is not so much sequestered by the toxic RNA, but rather plays a protective role in the aggregation of the mutant transcript, preventing its export to the cytoplasm and its translation into toxic proteins. When ICC is performed on DM cells with an anti-MBNL1 antibody, the protein appears as small spots in the cell nucleus, which co-locate with foci, in both DM1 and DM2 cells.

In order to determine whether Dinaciclib and SNS-032 affect the cellular localization of MBNL1, DM1 and DM2 cells were treated with the compounds for 24 hours and an ICC performed to detect MBNL1. As a control, cells were treated with DMSO. Figure 3.14 shows a strong nuclear staining for MBNL1 in the absence of the kinase inhibitors in DM1 and DM2 cells treated with DMSO. From the original images in the microscope software, tiny spots can be seen in the nucleus of the cells when zoomed in, which correspond to the MBNL1 sequestered in foci. However, these can not be appreciated here, but the attention

will be focused on the nuclear staining.

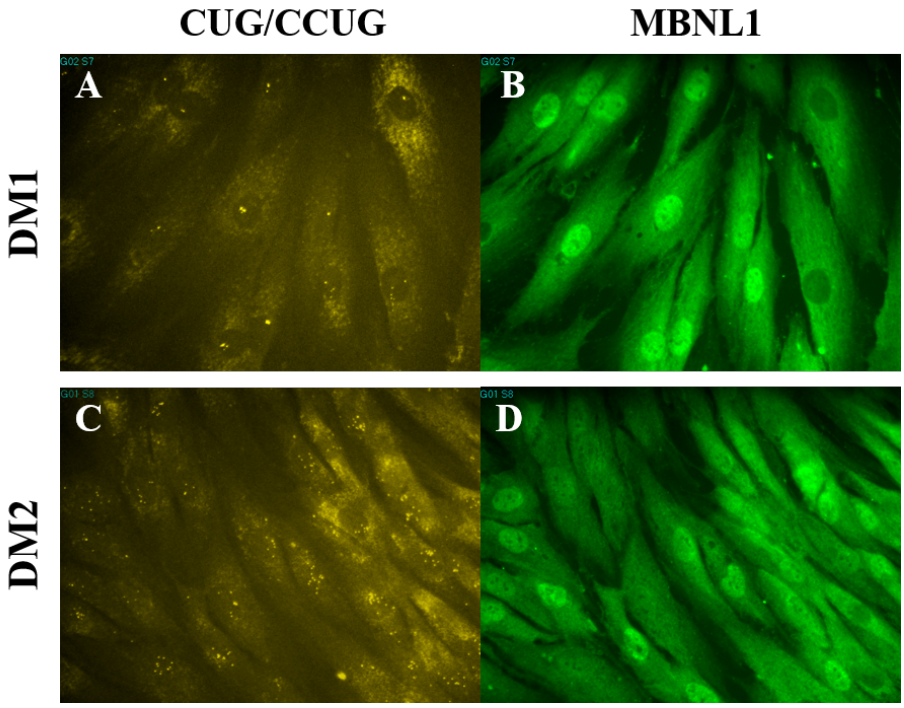


Figure 3.14: **Fluorescence images of DM1 and DM2 cells treated with DMSO for 24 hours.** Foci detection by FISH **(A, C)** and MBNL1 staining **(B, D)** in DM1 and DM2 cells treated with DMSO, used as a control.

Treatment of both DM1 and DM2 cells with Dinaciclib and SNS-032 produced a change in the staining of MBNL1 at the concentrations that are effective for foci reduction (Figures 3.15 and 3.16). At these concentrations the nucleus of the cells didn't show the strong MBNL1 nuclear staining observed at lower concentrations of the inhibitors or in DM cells treated with DMSO. In contrast, the signal from the MBNL1 antibody is much lower in the nucleus of the cells.

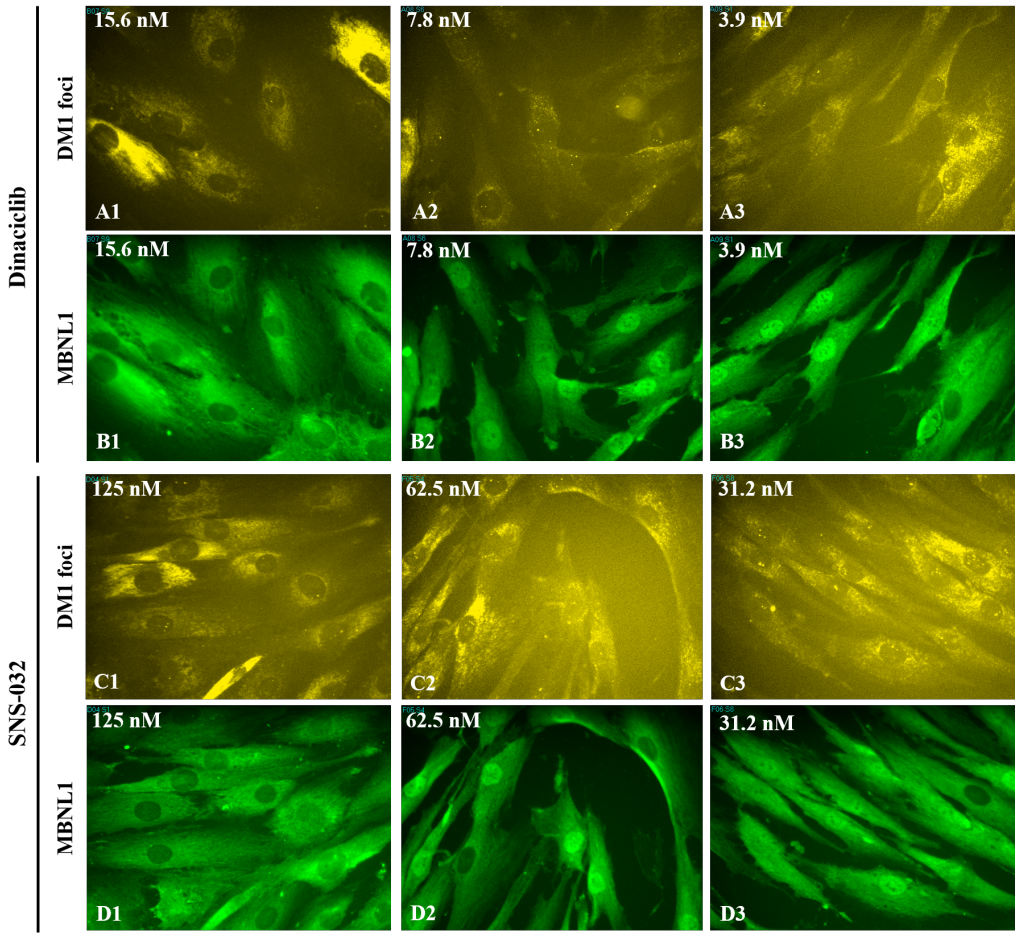


Figure 3.15: **FISH images and MBNL1 ICC after treatment of DM1 cells with Dinaciclib and SNS-032 for 24 hours.** Cells hybridized with the CAG probe are coloured yellow (**A1 - A3, C1 - C3**) and those treated with MBNL1 antibody are coloured green (**B1 - B3, D1 - D3**).

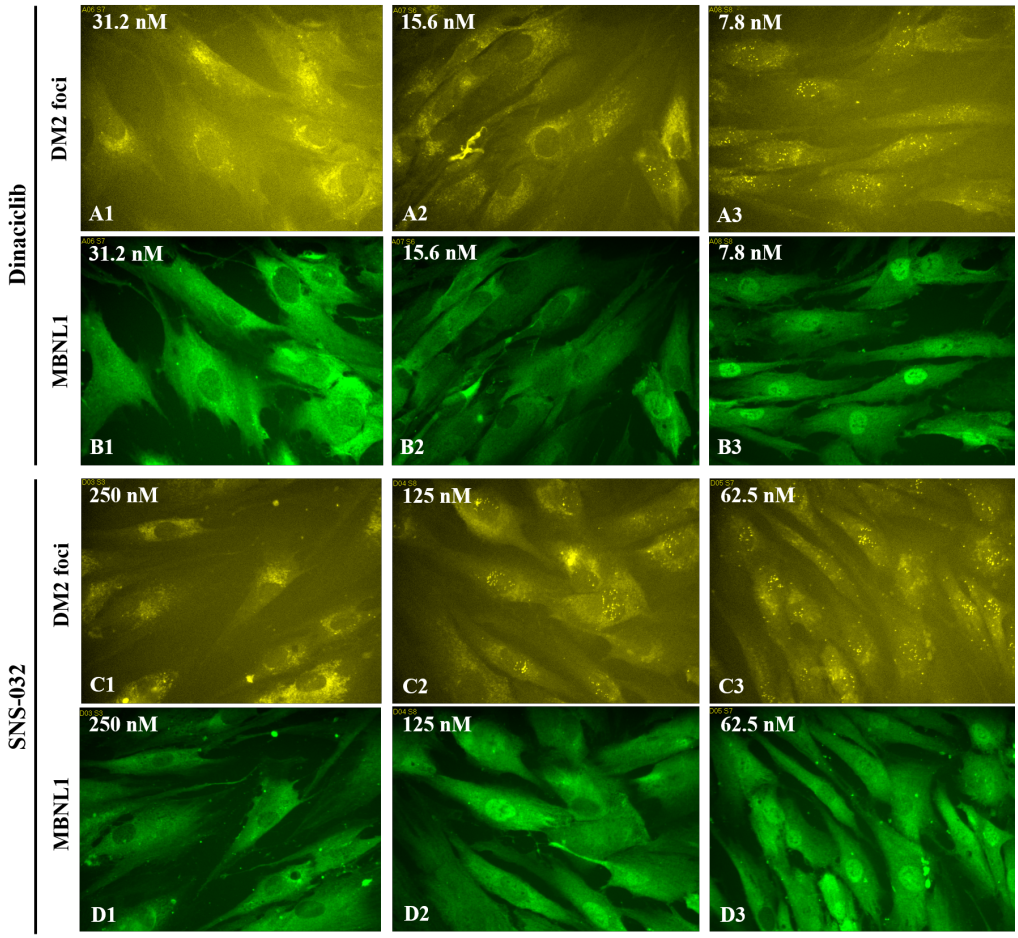


Figure 3.16: **FISH images and MBNL1 ICC after treatment of DM2 cells with Dinaciclib and SNS-032 for 24 hours.** Cells hybridized with the CAGG probe are coloured yellow (**A1 - A3, C1 - C3**) and those treated with MBNL1 antibody are coloured green (**B1 - B3, D1 - D3**).

The changes observed in the distribution of MBNL1 in cells treated with Dinaciclib and SNS-032 could have different causes. For example, inhibition of CDK12 could cause global transcription arrest, resulting in the production of less MBNL1. Alternatively, CDK12 could play a role in MBNL1 phosphorylation and thus in its cellular-nuclear localization. In this case, inhibition of CDK12 could alter the MBNL1 localization such that the MBNL1 previously sequestered in the nucleus is displaced and the repeat expansion transcripts are released for degradation. In order to see whether CDK12 inhibition has an effect on MBNL1 production, DM1 and DM2 cells were treated with 5 μM of Dinaciclib and SNS-032 for 24 hours, after which total proteins were extracted and the levels of MBNL1 analysed by western blot.

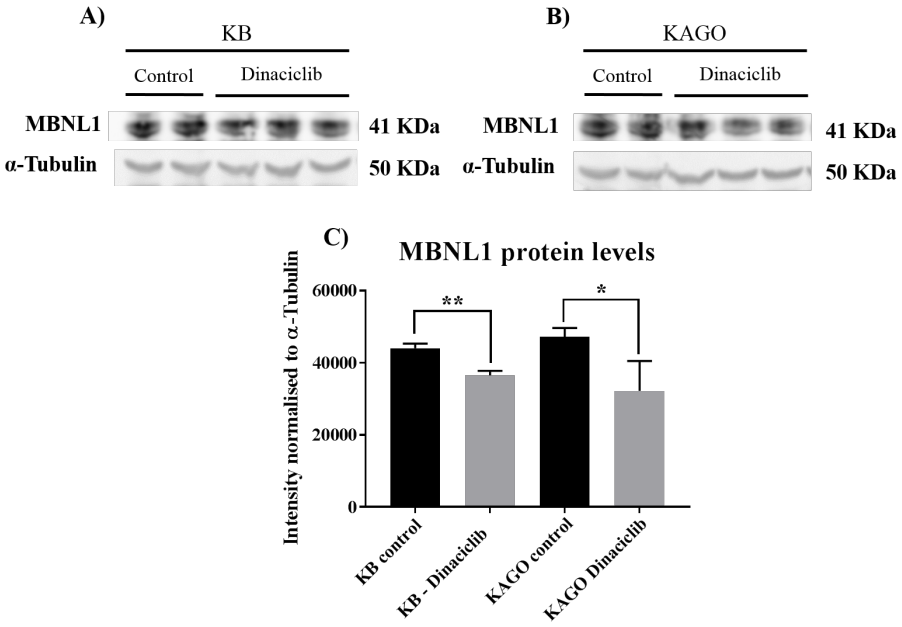


Figure 3.17: **Effect of Dinaciclib on MBNL1 protein expression in DM1 (KB) and DM2 (KAGO) cells.** Western blot for MBNL1 of protein extracted from DM1 **(A)** and DM2 **(B)** cells untreated and treated with 5 μM of Dinaciclib. α-Tubulin was used as loading control. **(C)** Histogram to quantify the levels of MBNL1, normalised to α-Tubulin. T-tests for each cell line comparing control cells with Dinaciclib-treated cells, revealed a statistically significant reduction of the MBNL1 protein levels in both DM1 and DM2 cells treated with Dinaciclib. **p=0.0019 (DM1); *p=0.0393 (DM2). n=3.

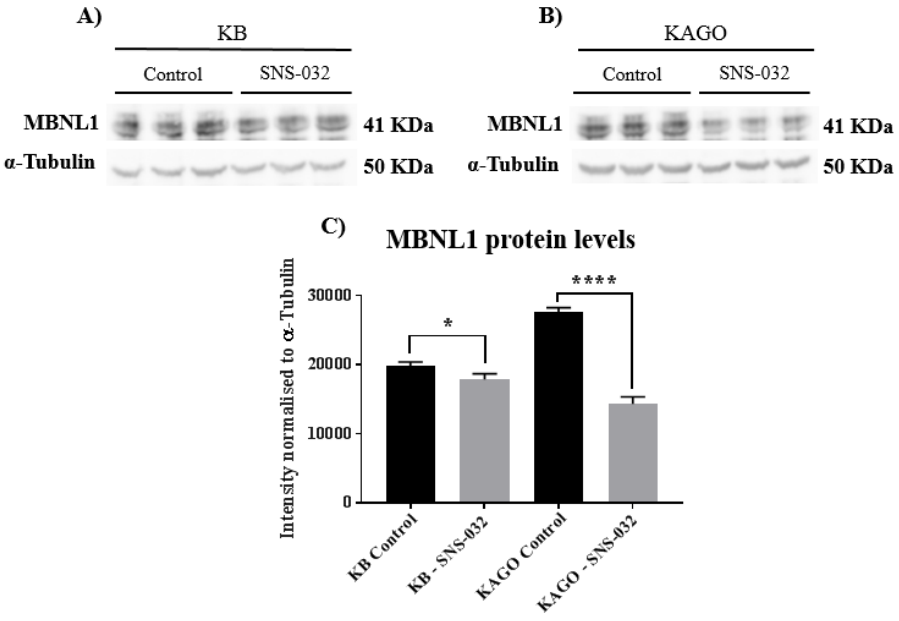


Figure 3.18: **Effect of SNS-032 on MBNL1 protein expression in DM1 (KB) and DM2 (KAGO) cells.** Western blot for MBNL1 of protein extracted from DM1 **(A)** and DM2 **(B)** cells untreated and treated with 5 μM of SNS-032. α-Tubulin was used as loading control. **(C)** Histogram to quantify the levels of MBNL1, normalised to α-Tubulin. T-tests for each cell line comparing control cells with SNS-032-treated cells, revealed a statistically significant reduction of the MBNL1 protein levels in both DM1 and DM2 cells treated with SNS-032. *p=0.0249 (DM1); ****p<0.0001 (DM2). n=3.

The histograms in Figures 3.17 and 3.18 show a statistically significant reduction of the levels of MBNL1 protein in both DM1 and DM2 cells following treatment with Dinaciclib and SNS-032. It is worth mentioning that the concentration used to treat the cells is much higher than the EC_{50} of the compounds to reduce nuclear foci, hence, conducting the same western blot experiments with concentrations of the kinase inhibitors around the EC_{50} would be informative, as it appears MBNL1 leaves the nucleus at the same concentration threshold that foci disappear (Figures 3.15 and 3.16). Taken all together, the previous results prove that inhibition of CDK12 by the kinase inhibitors affects MBNL1 production and subcellular distribution.

3.8 Effect of the kinase inhibitors on Alternative Splicing

The DM pathogenesis model suggests that MBNL1 is sequestered in foci, leading to alternative splicing impairment and mis-regulation of many genes. Previous experiments have shown that release of MBNL1 from nuclear foci can restore the alternative splicing of some genes (Wheeler et al., 2009; Warf et al., 2009b; Garcia-Lopez et al., 2011; Chakraborty et al., 2018). In work presented thus far, Dinaciclib and SNS-032 have proven to be effective for foci elimination and MBNL1 release from foci. In this section, DM2 cells were treated with Dinaciclib at a concentration of 5 μM for 24 hours and then RNA was extracted in order to analyse the alternative splicing of different genes. Seven genes misspliced in DM according to the literature were selected for the analysis: *INSR*, *PHKA1*, *SERCA*, *NCOR*, *MBNL1*, *NFIX* and *ITGA6*. Not all of the misspliced genes listed in the literature can be found in any DM model with the same alternative splicing impairment. Figures 3.19 - 3.22 show the pattern of alternative splicing of the selected genes, with the major isoform present in the non-DM population marked in bold.

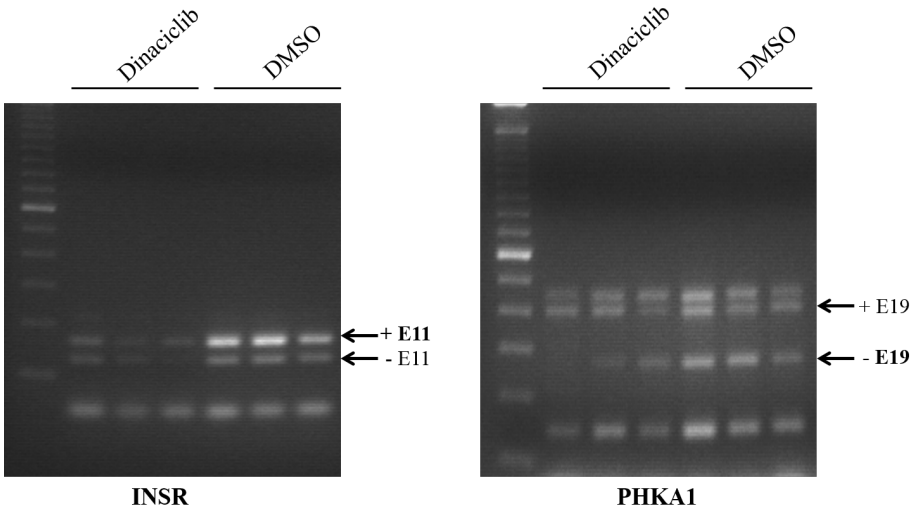


Figure 3.19: **Splicing analysis of *INSR* and *PHKA1* genes after treatment of DM2 cells with Dinaciclib.** The RT-PCR products corresponding to the isoforms for the exons included and excluded are shown. The major isoform is marked in bold. Both *INSR* and *PHKA1* genes present higher amounts of the major isoform of splicing in DM2 cells. n=3 for each treatment.

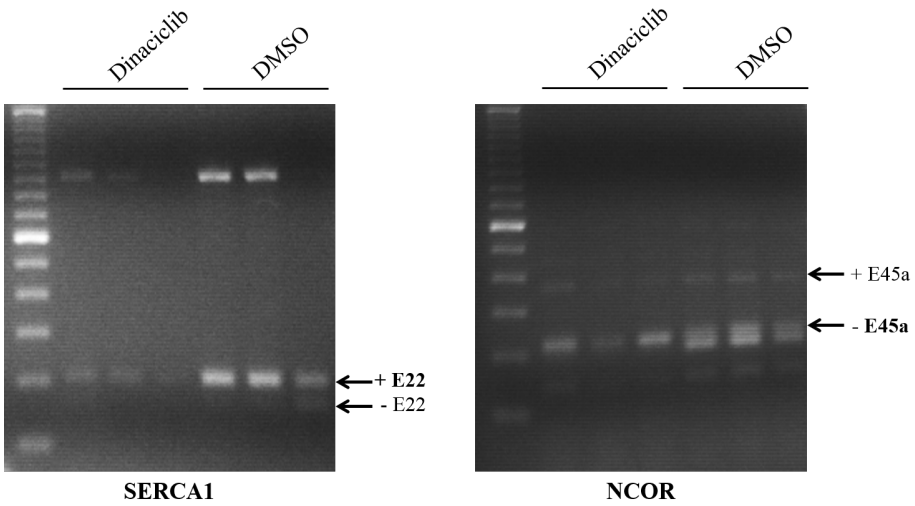


Figure 3.20: **Splicing analysis of *SERCA1* and *NCOR* genes after treatment of DM2 cells with Dinaciclib.** The RT-PCR products corresponding to the isoforms for the exons included and excluded are shown. The major isoform is marked in bold. Both *SERCA1* and *NCOR* present higher amounts of the major isoform of splicing in DM2 cells. n=3 for each treatment.

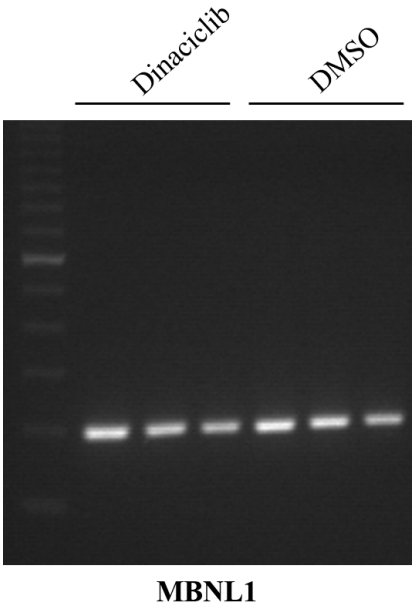


Figure 3.21: **Splicing analysis of the *MBNL1* gene after treatment of DM2 cells with Dinaciclib.** The RT-PCR products corresponding to MBNL1 isoform are shown. *MBNL1* only shows one isoform in DM2 cells. n=3 for each treatment.

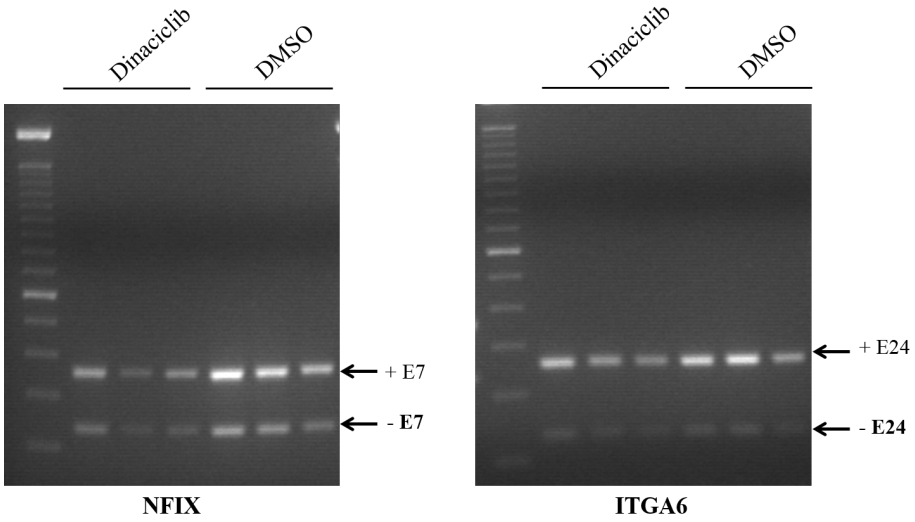


Figure 3.22: **Splicing analysis of *NFIX* and *ITGA6* genes after treatment of DM2 cells with Dinaciclib.** The RT-PCR products corresponding to the isoforms for the exons included and excluded are shown. The major isoform is marked in bold. Both *NFIX* and *ITGA6* show higher amounts of the minor isoform of splicing in DM2 cells. n=3 for each treatment.

Out of the 7 genes analysed, *MBNL1* showed just 1 isoform, so it couldn't be used to analyse the alternative splicing restoration. The ratio of the isoforms of the other genes was analysed in the DMSO-treated samples, in order to look for genes with an inverted ratio of the isoforms compared to the non-DM population. Figure 3.23 shows the ratio of the isoforms of the genes analysed in this chapter, with the bar corresponding to the major isoform being black and the bar corresponding to the minor isoform being red.

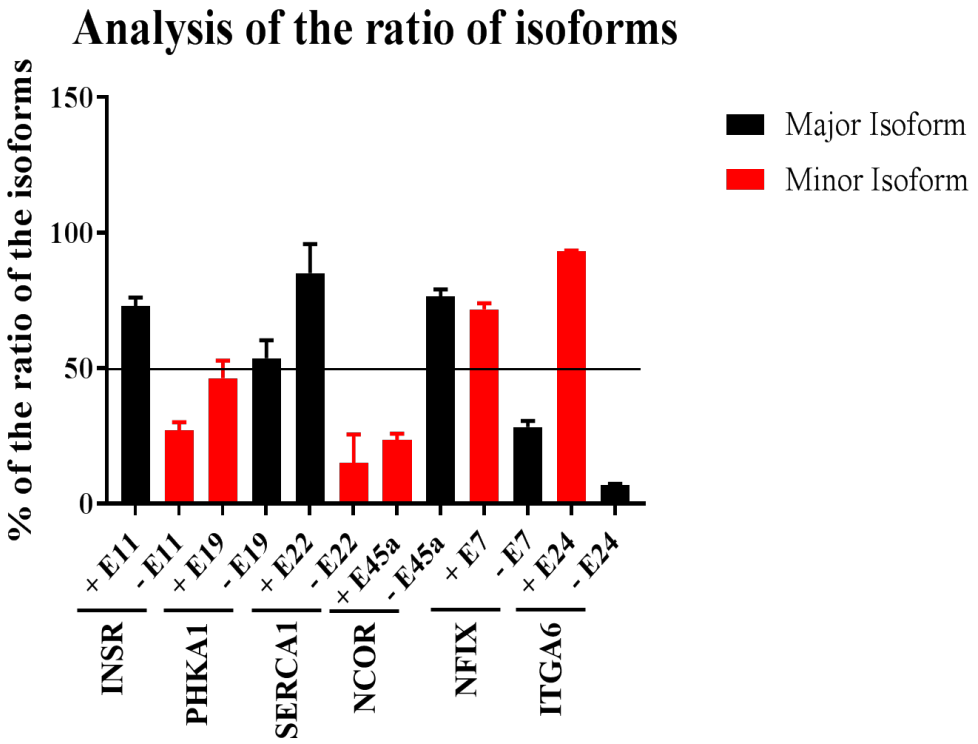


Figure 3.23: **Analysis of the ratio of the isoforms of *INSR*, *PHKA1*, *SERCA1*, *NCOR*, *NFIX* and *ITGA6* in DM2 cells treated with DMSO.** RT-PCR products of the genes mentioned were amplified and run on a gel. The bands were analysed using ImageJ and the graph shows the percentage of the ratio of the two isoforms for each of the genes. The major isoform in the non-DM population is represented by a black bar, whereas the minor isoform is represented by a red bar. A line across the graph marks the 50% of the ratio. Genes with the red bar above 50% such as *NFIX* and *ITGA6* show a pattern of isoforms characteristic of the DM patients.

In Figure 3.23 we can see that *INSR*, *PHKA1*, *SERCA1* and *NCOR* present a higher ratio of the major isoform, usually found in the non-DM population. Only *NFIX* and *ITGA6* presented a higher ratio of the minor isoform, as their red bar is above 50% in the graph. This pattern of splicing is reversed to what is common in the non-DM population, and has been reported to be characteristic of myotonic dystrophy. Nevertheless, it is important to mention that there does not need to be a completely switch in splice isoform to indicate pathology, so the effect of Dinaciclib in the alternative splicing of all the genes that showed two isoforms was studied.

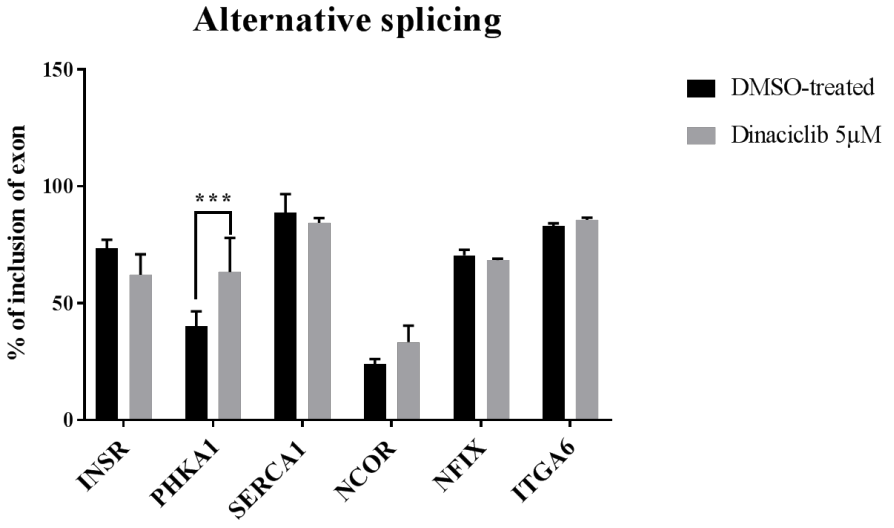


Figure 3.24: **Analysis of the alternative splicing of several genes in DM2 cells treated with 5 μM Dinaciclib and DMSO.** Data is shown as the percentage of inclusion of exon 11 of *INSR*, exon 19 of *PHKA1*, exon 22 of *SERCA1*, exon 45a of *NCOR*, exon 7 of *NFIX* and exon 24 of *ITGA6* in DM2 cells, comparing cells treated with DMSO or with Dinaciclib at a concentration of 5 μM. There is no significant difference between DMSO-treated samples and Dinaciclib-treated samples for all the genes, except for *PHKA1*, in which p value was p=0.0009. Nonetheless, inclusion of exon 19 of *PHKA1* is characteristic of the DM-population, so it appears to be a worsening effect of the alternative splicing instead of a rescue.

The graph from Figure 3.24 shows the percentage of inclusion of the exons analysed for each of the selected genes. Dinaciclib-treated bars in the case of *INSR*, *SERCA1*, *NFIX* and *ITGA6* above the DMSO-treated bars would indicate that Dinaciclib partially rescues the alternative splicing of these genes. On the contrary, in the cases of *PHKA1* and *NCOR*, Dinaciclib-treated bars below DMSO-treated bars, would also suggest a partial rescue of the alternative splicing. Nonetheless, as can be seen in the graph, and after running an ANOVA test, the differences between DMSO and Dinaciclib-treated cells regarding the alternative splicing are not statistically significant, except in the case of *PHKA1*. However, in this case, the alternative splicing appears to have been worsened, as inclusion of the exon 19 is characteristic of the DM-population. These results suggest that Dinaciclib was not able to rescue the aberrant splicing of the selected genes in this particular cell line.

3.9 Effect of newly synthesized kinase inhibitors in DM2 cells

An ongoing project in our laboratory consists of the development of new small molecules that specifically inhibit CDK12. The new compounds have been generated by modifying the structure of different kinase inhibitors in order to increase the selectivity towards CDK12 and the cell permeability. Some of the compounds show promising results regarding foci reduction in DM1 cells, and the best ones are currently being tested in the HSA^{LR} mouse. Four of the most effective compounds that reduced foci in DM1 cells were tested here in DM2 cells.

KAGO Telo cells seeded onto 96-well plates were treated with compounds ADS158081, ADS158784, ADS160181 and ADS160919 (the structure of the compounds can not be shown due to patent issues) for 48 hours and then the number of foci in the nucleus of the cells was analysed by performing an *in situ* hybridisation. Figure 3.25 shows the effect of these compounds on the number of foci in the nucleus of the cells.

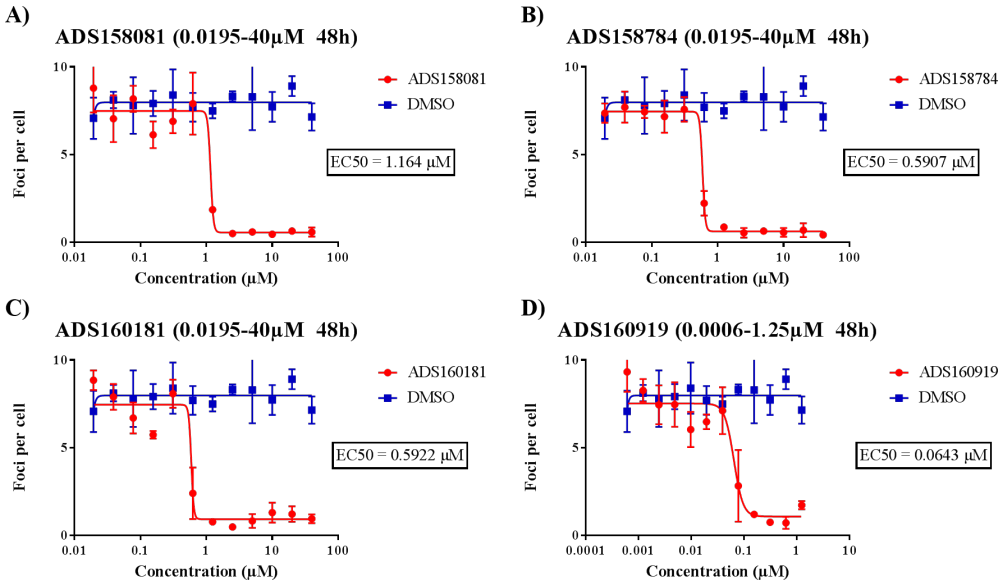


Figure 3.25: **Foci reduction in KAGO Telo (DM2) cells following treatment with newly synthesized kinase inhibitors.** The concentrations in a log scale of compounds used to treat the cells are plotted on the X-axis. The number of foci per nucleus are plotted on the Y-axis. The graphs represent a dose-response curve fitted with non-linear regression. Red lines correspond to results from cells treated with the compounds, whereas blue lines correspond to results from cells treated with DMSO control. **A), B) and C)** Cells treated with ADS158081, ADS158784 and ADS160181, respectively, for 48 hours at a range of concentrations from 40 μM to 0.0195 μM. **D)** Cells treated with ADS160919 for 48 hours at a range of concentrations from 1.25 μM to 0.6 nM.

The four compounds effectively reduced the number of foci in DM2 cells, and the EC_{50} values are very similar to those obtained in DM1 cells by others in the laboratory (data not shown). Even though Dinaciclib did not restore the alternative splicing of the genes tested in DM2 cells, this and similar kinase inhibitors can serve as a starting point to develop a suitable therapy for the disease. Indeed, so far the new molecules generated are producing promising effects in both DM1 and DM2 cells. Furthermore, the results in DM2 cells support the idea that inhibition of CDK12 produces the elimination of foci by inhibition of the transcription of the mutant transcript, as CDK12 is required for the elongation phase of transcription.

3.10 Summary

Myotonic dystrophy type 1 and 2 share a pathomechanism and some similarities. Nevertheless there are differences between the two which require them to be classified as different diseases. Some of these differences include a distal muscle weakness in DM1 patients, whereas DM2 patients show proximal muscle weakness. Myotonia is more severe in DM1 than in DM2, as is the cardiac involvement. There is no congenital form in DM2 and no anticipation effect either. However, the cause of the differences between DM1 and DM2 remains unclear. In this chapter my first objective was to study the localization of foci in the nucleus of DM1 and DM2 cells and examine their relationship with other proteins implicated in the pathogenesis of the disease. Foci are a hallmark of myotonic dystrophy and other microsatellite expansion disorders, and it has been demonstrated that foci associate with different structures in the nucleus of DM1 and DM2 cells.

In this chapter the relationship between repeat expansion foci, SC35 speckles and the CDK12 protein was analysed. In DM1 cells, foci strongly associate with SC35 speckles, and also with CDK12 proteins, which are located within the SC35 speckles. However, in DM2 cells, the majority of foci are free in the nucleus of the cells, and do not co-localize with either SC35 speckles or CDK12. Despite the difficulties to visualize both foci and CDK12 in the cells, a suitable protocol for combining CDK12 ICC with FISH in DM1 cells allowed the direct analysis of the co-localization of DM1 foci with CDK12. Direct analysis of the degree of co-localization of DM2 foci with CDK12 could not be established due to technical difficulties when combining ICC and FISH in this cell model, as DM2 cells appeared to be more sensitive to the chemistry of the experiment, which affects the mutant RNA or the CAGG probe and thus foci could not be detected. Results presented here from the analysis of SC35 speckles and CDK12 in non-DM, DM1 and DM2 cells showed a significant difference in the degree of co-localization between these proteins in each of the cell lines. In DM1 and DM2 cells, SC35 speckles and CDK12 co-localize more often than in non-DM cells. Thus, taken together, and by inference, these results suggest

that repeat expansion foci in DM2 do not co-localize with CDK12. This is particularly important as CDK12 inhibitors affect repeat expansion foci similarly in both DM1 and DM2 cells, suggesting their activity is not a consequence of their nuclear localisation.

Realization of the same analysis, regarding co-localization of foci with different proteins in the nucleus, in additional DM1 and DM2 cell lines would be of great benefit, as in this work only one DM1 cell line and one DM2 cell line were used, so the differences observed regarding foci localization in the nucleus of the cell could be specific of these two cell lines. The use of other DM1 and DM2 cell lines could help us clarify if this is a general difference between DM1 and DM2.

Once the location of foci in the nucleus of DM1 and DM2 cells was established, a set of kinase inhibitors that target mainly CDK12 were tested in DM2 cells. Similar to the results obtained previously in DM1 cells by others in the laboratory, four of the compounds (Dinaciclib, SNS-032, AT7519 and PHA-767491) were positive for foci reduction, whereas the other two (PD0332991 and CDK9 inhibitor II) didn't eliminate foci. Dinaciclib and SNS-032, the most potent inhibitors, were selected to perform downstream experiments and analyse other potential beneficial effects of the compounds in both DM1 and DM2 cells.

Dinaciclib and SNS-032 reduced the number of SC35 speckles up to 50% at the highest concentrations in both DM1 and DM2 cell lines. The effect on the speckles appears to be DM-specific as the compounds didn't reduce the number of speckles in non-DM cells. This result suggests that the presence of the mutant transcript is required for the compounds to produce a reduction in SC35 speckles.

The effect of the inhibitors was not limited to repeat expansion foci and SC35 speckles. Dinaciclib and SNS-032 treatment of DM1 and DM2 cells also changed the pattern of staining of the anti-MBNL1 antibody. In the presence of repeat expansion foci, MBNL1 shows strong nuclear staining. However, at the concentrations of the compounds that eliminate foci, the staining of the antibody changes and MBNL1 signal is much lower in the nucleus of the cells. Analysis of the levels of MBNL1 protein by western blot in DM1 and DM2 cells treated

with 5 μM of Dinaciclib and SNS-032 for 24 hours revealed reduced amounts of MBNL1 protein in both cell lines treated with the compounds, compared to controls. Additionally, analysis of the alternative splicing of a set of genes in DM2 cells, didn't show restoration of the aberrant splicing with 5 μM of Dinaciclib. Furthermore, the alternative splicing of PHKA1 appears to have been worsened by Dinaciclib, as its pattern of splicing is shifted towards what is usually found in the DM-population. This effect of the compound on the alternative splicing could be caused by the downregulation of MBNL1, which could be contributing to prevent a restoration of the alternative splicing, or even could worsen the splicing. Further studies are required to analyse if this would be a plausible explanation.

The results obtained in this chapter suggest that inhibition of CDK12 by the kinase inhibitors affects transcription of *DMPK* and *CNBP* as well as the protein levels of SC35 and MBNL1, by a yet to be determined mechanism. The reduction of repeat expansion foci in DM cells following inhibition of mutant transcript production has other downstream consequences that require further investigation.

Finally, even if these kinase inhibitors do not restore the alternative splicing and their mode of action is not yet fully understood, they can serve as a starting point to develop a molecule that eliminates foci and rescues the alternative splicing. Efforts to synthesize new kinase inhibitors are producing promising results. Four of the best compounds generated were tested in DM2 cells producing the same effects on repeat expansion foci reduction as in DM1 cells. These results open the door to a new potential therapy for both DM1 and DM2 and point to CDK12 as a novel therapeutic target.

Chapter 4

CDK12 CRISPR/Cas9-mediated gene knock out in DM1 cells

4.1 Aim

The aim of the work in this chapter was to generate DM1 and DM2 cell lines with stable knock out of *CDK12* to be used as a model system to study myotonic dystrophy and understand the implication of CDK12 in the pathogenesis of the disease.

4.2 Introduction

One of the hallmarks of myotonic dystrophy is the presence of foci in the nucleus of the cells. As explained in the Introduction, these foci are made of aggregations of the mutant mRNA transcripts, that can not be exported to the cytoplasm, and MBNL1 proteins. Several studies trying to eliminate foci have been published, some of which also alleviated some symptoms of the disease (Warf et al., 2009b; Wheeler et al., 2009; Garcia-Lopez et al., 2011; Wheeler et al., 2012; Chakraborty et al., 2018). Some of the technologies used to eliminate foci have proven to be effective, also producing alternative splicing restoration and amelioration of the symptoms. However, these molecules often have associated prob-

lems such as high levels of toxicity and poor oral availability, which make them unsuitable for drug development.

In a recent study from Ketley *et al.* (currently unpublished), CDK12 was identified as the main target of a subset of kinase inhibitors that successfully reduced foci in DM1 cells, and also in DM2 cells as shown in Chapter 3. Inhibition of CDK12 with shRNAs also eliminated foci from the nucleus of DM1 cells. Hence, CDK12 appears to have a role in foci maintenance and its inhibition could be used as a starting point to develop a therapeutic strategy. However, chemical inhibition usually comes associated with off-target effects. Despite the fact that CDK12 is the main target of the kinase inhibitors, they are not entirely selective and they inhibit other CDKs. It is not clear yet whether other CDKs, in addition to CDK12, contribute to the inhibitors effect on foci.

In order to assess the effect of CDK12 knock out on foci and myotonic dystrophy pathogenesis without having to worry about other off-target effects caused by the chemical inhibitors, efforts were made in this chapter to edit the DNA of DM1 and DM2 cells in order to create mutations in *CDK12* that disrupt the synthesis of the protein. Based on the results obtained with *CDK12* chemical inhibition, my hypothesis was that genomic knock out of *CDK12* should eliminate foci from DM cells.

Mutations in *CDK12* were created using the CRISPR/Cas9 genome editing technology. A gRNA with high activity *in vitro* targeting exon 1 of *CDK12* was selected to perform the experiments, and hygromycin resistant clones were expanded and screened for *CDK12* knock out. Exon 1 was selected as the target site with the aim to completely disrupt the production of the whole protein. Sometimes, mutations in the middle or at the end of the coding region of a gene can lead to the production of a truncated protein shorter than the original. Nonetheless, a gene with mutations in the exon 1 shouldn't produce any protein.

Initially, the experiment was performed in DM2 cells, KAGO Telo, but after a few rounds of transfection in which the cells failed to form colonies, DM1 cells, KB Telo MyoD, were used. In this case, the cells formed colonies that were later screened for mutations in *CDK12*.

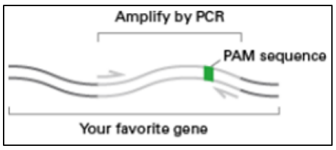
4.3 Analysis of *in vitro* cleavage efficiency of the gRNAs

The exon 1 of *CDK12* was scanned using the CRISPR softwares to predict potential gRNAs. Exon 1 was selected as the target site in order to completely disrupt the production of the protein from the start, and to avoid having truncated proteins of a shorter length. Different isoforms of CDK12 are produced, all from the same ATG start codon present at the beginning of the exon 1 and all have the same N-terminal region, but they differ in length and in the C-terminal region. Hence, the mutations introduced in exon 1 of *CDK12* by the gRNA selected and the Cas9 were expected to disrupt the production of all CDK12 isoforms, and no protein was expected to be present in the cells should the experiment work properly.

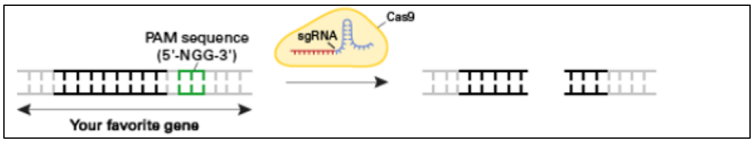
It has been demonstrated that not all gRNAs predicted by a CRISPR software are equally efficient, probably due to differences in the expression of RNAs, secondary structures leading to inefficient Cas9 target recognition or degradation of the gRNA by endogenous RNA interference machinery, so every new gRNA needs to be tested *in vitro* before its transfection into cells. gRNAs that work *in vitro* are more likely to work *in vivo* than those that don't show any activity *in vitro* (Cong et al., 2013). The generation of a human stable cell line takes around 3-4 months, so the *in vitro* assay was conducted in order to save time, money and effort.

The eight gRNAs predicted by the CRISPR Design web tool (sequences shown in Table 2.2 and mapped in the Appendix 2) were incubated in a PCR thermocycler with the Cas9 protein and a PCR fragment harbouring its target site, and the resulting PCR fragments were analysed on a gel. If a gRNA possesses high activity *in vitro*, the Cas9 will be able to cleave the PCR fragment into two fragments of smaller size when complexed with that particular gRNA (Figure 4.1).

A) PCR generated target for cleavage



B) *In vitro* cleavage of target sequence by Cas9 and synthesized gRNA



C) Separation of cleaved products on an agarose gel

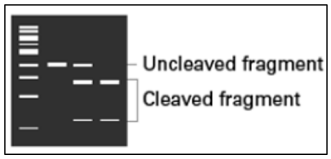


Figure 4.1: **Overview of the *in vitro* assay protocol. A)** A PCR fragment harbouring the gRNA target site is synthesized from genomic DNA. **B)** The PCR fragment is then combined with each potential gRNA and recombinant Cas9. **C)** The reaction is separated by agarose gel electrophoresis and fragments analysed. Cleavage by the Cas9 will result in two or three bands that can be distinguished in the agarose gel.

Figure 4.2 shows good activity *in vitro* for the gRNAs 2 and 8 as the PCR fragment was cut, whereas the Cas9 was unable to cleave the fragment when complexed with any of the other gRNAs. The cleavage of the PCR fragment produced with gRNA 2 was cleaner than the smear caused with gRNA 8, which was likely to be caused by higher amounts of the Cas9 protein present in the sample, which affects the movement of the DNA bands through the gel.

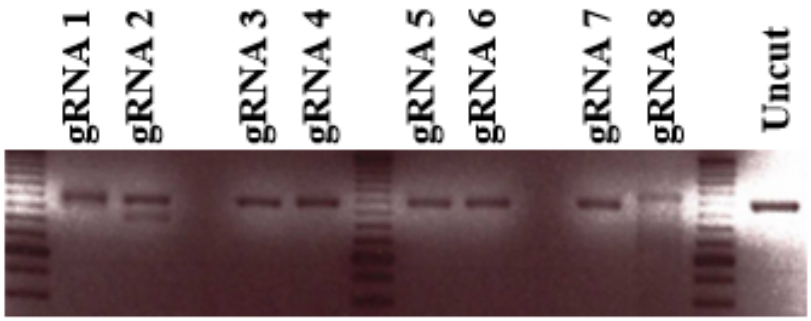


Figure 4.2: ***In vitro* assay of the predicted gRNAs targeting *CDK12***. gRNAs 1, 3, 4, 5, 6 and 7 show a single band of the same size as the uncut band, indicating that no cleavage has occurred in these samples because the gRNAs don't have high activity *in vitro*. However, gRNA 2 shows 2 bands, indicating that the original PCR fragment was cut by the Cas9 protein. gRNA 8 also shows cleavage activity *in vitro*, although the DNA appears degraded with a smear in the lane.

One of the biggest concerns about the CRISPR system is the possibility of off-target effects. Cas9 generally has very few off-targets effects, although the frequency depends on the sequence of the gRNA. In 2013, Hsu *et al.* demonstrated that Cas9 can tolerate mismatches in the sequence, and it retains its cleavage activity unless the mismatches are in the PAM sequence (Hsu et al., 2013). However, Cong *et al.*, reported that single-base mismatches in the 8-bp of the PAM-proximal region abolish all Cas9 activity (Cong et al., 2013).

The CRISPR Design web tool predicted 6 off-target effects for the gRNA 2, all of which are listed in Table 4.1. All the additional sequences found have 4 mismatches and usually, with more than 3 mismatches, the Cas9 doesn't work, so gRNA 2 should be specific to its target. The number of off-target effects was very different between gRNA 2 and gRNA 8, as gRNA 8 has 26 predicted off-target effects, much more than gRNA 2. Thus, based on the quality of the cleavage of the PCR fragment and the lower number of off-target sequences, gRNA 2 was selected as the gRNA to be used for the experiment.

Table 4.1: **Off-target effects predicted for gRNA 2 targeting *CDK12***. The table shows similar sequences to the gRNA 2 target site present in other chromosomal locations, along with their score. The score is calculated by the software based on the number of mismatches and thus it is the probability that the gRNA will bind to that sequence. Each base underlined corresponds to a mismatch with the original gRNA 2 sequence. The sequence in bold corresponds to the PAM site.

Chromosome	Sequence	n_mismatches	Score	Gene
Chr 17	GCTTGGTATCGAAGCACAAG**CGG**	0	100	NM_016507
Chr 4	ACGTGCTATCAAAGCACAAG**CAG**	4	0.481408840561	NM_014247
Chr 19	GCTTGGTGTTGAAGCAGGAG**TGG**	4	0.101056725	NM_019884
Chr 5	GTTTGGTATCAAAGAAAAAG**CAG**	4	0.090669425	NM_001037332
Chr 22	GCTTGTTATGGAGGCACCAG**CAG**	4	0.0635314208497	NM_004286
Chr 3	GCTTGGTAACGCAGTGCAAG**AAG**	4	0.0192084049821	NM_175056

4.4 Transfection of cells and expansion of the clones

The Cas9 protein and the gRNA can be transfected into cells in three different forms. The simplest option is to transfect the cells with plasmids encoding the Cas9 protein and the gRNA. A disadvantage of this type of delivery is that the Cas9 is expressed from the plasmid for a longer time, which could cause off-target effects. The second option is to transfect the cells with the mRNAs to express the Cas9 protein and the gRNA. And the last option is to prepare a complex of the Cas9 with the gRNA *in vitro* and then transfect the cells with it. The last option can achieve higher success and less off-target effects, because the Cas9 is degraded more rapidly, however due to the size of the Cas9 (163 KDa) the delivery into the cells is more difficult than transfecting plasmids (Wang et al., 2016; Kouranova et al., 2016). For this work, a plasmid-based strategy was used to deliver the Cas9 and gRNA to the cells.

Each of the transfections was performed by electroporation. KAGO Telo cells were the first to be transfected. After four different rounds of transfections, the cells survived the eletroporation process every time but they didn't form colonies during the hygromycin selection. Cells transfected with the pScramble_Cas9 plasmid didn't form colonies either. It is unclear why KAGO Telo cells

didn't form colonies easily. Thus, KB Telo MyoD cells were used to perform the next rounds of transfections.

Initially, the aim was to knock out *CDK12* by using the pCDK12_HDR plasmid, which contains the neomycin resistance gene flanked by 1 Kb homology arms. This plasmid was going to be transfected into the cells, expecting an homologous recombination of the 1 Kb arms with their complementary sequence in the exon 1 of *CDK12* in the DM1 cells used, which would cause the knock in of the neomycin resistance gene. This neomycin resistance gene, besides causing the knock out of the CDK12 protein, was going to be used as a selectable marker, to select the cells that had been transfected. However, experiments described in chapter 5 were performed first, and the outcome of those experiments made me re-design my *CDK12* knock out strategy, such as, instead of using the pCDK12_HDR plasmid, I attempted to knock out *CDK12* by introducing a hygromycin linear marker (Clontech) by NHEJ-mediated knock in.

In order to disrupt *CDK12* and 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 transfected cells and knock out *CDK12*. To increase the probability of mutating *CDK12*, 4 T75 flasks of KB Telo MyoD (passage 22) cells were electroporated with the pCas9_gRNA_CDK12 plasmid, which contains the Cas9 cDNA and the gRNA 2 sequence targeting *CDK12*, and the hygromycin linear marker. The Cas9 protein and the gRNA targeting exon 1 of *CDK12* were expected to be expressed from the pCas9_gRNA_CDK12 plasmid and cut the target site, producing a double-strand break in exon 1 of *CDK12*. Then, the hygromycin linear fragment was expected to be introduced, by NHEJ-mediated knock in, in the target site. Additionaly, other mutations such as insertions or deletions were expected in the target site, as repair of the double-strand break by NHEJ can lead to indels. These mutations in the target site were supposed to disrupt the production of the CDK12 protein.

Additional T75 flasks of KB Telo MyoD cells were electroporated and used as different controls. pCas9_gRNA_CDK12 without hygromycin lin-

ear marker was transfected into the cells and used as a negative control for the hygromycin selection process. pScramble_Cas9 and the hygromycin linear molecule were transfected into another T75 flask of KB Telo MyoD cells to serve as a lethality control. As CDK12 is involved in transcription elongation, its knock out could be lethal for the cells. If cells transfected with pCas9_gRNA_CDK12 don't survive whereas cells transfected with pScramble_Cas9 do survive, it may mean that *CDK12* knock out is actually lethal. The positive transfection control consisted of KB Telo MyoD cells transfected with pmaxGFP, and an additional T75 flask was transfected just with the hygromycin linear molecule to determine transfection efficiency and the frequency of random integration of the hygromycin resistance gene. Concentrations of each of the plasmids and molecules transfected into the cells are shown in Table 4.2.

Table 4.2: **Concentrations of the reagents transfected into KB Telo MyoD cells in order to mutate *CDK12*.**

pCas9_gRNA_CDK12 + Hygromycin	**Concentration**	**Number of replicates**
pCas9_gRNA_CDK12 [1µg/µl]	2 µg	4
Hygromycin linear molecule [50ng/µl]	100 ng	

Lethality Control	**Concentration**	**Number of replicates**
pScramble_Cas9 [1.9µg/µl]	2 µg	1
Hygromycin linear molecule [50ng/µl]	100 ng	

Negative Control	**Concentration**	**Number of replicates**
pCas9_gRNA_CDK12 [1µg/µl]	2 µg	1

pmaxGFP	**Concentration**	**Number of replicates**
pmaxGFP [0.5µg/µl]	2 µg	1

Hygromycin Random Integration Control	**Concentration**	**Number of replicates**
Hygromycin linear molecule [50ng/µl]	100 ng	1

The cells transfected with pmaxGFP were seeded into a 96-well plate, 48-hours post-transfection, and the GFP expression was analysed using the Micro Widefield Plate Reader. Figure 4.3 shows around 40% of GFP expression, so this suggests that the electroporation had worked and successful transfection with the other plasmids could be expected.

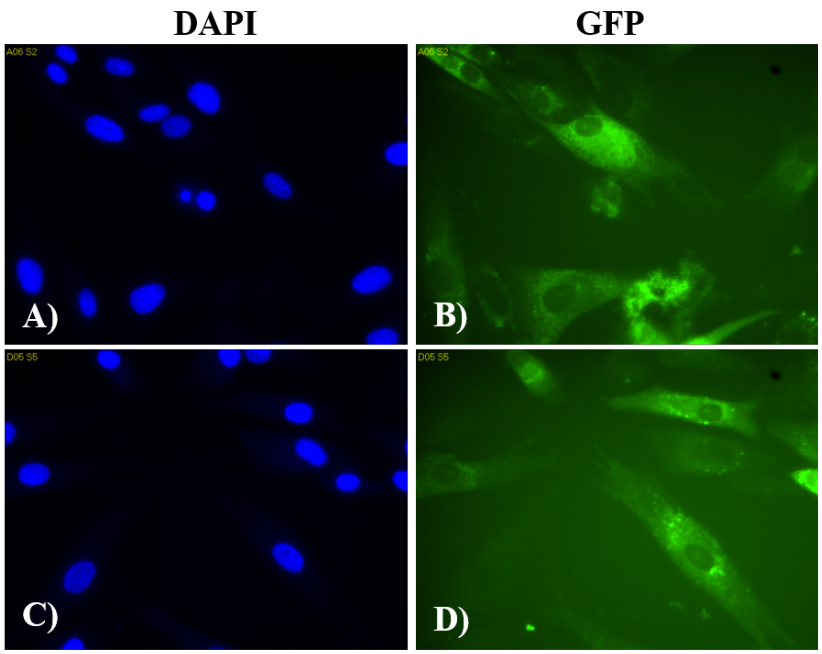


Figure 4.3: **Analysis of GFP expression after electroporation of KB Telo MyoD cells with pmaxGFP. A), C)** The nucleus of the cell is stained in blue, and **B), D)** GFP expression is shown in green.

Each of the other T75 flasks was subcultured into 8 petri dishes containing regular media. Three days later the media was replaced by fresh media containing 50 μg/ml of Hygromycin B and the process of selection was started. Media with fresh Hygromycin B was replaced every 3-4 days and after 8-9 days, all the cells in the negative control were dead. Cells from the other transfections started to form colonies 11-12 days after starting the selection process and colonies were picked up and kept growing in 24-well plates. Table 4.3 shows how many colonies were formed and picked from each of the transfections.

Table 4.3: **Number of colonies picked from each transfection event.**

Samples	**Number of colonies**
pCas9_gRNA_CDK12 + Hygromycin (4 replicates)	133
Lethality Control	65
Hygromycin Random Integration Control	102

Cells transfected with the gRNA 2 targeting *CDK12* survived and formed multiple colonies. However, all of the colonies had to be screened and the target site fully analysed, as cells transfected with the hygromycin linear molecule alone also formed many colonies. Thus, it is possible that many of the colonies transfected with pCas9_gRNA_CDK12 may have just contained the hygromycin linear marker but without the desired plasmid, pCas9_gRNA_CDK12.

4.5 Determining of the efficiency of genome editing

The colonies that survived hygromycin selection were kept growing in 24-well plates, and once they were confluent colony PCR (described in the Methods section) was performed. Conventional DNA extraction followed by PCR amplification was replaced by colony PCR amplification to speed up the screening for large sample number. The unmodified target site should produce a fragment of 968-bp. If the hygromycin linear molecule was integrated in the target site, its amplification should produce a band of around 3Kb, so the PCR amplification was performed using Long Amp Taq, which is a *Taq* polymerase optimized for amplification of fragments up to 30Kb. Different set of primers were used for various experiments of this chapter. The diagram in Figure 4.4 shows the location of each set of primers in the *CDK12* genomic DNA and cDNA.

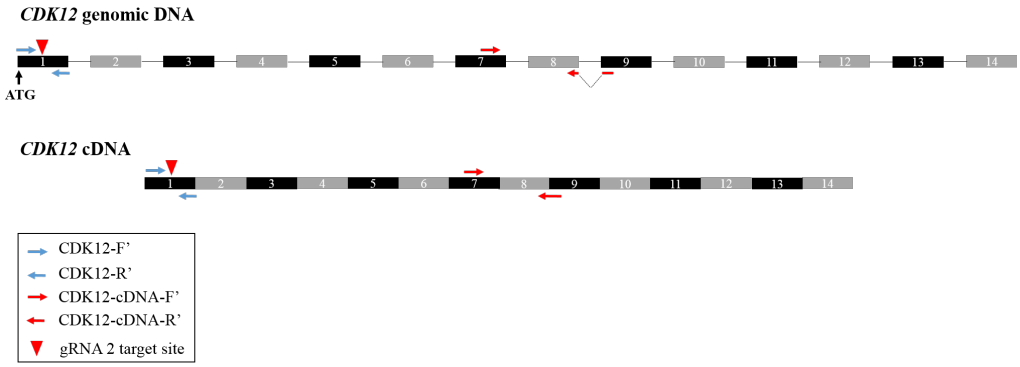


Figure 4.4: **Diagram of *CDK12* genomic DNA and cDNA showing the location of the set of primers used in this chapter.** The red triangle marks the target site of the gRNA 2 in exon 1 of *CDK12*. Blue arrows represent the primers CDK12-F' and CDK12-R' used to amplify the target site of the cells electroporated. Red arrows represent the primers described on section 4.6, CDK12-cDNA-F' and CDK12-cDNA-R', to amplify the *CDK12* cDNA and analyse the *CDK12* mRNA levels in the CRISPR-modified clones.

Out of the 133 colonies obtained from the cells transfected with pCas9_gRNA_CDK12 and the hygromycin linear marker, 23 of them were subjected to PCR amplification of the target site using the primers CDK12-F' and CDK12-R' (blue arrows in Figure 4.4). One of the colonies from the cells transfected with the Scramble gRNA was included in this first screening, as a control. The 4 T75 flasks used in this experiment were labelled replicates 1, 2, 3 and 4. Each of the petri dishes derived from each experiment were labelled with letters (A-H), and the colonies picked up from each plate were given individual numbers. Thus, colony 3-D2 is the second colony picked from petri dish D of replicate 3. The colony arising from cells transfected with the scramble gRNA was labelled as Scramble. Figure 4.5 shows the pattern of bands obtained for all the colonies screened after PCR amplification of the target site.

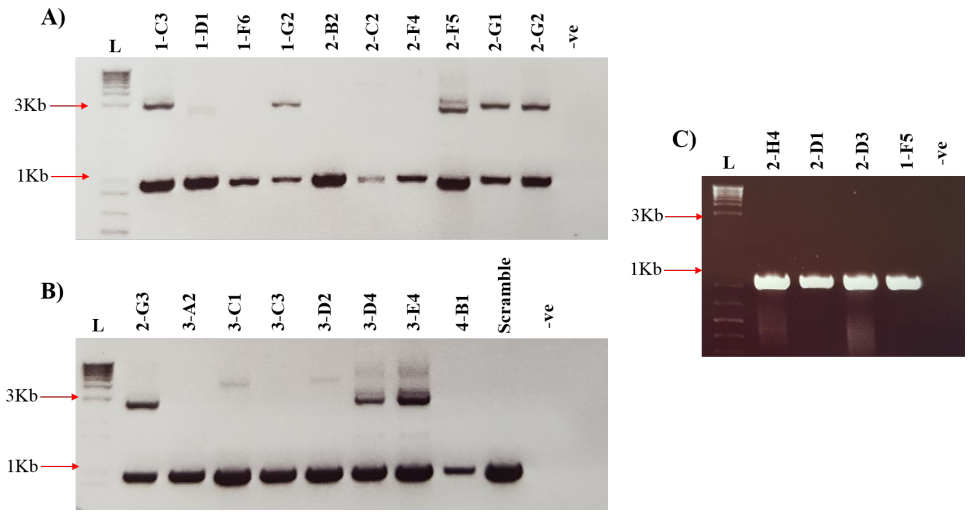


Figure 4.5: **Analysis of colonies arising from the pCas9_gRNA_CDK12 + hygromycin linear marker transfections in KB Telo MyoD cells.** Each lane shows the product of PCR amplification of DNA from a single colony with primers flanking the target site. Scramble represents a colony transfected with pScramble_Cas9 and the hygromycin linear marker. L: 1kb ladder. Arrows indicate the products of 968-bp and 3 kb where a possible insertion had occurred.

Figure 4.5 shows that all samples analysed have a 968-bp band, which corresponds to at least one of the genomic alleles. Samples with an upper band of around 3kb or bigger, were hypothesized to contain the hygromycin linear marker inserted at the target site. In order to characterise these fragments further, the gel bands were excised, DNA extracted and sequenced by Sanger sequencing. The mutations identified in the target site of the various colonies are shown in table 4.4.

Table 4.4: **Mutations identified in DM1 cells transfected with pCas9_gRNA_CDK12 and the hygromycin linear marker.** Deletions and other changes were noted in the DNA extracted from the 1 kb fragment, whereas the upper bands (3 kb) correspond to the hygromycin linear molecule integrated in the target site. Clones with a "-" symbol had no mutations in the target site in either of the alleles. Allele 1 refers to the 1 Kb PCR products, whereas Allele 2 refers to the 3 Kb PCR products. Colonies with "Sequence change" had the same number of base pairs as in control cells, but with different nucleotides. The number in brackets represents how many nucleotides were changed compared to the control sequence.

Clone	Type of mutation	
	Allele 1	Allele 2
1-C3	17-bp deletion	Hygromycin integration
1-D1	Sequence change (4 bp)	-
1-F2	Unknown	-
1-F5	-	-
1-F6	17-bp deletion	-
1-G2	Hygromycin integration	-
2-B2	-	-
2-C2	Sequence change (>10 bp)	-
2-D1	-	-
2-D3	-	-
2-F4	-	-
2-F5	35-bp deletion	Hygromycin integration
2-G1	Hygromycin integration	-
2-G2	Hygromycin integration	-
2-G3	Sequence change (>10 bp)	Hygromycin integration
2-H4	Sequence change (7 bp)	-
3-A2	Hygromycin integration	-
3-C1	Hygromycin integration	-
3-C3	-	-
3-D2	Sequence change (>10 bp)	-
3-D4	2-bp deletion	Hygromycin integration
3-E4	4-bp deletion	Hygromycin integration
4-B1	-	-

Clones 1-F5, 2-B2, 2-D1, 2-D3, 2-F4, 3-C3 and 4-B1 were not mutated by the Cas9, so most of them were excluded from further analysis, except for 2-D1 and 1-F5 which were used as a negative control. Clone 1-F2 has an unknown type of mutation in the target site because it couldn't be amplified by PCR and sent for sequencing. Clones 1-G2 and 3-D2 stopped growing after a few passages and eventually died. Clones with two different types of mutations (i.e. 1-C3, 2-F5, 2-G3, 3-D4 and 3-E4) have both alleles mutated. Overall, the experiment was

successful and exon 1 of *CDK12* was mutated in 16 of the 23 colonies resistant to hygromycin analysed. The integration of the hygromycin linear marker in the target site should disrupt *CDK12*. Furthermore the deletions identified in the target site are not multiples of 3, so the frame of translation won't be maintained and CDK12 protein should not be produced or the protein produced should not be functional.

Tables 4.5 and 4.6 summarise the number of colonies and alleles mutated along with the efficiency of the CRISPR system in this experiment. The remaining colonies produced from these transfections were not screened as the 16 shown in Table 4.4 provided sufficient material to study the effect of mutations in *CDK12* on foci and myotonic dystrophy pathogenesis.

Table 4.5: **Summary of the number of hygromycin-resistant colonies and alleles mutated after CRISPR/Cas9-genome editing of the *CDK12* gene in DM1 cells.**

	No PCR amplification	No cleavage	Mutations	Fragment insertion	Total
Colonies	1	7	16	10	23
Alleles	2	24	10	10	46

Table 4.6: **Efficiency of Cas9 cleavage and hygromycin integration into the target site of *CDK12* in DM1 cells.**

	Alleles	% efficiency
Cas9 cleavage efficiency	20/44	45.4%
Recombination efficiency	10/44	22.7%

Stable DM1 cell lines with mutations in *CDK12* were successfully generated. The gRNA identified for exon 1 cleavage allows for efficiencies of around 45% of the clones selected with hygromycin (Table 4.6). The indels produced in the target site can knockdown or knock out a gene. Furthermore, integration of the selectable marker into the target site also contributes to the disruption of the gene. Thus, these cell lines allow the quantification of CDK12 transcript levels by RT-PCR and analysis of the extent of protein knock out by western blot, as well as the study of the effect of the mutations on foci formation.

4.6 Assessment of *CDK12* knock out by RT-PCR

Mutations in a gene can prevent protein translation by altering the reading frame (resulting in a premature stop codon) or through introduction of a mis-sense mutation even though transcription of the mRNA is maintained. In order to assess the level of *CDK12* mRNA reduction, total RNA was extracted from each clone and cDNA was synthesized. The cDNA was used to perform a multiplex PCR with primers against *CDK12* (red arrows in Figure 4.4) and *GAPDH*. PCR products were subjected to gel electrophoresis (Figure 4.6).

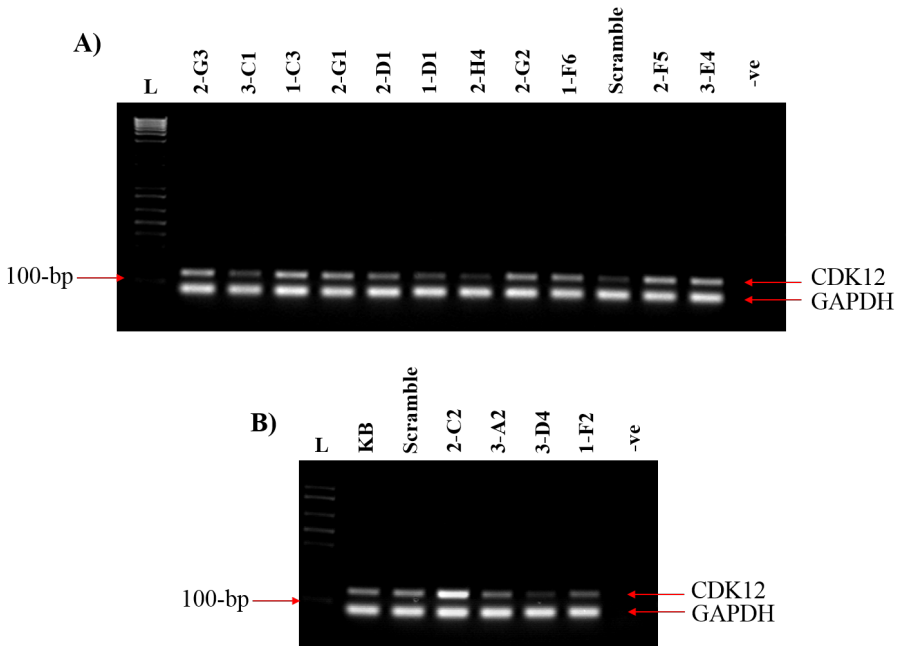


Figure 4.6: **RT-PCR amplification of *CDK12* and *GAPDH* RNA in various cell clones.** Lower band, GAPDH; upper band, CDK12; L, 1Kb ladder. Samples were split between two gels (**A) and B**)), both of which included Scramble and the negative control for PCR amplification.

The intensity of the bands was analysed using ImageJ. The intensity of each of the GAPDH bands was normalised against Scramble and the CDK12 bands were normalised against each GAPDH band. The experiment was performed 3 times and statistical data was obtained after performing an one-way ANOVA. Changes in the *CDK12* mRNA expression are shown in figure 4.7.

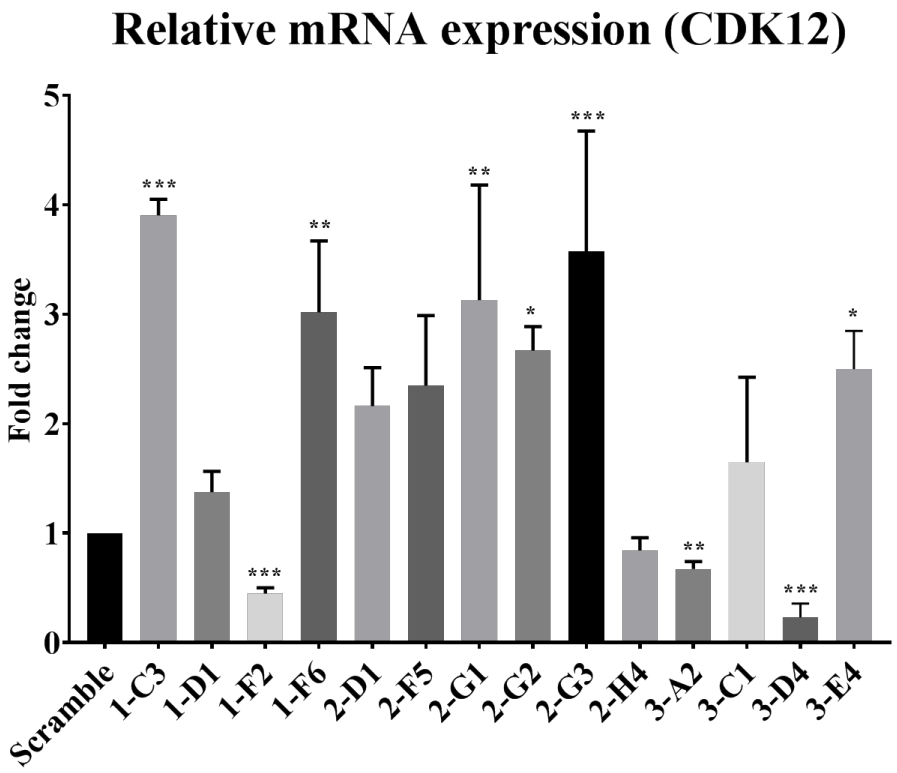


Figure 4.7: **RT-PCR analysis of *CDK12* expression in CRISPR-engineered DM1 cells.** Data are expressed as *CDK12* RT-PCR products normalized to *GAPDH*. The arbitrary value of 1 being assigned for the level of *CDK12* in Scramble. Error bars represent the average fold change ± SD for n=3 (technical replicates). $^{*}p$=0.05>x>0.01; $^{**}p$=0.01>x>0.001; $^{***}p$=<0.001.

In cell clones 2-H4, 1-F2, 3-A2 and 3-D4, *CDK12* mRNA levels are decreased. This doesn't appear to be related to the type of mutation observed in *CDK12*, as the 4 colonies present very different types of mutations. In all other colonies, *CDK12* mRNA level was increased compared to levels in control.

Mutations in the genome can cause unexpected consequences. For example, it has been reported that CRISPR-generated mutations can cause exon skipping (Mou et al., 2017). Thus, I selected some of the colonies which showed a 3 Kb band when the genomic DNA was amplified (Figure 4.5) and also the colonies with reduced CDK12 mRNA levels (Figure 4.7). RT-PCR was performed with CDK12-F' and CDK12-cDNA-R' primers for *CDK12* cDNA that were 2.8Kb apart (Figure 4.8 **A)**). The aim of this experiment was to examine in the selected CRISPR-modified clones any potential exon skipping event, and to

see whether the hygromycin resistant gene is transcribed as part of the CDK12 cDNA or it stops transcription of the allele it was integrated in.

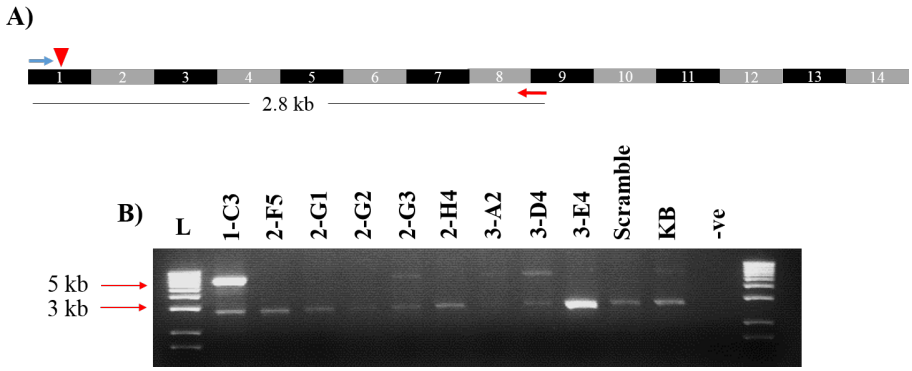


Figure 4.8: **Exon skipping analysis by RT-PCR of CRISPR-mutated DM1 cells. A)** *CDK12* cDNA diagram showing the location of the primers used to amplify the cDNA. **B)** Agarose gel of RT-PCR products from CDK12 cDNA amplification in CRISPR-generated clones with reduced CDK12 mRNA levels or with hygromycin linear marker integrated in the target site. Lane L: 1 Kb ladder.

Colonies 2-F5, 2-G1, 2-G2 and 3-E4 in Figure 4.8 (**B**) didn't show an upper band in the RT-PCR amplification, whereas they did when genomic DNA was amplified (Figure 4.5 **A)** and **B)**). The PCR fragments from this amplification were all of the expected size suggesting that exon skipping of the first exons of CDK12 was unlikely (Figure 4.8 (**B**)).

The cDNA was then amplified with the same primers used to amplify the genomic DNA (blue arrows in Figure 4.4), and the PCR product was sent for sequencing by Sanger sequencing. The aim of this experiment was to compare the genomic DNA sequence and the cDNA sequence of the clones in order to look for changes that could explain the differences observed in the *CDK12* mRNA levels. Indeed, all of the colonies showed the same sequence on BLAST searches, except for 2-G3 and 2-H4 (Figure 4.9). Nonetheless, the sequencing results for these two clones showed poor quality sequence, therefore unreliable and not worth speculation. Overall, the other colonies sequenced didn't show any post-transcriptional modifications that could explain the changes observed in the *CDK12* mRNA levels.

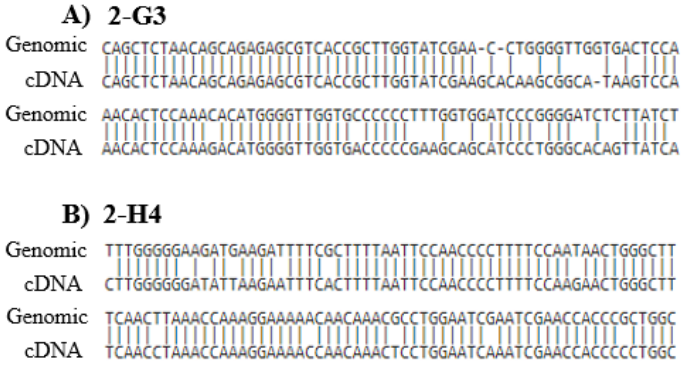


Figure 4.9: **Alignment of the genomic DNA and the cDNA sequences of the clones 2-G3 and 2-H4.** Alignment of the genomic DNA sequence from the clones **A)** 2-G3 and **B)** 2-H4 with their cDNA sequence.

4.7 *CDK12* knock out quantification by Western Blot

In view of the changes observed in *CDK12* mRNA levels I decided to examine whether CDK12 protein levels were affected. Protein was extracted from each colony and 30 μg of protein from each sample was used to perform western blots. Due to the high molecular weight of the protein, CDK12 is difficult to quantify and it was not possible to analyse protein from all colonies. Results from the CDK12 western blots in the clones in which the experiment worked, are shown in Figure 4.10 **A)**. The bands were quantified and Figure 4.10 **B)** shows the levels of CDK12 protein in each of the clones analysed, compared to the levels in KB Telo MyoD cells. These particular clones were selected because some of them have reduced levels of CDK12 mRNA whereas others have increased levels.

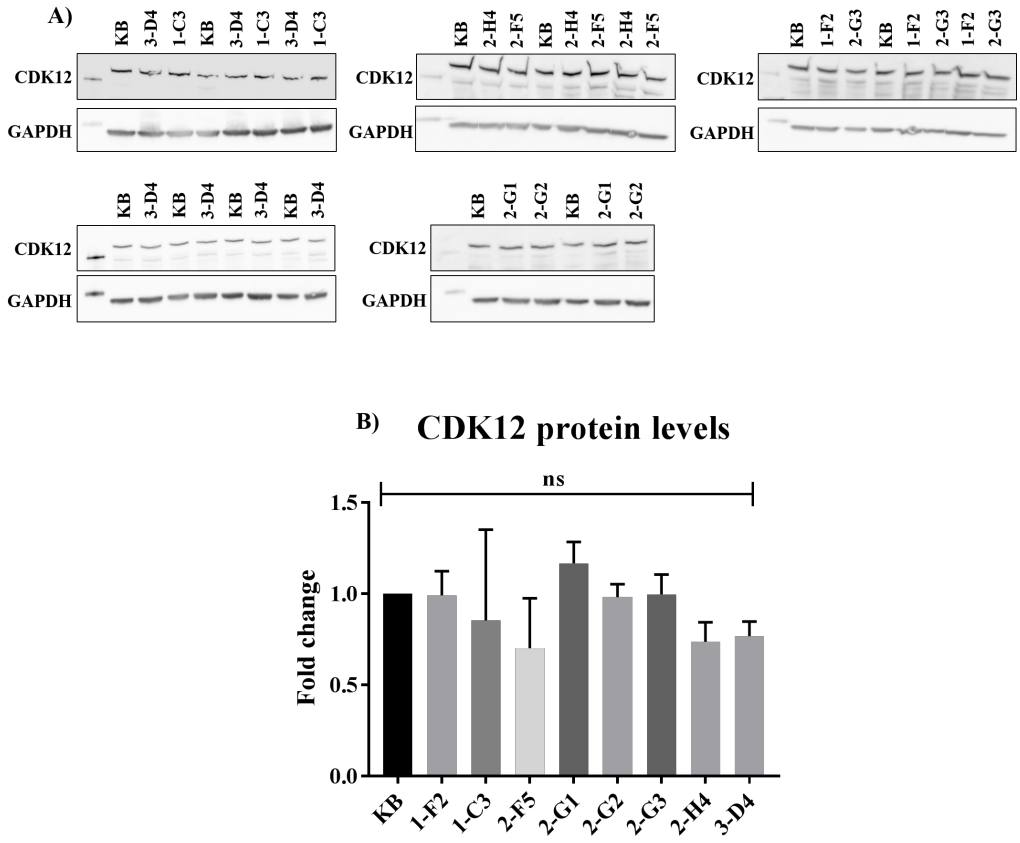


Figure 4.10: **CDK12 western blot images and protein levels in the CRISPR-modified clones.** **A)** Western blot images for *CDK12* and *GAPDH* in the selected clones. **B)** Analysis of the CDK12 protein levels from the western blot images. Each column shows the average value of CDK12 with ± SD (from 3 technical replicates) compared to protein levels in KB Telo MyoD cells. An ANOVA test showed no significant difference between the different clones.

Figure 4.10 **B)** shows no significant difference in the levels of CDK12 between the different CRISPR-modified clones and control cells when analysed by western blot and after performing an ANOVA test. As previously said, CDK12 is a protein difficult to analyse by western blot, so in addition to these analysis, CDK12 was also detected by ICC to examine the number of spots or localization of the protein in the cell. ICC was performed in four of the CRISPR-modified clones and cells were imaged using confocal microscopy. Two of the four clones used in this experiment (1-F2 and 3-D4) showed the highest reduction in the *CDK12* mRNA levels, 2-G2 showed increased *CDK12* mRNA levels and 1-F5 didn't present mutations in the target site, it was used as a negative control. Figure 4.11 shows the images obtained from the clones.

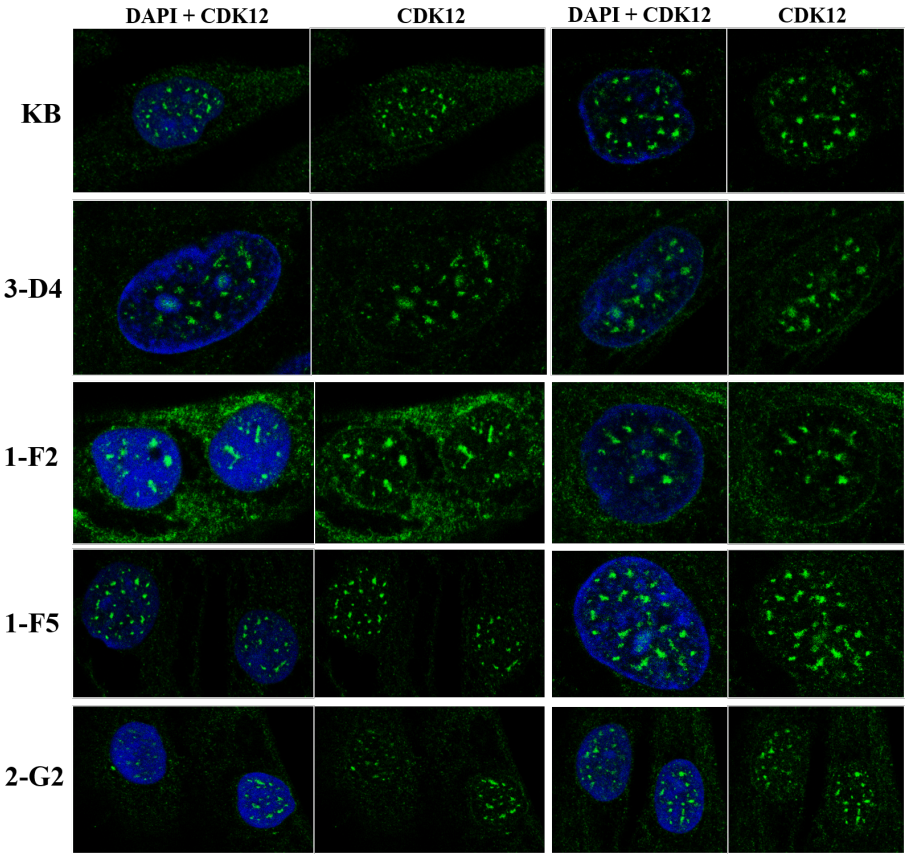


Figure 4.11: **CDK12 immunostaining in control cells (KB) and 4 CRISPR-modified cell clones.** 2 sets of images of different cells are shown per each sample, with 2 pictures per image, one combining DAPI staining of the nucleus and CDK12 staining and the other one just showing CDK12 staining.

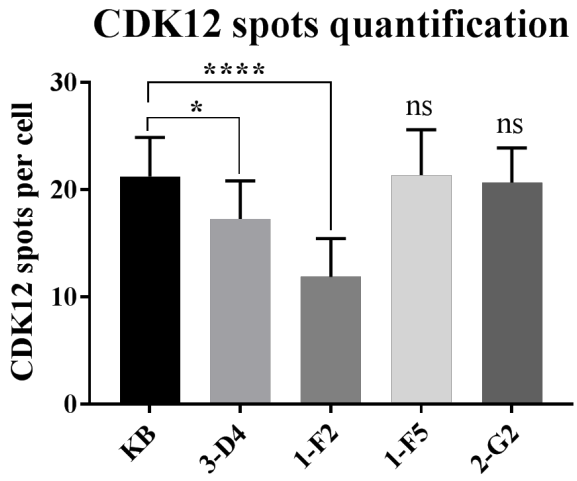


Figure 4.12: **Quantification of the number of CDK12 spots in control cells (KB) and 4 CRISPR-modified cell clones.** The number of CDK12 spots per nucleus was quantified in the CRISPR-modified clones in which CDK12 ICC was performed. 3-D4 and 1-F2 show a statistically significant reduction in the number of CDK12 spots per nucleus, compared to control cells (KB), whereas 1-F5 and 2-G2 have a similar number of CDK12 spots compared to control cells. Statistical data was obtained performing an ANOVA. *p=0.0117, ****p<0.0001.

Quantification of the number of CDK12 spots per nucleus in the CRISPR-modified clones showed a statistically significant reduction in the number of spots per cell in the clones 3-D4 and 1-F2, compared to control cells (KB) (Figure 4.12). 1-F5 showed very similar number of CDK12 spots compared to KB cells, and there was a slight reduction in the number of CDK12 spots in the clone 2-G2, nonetheless, it was not statistically significant. These results differ slightly from those obtained by western blot. Despite not observing a statistically significant difference when CDK12 western blot bands were quantified, the reduction in the number of CDK12 spots per cell in 3-D4 match the slightly reduced levels of CDK12 in Figure 4.10. The same can be observed for 2-G2, in which the levels of CDK12 in the western blot analysis and in the ICC quantification are very similar to those obtained in control cells (KB). Nonetheless, in the case of 1-F2, western blot quantification of CDK12 protein levels didn't show any difference with control cells, whereas when analysed by ICC, the number of CDK12 spots in the cell was significantly reduced, compared to control cells. The results from the CDK12 ICC in 3-D4 and 1-F2 match those observed when the CDK12

mRNA levels were analysed.

The differences observed may be due to the technical problems faced during the performance of the western blots, as the quality of the blots is not great, so further optimization of the quantification by western blot would be needed in order to obtain a robust conclusion. However, what can be concluded from the CDK12 ICC images is that even if the number of CDK12 spots per nucleus differs from control cells, all clones analysed present CDK12 spots in the nucleus, which means that *CDK12* wasn't fully knocked out and the protein is still being produced.

4.8 Effect of *CDK12* genome editing on the mutant transcript

Modification of *CDK12* exon 1 didn't appear to significantly alter the levels of CDK12 protein in most of the CRISPR-modified clones. In order to determine whether it had any effect on the production of foci in the nucleus, the cell clones were plated on 96-well plates and *in situ* hybridization performed, using a Cy3-$(CAG)_{10}$ labelled probe. Cells were imaged using the Widefield Plate Reader and the number, area and intensity of foci per cell was quantified. Table 4.7 contains the values obtained for each of the parameters analysed. The values are the mean of the results from 6 different wells in the 96-well plate (around 350 - 500 cells) for each of the colonies. These values were then plotted on a graph (Figure 4.13) and the statistical data was obtained performing an ANOVA.

Table 4.7: **Foci number, area and intensity values after performing *in situ* hybridization in the CRISPR-generated clones.**

Colony	Average foci number per cell	Average foci area	Average intensity per cell
1-D1	5.01	1.23	227.84
1-F2	7.37	1.93	99.57
1-F6	4.47	1.17	60.21
2-D1	5.59	1.47	125.36
2-F5	6.32	1.37	74.04
2-G1	6.69	1.94	134.79
2-G2	4.14	1.02	92
2-G3	5.75	1.63	43.80
2-H4	4.45	1.08	89.13
3-C1	4.87	1.27	57.35
3-D4	9.47	2.27	134.75
3-E4	6.3	1.55	84.38
Scramble	4.09	1.1	93.31
KB (control)	**4.45**	**1.15**	**74.2**

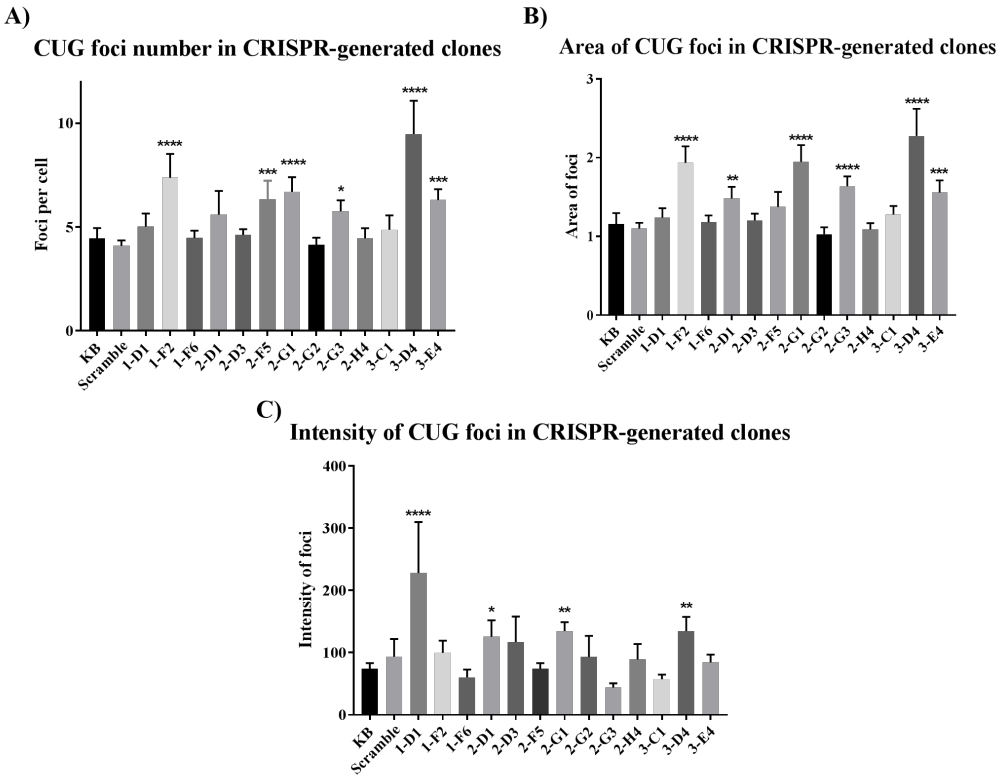


Figure 4.13: **Statistical analysis of the number, area and intensity of foci per cell in the CRISPR-generated clones.** Plotted in the graph are the values from table 4.7 and the statistical significance obtained performing an ANOVA for each of the colonies' values compared with KB cells' values. $^{*}p$=0.05>x>0.01; $^{**}p$=0.01>x>0.001; $^{****}p$=<0.0001.

Contrary to my intention, genome editing to inactivate CDK12 was not successful and therefore didn't reduce the number of foci. In fact, most of the colonies had a similar number of foci per cell compared to control. Furthermore, 6 of the colonies showed a statistically significant increase in the number of foci per cell (Figure 4.13). There was also a correlation between the number of foci per cell and the area of foci (Figure 4.14), so colonies with increased number of foci also had bigger foci. Interestingly, 5 of the clones which showed increased number of foci, have the hygromycin linear fragment integrated into exon 1 of *CDK12*. The other colony, 1-F2, could not be sequenced as it didn't produce a PCR product when the target site was amplified. These results suggest that the presence of the hygromycin linear marker in *CDK12* increases the production of foci. Nonetheless, other possibilities such as off-target effects can not be completely ruled out.

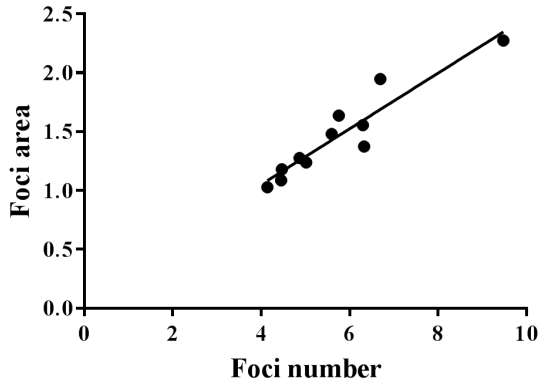


Figure 4.14: **Relationship between the number of foci per cell and the area of foci in the CRISPR-generated clones.** There is a correlation between the number of foci per cell and the area of foci. Equation y=0.2364x + 0.1051, $^{****}p$=<0.0001, r^2=0.9395, n=11.

The next table (Table 4.8) summarizes the data known from each of the colonies, including type of mutations found in the target site, levels of CDK12 mRNA and protein and number of foci per cell. In the table, "Unknown" is used for colonies in which some or any of the data was not analysed, either by technical difficulties or lack of time. "Not altered" refers to colonies showing results very similar to those obtained in controls. Colonies with a "-" symbol in the mutations column had no mutations in the target site.

Table 4.8: **Data from the type of mutations found in the target site, CDK12 mRNA and protein levels and number of foci per cell for each of the colonies screened in this chapter.** Type of mutations found in each of the CRISPR-generated colonies, along with their CDK12 mRNA and protein levels and the number of foci per cell. "Unknown" was used for colonies in which a specific data was not measured (e.g. CDK12 mRNA or protein levels, or number of foci per cell). "Not altered" refers to colonies showing results very similar to those obtained in control cells. Colonies with a "-" symbol in the Type of mutations column, had no mutations in that allele in the target site.

Clone	Type of mutation		CDK12 mRNA level	CDK12 protein level	Number of foci/cell
	Allele 1	Allele 2			
1-C3	17-bp deletion	Hygromycin integration	Increased	Not altered	Unknown
1-D1	Sequence change (4 bp)	-	Not altered	Unknown	Not altered
1-F2	Unknown	-	Reduced	Reduced	Increased
1-F5	-	-	Unknown	Not altered	Unknown
1-F6	17-bp deletion	-	Increased	Unknown	Not altered
1-G2	Hygromycin integration	-	Unknown	Unknown	Unknown
2-B2	-	-	Unknown	Unknown	Unknown
2-C2	Sequence change (>10 bp)	-	Unknown	Unknown	Unknown
2-D1	-	-	Increased	Unknown	Not altered
2-D3	-	-	Unknown	Unknown	Not altered
2-F4	-	-	Unknown	Unknown	Unknown
2-F5	35-bp deletion	Hygromycin integration	Increased	Not altered	Increased
2-G1	Hygromycin integration	-	Increased	Not altered	Increased
2-G2	Hygromycin integration	-	Increased	Not altered	Not altered
2-G3	Sequence change (>10 bp)	Hygromycin integration	Increased	Not altered	Increased
2-H4	Sequence change (7 bp)	-	Reduced	Not altered	Not altered
3-A2	Hygromycin integration	-	Reduced	Unknown	Not altered
3-C1	Hygromycin integration	-	Increased	Unknown	Not altered
3-C3	-	-	Unknown	Unknown	Unknown
3-D2	Sequence change (>10 bp)	-	Unknown	Unknown	Unknown
3-D4	2-bp deletion	Hygromycin integration	Reduced	Reduced	Increased
3-E4	4-bp deletion	Hygromycin integration	Increased	Unknown	Increased
4-B1	-	-	Unknown	Unknown	Unknown

4.9 Effect of *CDK12* overexpression on the mutant transcript

Previous results of the effect of genome editing *CDK12* exon 1 in the *CDK12* mRNA and protein levels and in the numbers of nuclear foci were unexpected, as it doesn't appear to be a link between the kind of mutation observed in the target site of each clone and the effects produced in the cell regarding CDK12 production (Table 4.8). It is not clear why the mutations created by the CRISPR system produce such different results, but what can be concluded is that *CDK12* was not properly knocked out in this experiment. Thus, an additional experiment was performed in order to see whether CDK12 transient overexpression would affect the number of foci per cell in DM1 and DM2 cells.

DM1 and DM2 cells were electroporated with a plasmid that contains the CDK12 cDNA (a 5935 bp transcript) cloned into the ORF of the pReceiver-M12 Expression clone plasmid (GeneCopoeia) in order to transiently overexpress the protein (Figure 4.15). Three different concentrations of the plasmid were transfected into the cells, 3, 4 and 5 μg. Two days post-transfection, cells were seeded onto 96-well plates and grown for 24 hours before performing the *in situ* hybridization protocol. Foci were detected using a Cy3-labelled $(CAG)_{10}$ or $(CAGG)_8$ probe (Figure 4.16) and quantified (Figure 4.17). Untransfected cells were used as a control.



Figure 4.15: **Structure of the plasmid used to transiently overexpress CDK12 in DM1 and DM2 cells.** *CDK12* cDNA was cloned into the ORF of the pReceiver-M12 Expression clone plasmid and the resulting plasmid was transfected into DM1 and DM2 cells to overexpress *CDK12*.

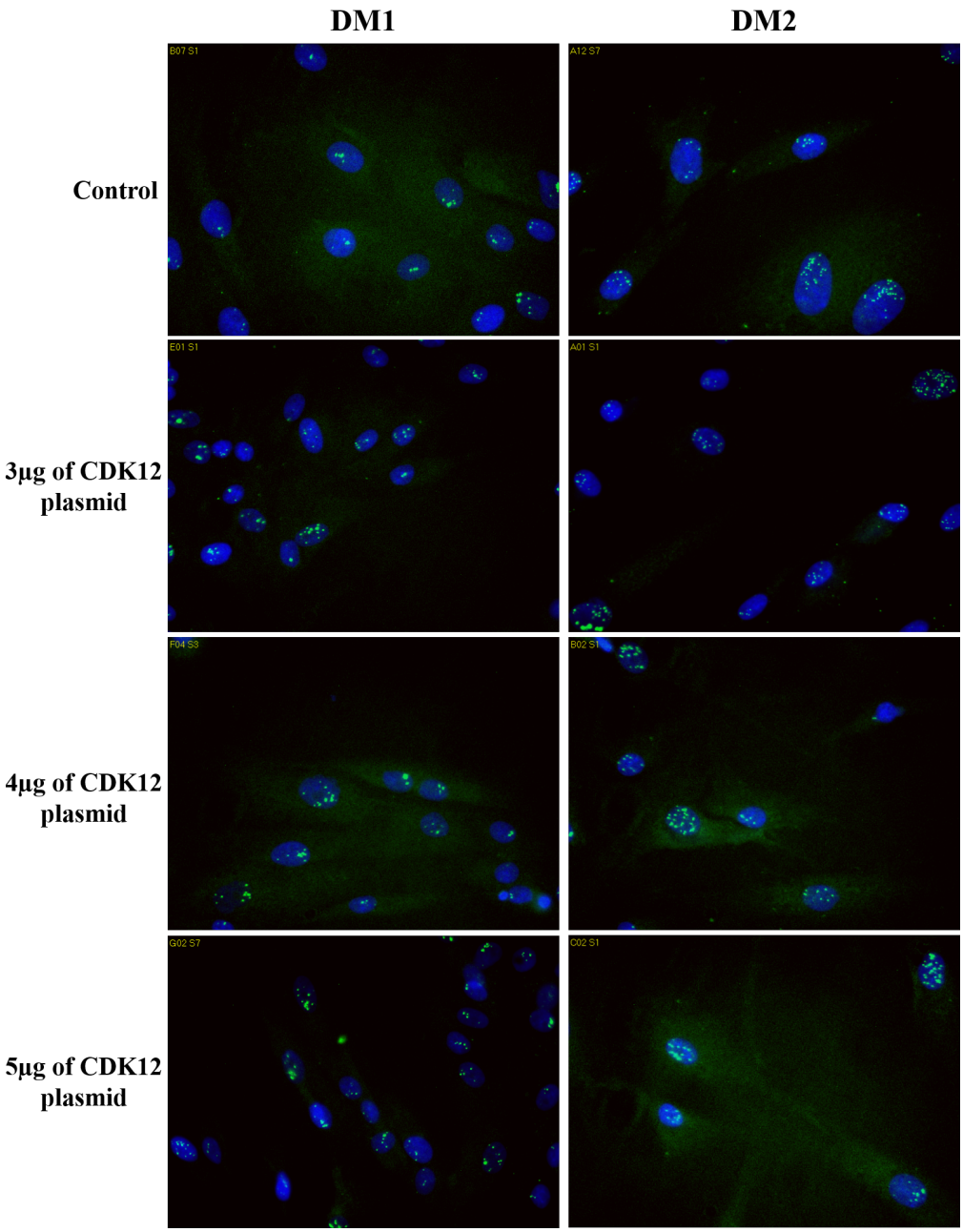


Figure 4.16: **Representative images of foci in DM1 and DM2 cells transfected with a plasmid to overexpress CDK12.** The nucleus of the cells is shown in blue whereas foci can be seen as green spots.

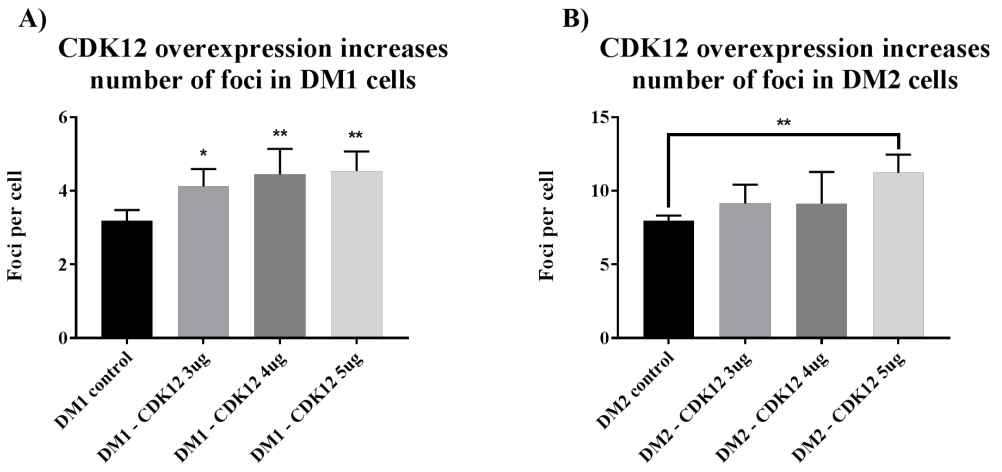


Figure 4.17: **Transient overexpression of CDK12 increases the number of foci per cell in DM cells.** The graph shows the number of foci per cell quantified in **A)** DM1 and **B)** DM2 cells transiently transfected with a plasmid expressing CDK12. Statistical data was obtained performing an ANOVA. $^{*}p=0.0146$; $^{**}p=0.0006<p<0.0021$.

Data from the quantification of the number of foci in Figure 4.17 shows a dose-response increase in the number of foci per cell in both DM1 and DM2 cells when increasing concentrations of the plasmid that expresses CDK12 were transfected into the cells. DM1 cells transfected with 3 μg of the CDK12 plasmid already showed a statistically significant increase in the number of foci per cell compared to control cells. In the case of DM2 cells, the increase was statistically significant with cells transfected with 5 μg of the CDK12 plasmid. These results suggest, on the one hand, that CDK12 is involved in the transcription of *DMPK* and *CNBP*, and probably other genes, and therefore in the mutant transcript. On the other hand, these results support the idea that *CDK12* was not properly knocked out in any of the CRISPR-modified clones, as none presented a reduced number of foci per cell.

4.10 Summary

The aim of this chapter was to generate stable DM1 and DM2 cell lines in which *CDK12* had been knocked out in order to study subsequent molecular events and clarify the role of CDK12 in the DM pathology. A suitable

gRNA with high activity *in vitro* to target exon 1 of the gene was identified and pCas9_gRNA_CDK12, a plasmid from which this gRNA and the Cas9 protein could be expressed, was generated. Electroporation of DM1 cells with this plasmid and the hygromycin linear fragment produced 133 resistant colonies. Colony PCR of 23 of these cell clones revealed modifications in the target site in 16 colonies. Modifications observed included deletions of a few base pairs and other changes, and insertion of the Hygromycin linear marker into the target site, at an efficiency of around 22% in the alleles of the colonies screened. Based on these results, Cas9 cleavage efficiency was around 45% in the hygromycin-resistant colonies.

Analysis of the *CDK12* mRNA and protein levels revealed unexpected results. Most of the colonies showed *CDK12* mRNA levels to be increased compared to levels in control cells, whereas just 3 colonies (1-F2, 3-A2 and 3-D4) showed statistically significant reduced mRNA levels. However, the mRNA level results didn't correlate with the estimated levels of CDK12 protein, as the colonies in which CDK12 was analysed by western blot didn't show any statistically significant difference with the CDK12 levels in control cells. Nonetheless, CDK12 is a protein difficult to analyse by western blot due to its high molecular weight and the technical problems that arise from it when transferring the protein from the gel to the membrane. As an alternative approach, CDK12 was examined by ICC in a few clones. Using this method, statistically significant reduced number of CDK12 spots in the nucleus were found in the clones 1-F2 and 3-D4, which also presented reduced levels of *CDK12* mRNA. No significant difference was found in the number of CDK12 spots in the colonies 1-F5 and 2-G3. Overall, it doesn't appear to be a link between the mutations found in the target site and the effect on *CDK12* mRNA or protein levels, as there isn't any logical correlation between these three variables. There are a number of possible explanations for the unexpected results obtained: 1) The RNA levels may have not been measured accurately; 2) The measurement of CDK12 protein levels and ICC data may have been inaccurate; 3) The mutations generated in exon 1 of *CDK12* do not significantly compromise the production of CDK12 protein; 4) CDK12 RNA and/or protein stability is affected by the genome modification in

ways that are beyond detection in the present study. Further characterization of the CRISPR-modified clones would be required to fully understand the effect of each type of mutation on CDK12 production and its effect.

Quantification of the number of repeat expansion foci in the CRISPR-modified clones revealed similar number of foci per cell in most of the clones, compared to control cells. Surprisingly, 6 of the clones showed a significantly increased number of foci in the nucleus. These clones have in common that they have the hygromycin linear marker inserted into the gRNA target site, in exon 1 of *CDK12*. Some of the other clones also have hygromycin integrated into the target site, but apparently in these 6 particular clones, integration of the hygromycin appears to be increasing the transcription of mutant RNA and therefore the number of foci per cell. This may be because transcription of CDK12 is being driven from the promoter on the hygromycin B fragment.

Taken together all the results for each of the CRISPR-generated colonies, there is no evident explanation that could relate the type of mutations found in each of the colonies and the effects observed in the CDK12 mRNA and protein levels and in the production of foci. Further research would be required in order to understand how the mutations are affecting the cells. Nonetheless, this was not the aim of this chapter, so for the future, efforts should be focused on obtaining a *CDK12* knock out DM1 cell line to study the role of CDK12 in DM pathogenesis.

As the effect of CDK12 knock out could not be assessed in this chapter, because the mutations introduced in exon 1 of the gene didn't compromise the production of the protein, an additional experiment was performed in order to overexpress CDK12 and quantify the number of foci. Furthermore, based on the results from the levels of *CDK12* mRNA and protein and the number of foci per cell in the CRISPR-generated clones, concerns arose that function of CDK12 in the production of foci might be the opposite to the initially hypothesized. Nevertheless, transfection of DM1 and DM2 cells with a plasmid containing *CDK12* cDNA revealed a statistically significant dose-response increase in the number of repeat expansion foci per cell in both cell types. These results support the hypothesis that CDK12 is involved in the transcription of the mutant transcript,

and probably of other genes, as based on the literature it is involved in the elongation phase of transcription. This observation also supports the mechanism of action proposed in the previous chapter, that inhibition of the function of CDK12 by the kinase inhibitors causes transcription inhibition of a subset of genes, and therefore the reason why foci are eliminated from the nucleus is the inhibition of the transcription of the mutant transcript.

Chapter 5

Experiments to control foci expression in DM cells by CRISPR/Cas9-mediated knock in of an inducible promoter

5.1 Aim

The aim of the work in this chapter was to generate a DM2 cell line with stable integration of an inducible promoter to control expression of *CNBP* using the CRISPR/Cas9 technology and the pTet-One Inducible Expression System. The inducible system should allow control of the expression of the endogenous CNBP wild type and mutant transcripts. The cell lines generated could be used as a disease model to understand the foci formation process and study the molecular events downstream of mutant transcript expression.

5.2 Introduction

In DM2 the CCTG repeat expansions present in intron 1 of the *CNBP* gene, once transcribed into pre-mRNA and spliced out of the transcript, are retained in the

nucleus where they aggregate and sequester MBNL1 proteins, altering the alternative splicing of many genes. These aggregations are termed foci and are a hallmark of myotonic dystrophy (Taneja et al., 1995; Margolis et al., 2006). There is currently no treatment for myotonic dystrophy, but many of the therapeutic strategies are focused on either the elimination or reduction of the expression of toxic transcripts or in the inhibition or disruption of the MBNL1-$(C/CTG)_n$ complex. This would release MBNL1 from pathogenic sequestration, boosting the pool of available MBNL1 for alternative splicing of its pre- and mRNA targets. In fact, the overexpression of MBNL1 in DM1 mouse model has been shown to rescue alternative splicing and alleviate myotonia (Kanadia et al., 2006).

Despite the efforts to eliminate foci or develop a therapeutic strategy, the full composition and role of foci is not yet known. The lack of information is due in part to the absence of a good disease model for myotonic dystrophy. Such disease models are difficult to make because of the instability of the repeats and the difficulty to clone large numbers of repeats into an organism or a cell (Ashizawa et al., 1993). Currently there are a few disease models available for DM1 and DM2, but none of them completely recapitulates the full phenotype of the myotonic dystrophies. As explained in the Introduction the most significant models result from the inactivation of genes in either DM1 or DM2 loci, the overexpression of toxic repeats, or the inactivation of *MBNL1* or overexpression of CELF1 to study the alternative splicing regulation.

In order to create a disease model for DM2 overcoming the difficulties associated with cloning large copy number repeats into an organism, this chapter describes efforts to knock-in, via CRISPR/Cas9 genome editing, an inducible promoter in intron 1 of the *CNBP* gene to regulate the expression of the mutant transcript in DM2 fibroblasts. The inducible promoter, engineered with an additional pause site to stop transcription from the endogenous promoter of *CNBP*, was targeted to integrate upstream of the CCTG repeats to stop expression of the mutant transcript, thus these cells wouldn't have foci in their nuclei. In the presence of an inducer such as doxycycline, the mutant transcript would be expressed again and foci would reappear in the nucleus of the cells. Furthermore, by modifying the dose of Doxycycline, the expression of the mutant transcript

could be controlled by a dose-response effect, and the ability to express the mutant transcript at any time could give valuable insight into the full pathogenic mechanism and the role of foci in the disease progression. A schematic of the integration event and expression control is shown in Figure 5.1.

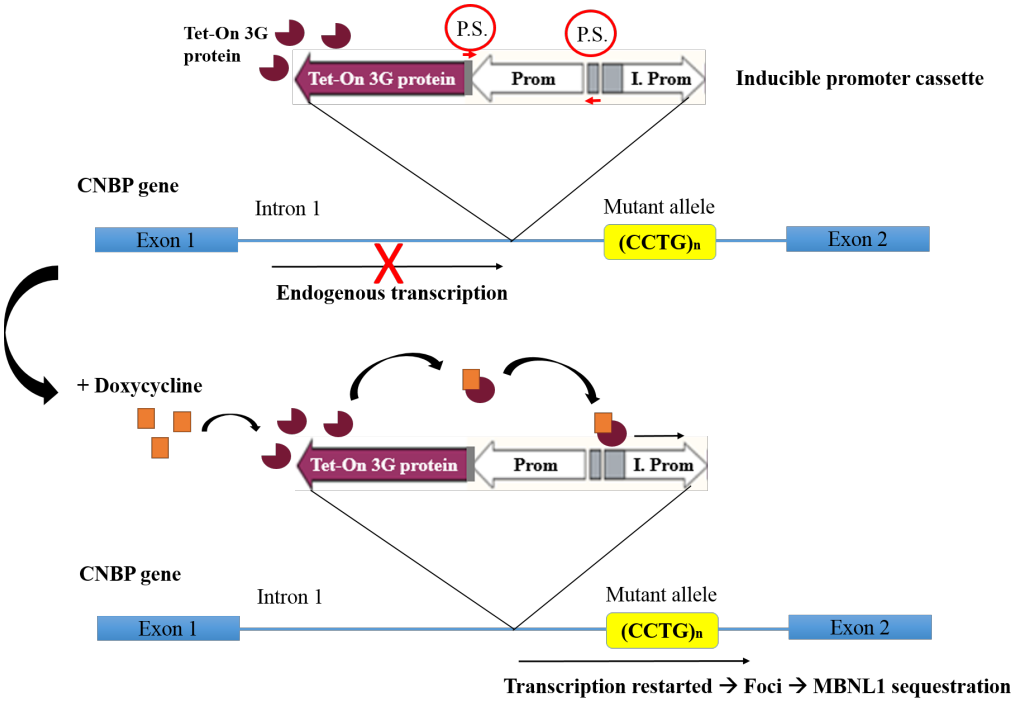


Figure 5.1: **Schematic of the integration event and control of the expression of the mutant transcript in DM2 cells.** P.S: Pause Site, Prom: Promoter, I. Prom: Inducible Promoter. Red arrows indicate the directionality of the pause sites. Tet-On 3G protein is constitutively expressed from the cassette. The first pause site will stop the endogenous transcription of the *CNBP* gene once the fragment is integrated into intron 1. Transcription of the mutant transcript will therefore be stopped and cells won't have foci. They should behave as non-DM cells. Addition of doxycycline will cause a conformational change of the Tet-On 3G protein, and this complex will bind to the inducible promoter to restart transcription of *CNBP*, allowing the expression of the mutant transcript.

The pTet-One inducible vector (Clontech) (map shown in Figure 5.2) will be used to deliver the inducible promoter into the DM2 cells. The experimental approach was modified following the outcome of each round of transfections. The first attempt to integrate the inducible promoter into the target site involved engineering the original pTet-One vector with 1 Kb homology arms at either side of the inducible promoter fragment to exploit homologous recombination to in-

corporate the inducible promoter into the genome. In the second approach, a 2.1 Kb fragment containing the sequence of the inducible promoter was amplified from the pTet-One vector and transfected into the cells as a linear fragment without flanking homology arms. In order to optimize the integration of the promoter based on the results of this new transfection, an additional round of transfection was performed using an engineered plasmid containing the sequence of the inducible promoter flanked by micro-homology arms of around 20 - 25-bp, and followed by the gRNA sequence including the PAM sequence. The presence of the micro-homology arms has been reported to allow precise integration of an exogenous DNA fragment (Yao et al., 2017).

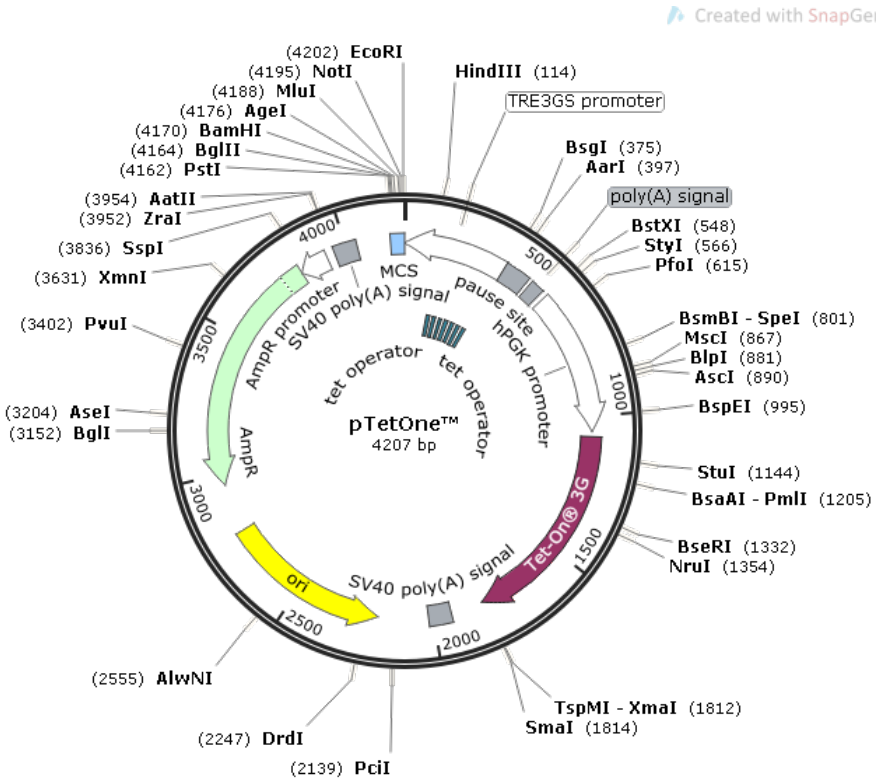


Figure 5.2: **Map of the pTet-One plasmid (Clontech).** Map of the original pTet-One vector that was later engineered with homology arms to be used as an homologous recombination template to integrate the inducible promoter into intron 1 of *CNBP*.

5.3 Analysis of *in vitro* cleavage efficiency of the gRNAs

In order to use the CRISPR/Cas9 system effectively, each gRNA needs to be tested *in vitro* before being transfected into the cells. Two gRNAs predicted by the CRISPR Design web tool (sequence shown in Table 2.3) to cleave *CNBP* were incubated with the Cas9 protein and a PCR fragment harbouring their target site. The fragments were resolved on a 2% agarose gel. Figure 5.3 shows high activity *in vitro* of the gRNA 2 as cleavage of the PCR fragment into two bands of different size was observed in the gel. However, Cas9 did not cleave the PCR fragment when complexed with the gRNA 1, meaning that gRNA 1 does not have high activity *in vitro* and most likely won't work *in vivo*.

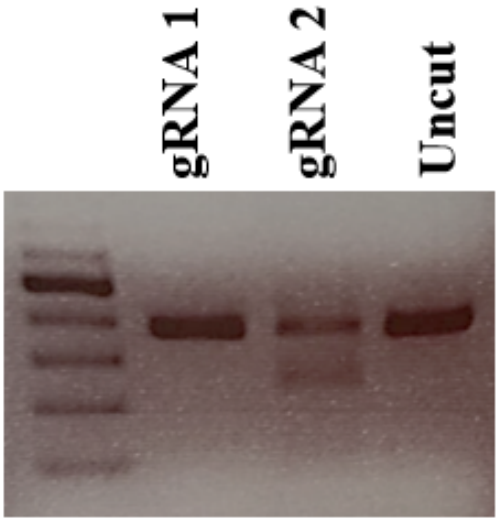


Figure 5.3: ***In vitro* assay of the predicted gRNAs targeting *CNBP***. gRNA 1 shows a single band of the same size as the uncut band, indicating that no cleavage has happened in this sample because this gRNA does not have high activity *in vitro*. gRNA 2 shows 2 bands, meaning that the original PCR fragment was cut by the Cas9 protein and therefore gRNA 2 has high activity *in vitro*.

Analysis of potential off-target effects for the gRNA 2 predicted by the CRISPR Design web tool and listed in Table 5.1, showed that all the other sequences found in the genome have four mismatches compared to the original sequence, and in all of them at least one of the mismatches is in the PAM-

proximal region. More than three mismatches, and particularly if they are in the PAM-proximal region, are likely to prevent the gRNA from binding to the DNA. Hence, based on the potential specificity and high activity *in vitro* of the gRNA 2, it was selected to perform the CRISPR-genome editing of the *CNBP* gene.

Table 5.1: **Off-target effects predicted for the gRNA 2 targeting *CNBP***. The table shows possible off-target sequences found in other chromosomes of the genome, along with their score. The score is calculated by the software based on the number and position of the mismatches and hence the probabilities of that gRNA of binding to a different locus in the genome. Every base underlined corresponds to a mismatch with the original gRNA 2. The sequence in bold corresponds to the PAM sequence.

Chromosome	Sequence	n_mismatches	Score	Gene
Chr 3	GATTGCATACATAGTACCTA**AGG**	0	100	none
Chr 5	GACTGCTTTCATGGTACCTA**AAG**	4	0.23154669292	NM_019030
Chr 14	TAATGCATGCATAGTGCCTA**AGG**	4	0.164066127333	NM_013382
Chr 12	GATTGTTTCCATAGTACCTT**CAG**	4	0.163784475483	NM_005337
Chr 12	GATTACATCCACAGTACCCA**CGG**	4	0.147311366266	NM_001122772

As explained in Chapter 4, the Cas9 protein and the gRNA can be transfected into cells in 3 different forms. For this work, a plasmid-based strategy was used again to deliver the Cas9 and gRNA to the cells.

5.4 Transfection Experiment 1

5.4.1 Transfection of the cells

Electroporation was the selected method to deliver the CRISPR reagents to the cells in order to edit their genome. Details of the protocol of electroporation can be found in the Methods section.

The pCas9_gRNA_ZNF9 plasmid, containing the Cas9 cDNA and the gRNA sequence to target *CNBP*, the pZNF9_HDR plasmid, containing the inducible promoter flanked by 1Kb homology arms, and a Hygromycin linear marker (clontech) were electroporated into KAGO Telo cells (passage 20). Both plasmids were transfected at a ratio 1:1, whereas the concentration of the Hy-

gromycin marker was calculated to be 20 times less the concentration of the pCas9_gRNA_ZNF9 plasmid. The expected outcome of this transfection was the homologous recombination of the 1 Kb arms, flanking the inducible promoter, from the pZNF9_HDR plasmid with the complementary sequence in intron 1 of *CNBP*, in order to integrate the inducible promoter upstream the CCTG repeats. The Hygromycin marker was expected to integrate somewhere else in the genome of the cells and be used as a selectable marker to ease the process of selection of engineered cells.

Additional combinations of plasmids were transfected into DM2 cells to be used as controls. Cells transfected with the pCas9_gRNA_ZNF9 plasmid and the Hygromycin marker were used to compare the efficiency of the Cas9 when transfected on its own and when co-transfected with a repair template. Cells transfected with pmaxGFP (supplied by Lonza in the electroporation kit) and the Hygromycin marker were used as a positive control for transfection efficiency. Finally, cells transfected only with pmaxGFP were used as a negative control during the Hygromycin selection process. The reagents, concentrations and number of replicates of each of the transfections can be found in Table 5.2.

Table 5.2: **Concentrations of the reagents transfected into DM2 cells in Transfection Experiment 1.**

pCas9_gRNA_ZNF9 + pZNF9_HDR + Hygromycin	**Concentration**	**Number of replicates**
pCas9_gRNA_ZNF9 [0.5µg/µl]	2 µg	4
pZNF9_HDR [1µg/µl]	1.5 µg	
Hygromycin linear molecule [50ng/µl]	100 ng	

pCas9_gRNA_ZNF9 + Hygromycin	**Concentration**	**Number of replicates**
pCas9_gRNA_ZNF9 [0.5µg/µl]	2 µg	1
Hygromycin linear molecule [50ng/µl]	100 ng	

pmaxGFP + Hygromycin	**Concentration**	**Number of replicates**
pmaxGFP [0.5µg/µl]	2 µg	1
Hygromycin linear molecule [50ng/µl]	100 ng	

pmaxGFP	**Concentration**	**Number of replicates**
pmaxGFP [0.5µg/µl]	2 µg	1

5.4.2 Analysis of transfection efficiency by GFP detection

Cells transfected with pmaxGFP and the Hygromycin marker were seeded onto a 96-well plate, 48 hours post-transfection, and grown for 24 hours, whereas the rest of the cells were transferred to petri dishes. The next day, cells in the 96-well plate were fixed with 4% PFA and nuclei were stained with DAPI. Cells were imaged using the Micro Widefield Plate Reader in order to analyse the expression of GFP. Figure 5.4 shows images of the cells acquired using fluorescence microscopy in order to detect GFP expression.

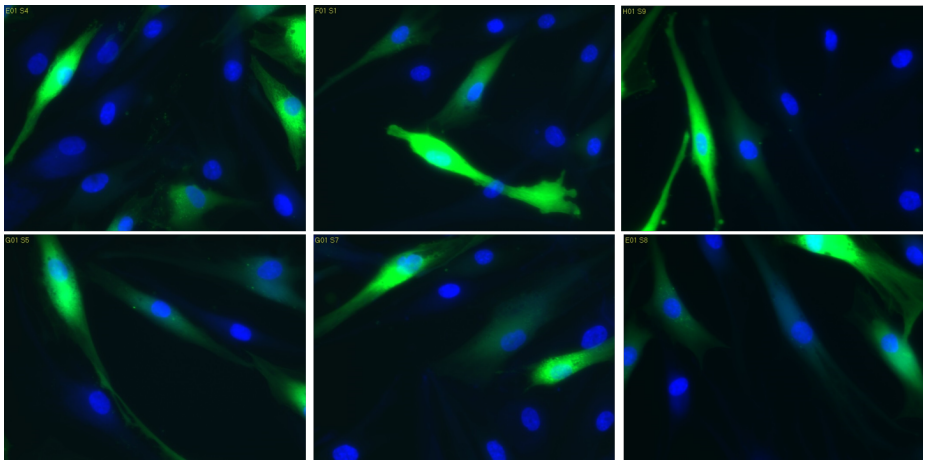


Figure 5.4: **Expression of GFP in DM2 cells transfected with pmaxGFP in Transfection Experiment 1.** Fluorescence images of cells transfected with pmaxGFP, showing the nuclei stained in blue and GFP expression in green.

Some of the cells in Figure 5.4 show GFP expression, indicating that the electroporation worked properly. By counting the number of nuclei in the images and the number of cells expressing GFP, the transfection efficiency can be estimated to be around 35% for this transfection event. These results suggested that the transfection experiment worked and successful transfection of the others plasmids should be expected.

5.4.3 Expansion of the clones

Three days after transferring the cells into petri dishes, the process of selection was started by culturing the cells with media containing 50 μg/ml of Hygromycin B. Media was replaced every 3-4 days with fresh antibiotic. After 2 weeks, single cells that were not killed by the hygromycin formed colonies which were picked and kept growing in 24-well plates. Table 5.3 shows the number of colonies picked up from each transfection. One of the four replicates of cells transfected with pCas9_gRNA_ZNF9 + pZNF9_HDR + Hygromycin didn't survive.

Table 5.3: **Number of colonies picked up from each transfection event in the Transfection Experiment 1.**

Samples	Number of colonies
pCas9_gRNA_ZNF9 +pZNF9_HDR + Hygromycin 1	15
pCas9_gRNA_ZNF9 + pZNF9_HDR + Hygromycin 2	14
pCas9_gRNA_ZNF9 + pZNF9_HDR + Hygromycin 3	11
pCas9_gRNA_ZNF9 + Hygromycin	13
Total	**53**

The number of colonies obtained from each transfection is very similar. Co-transfection of pCas9_gRNA_ZNF9 with the repair template does not have a negative effect on the cells transfection and colony growth, even though higher amounts of DNA were transfected into the cells in this case. Characterisation of the colonies was performed once they were well established in the 24-well plate.

5.4.4 Determining of the efficiency of genome editing in Transfection experiment 1

Following the expansion of the colonies resistant to hygromycin, the target site was analysed for mutations. In order to speed up colony screening, conventional DNA extraction from the cells was replaced by colony PCR of fibroblasts (described in the Methods section). PCR products were resolved on a gel and each of the bands was excised, the DNA extracted and sequenced by Sanger sequencing in order to characterise the mutations created in the target site after the cleavage by the Cas9.

The target site was amplified in the 53 colonies that survived the hygromycin selection process. From the 4 flasks of cells transfected with the pCas9_gRNA_ZNF9, the pZNF9_HDR and the hygromycin linear molecule, cells from one flask didn't survive the transfection. The remaining 3 replicates are referred as replicate 1, 2 and 3, and colonies from them will be labelled as 1-, 2-, 3-. From each replicate, cells were plated in 5 petri dishes, which will

be referred to with the letters A-E. Each colony from each petri dish received a number after the corresponding letter. Thus, 2-B4 would refer to the 4^{th} colony from petri dish B of replicate 2.

Colonies from cells transfected just with the pCas9_gRNA_ZNF9 plasmid and the hygromycin linear molecule will follow the same labelling guidelines, but will start with a C-. For example, C-A3, C-E2...

After the PCR amplification, expected size of the bands was 1 Kb. However, a 3 Kb band would suggest an integration event had happened in that sample. Figures 5.5 and 5.6 show the pattern of bands obtained for each of the colonies.

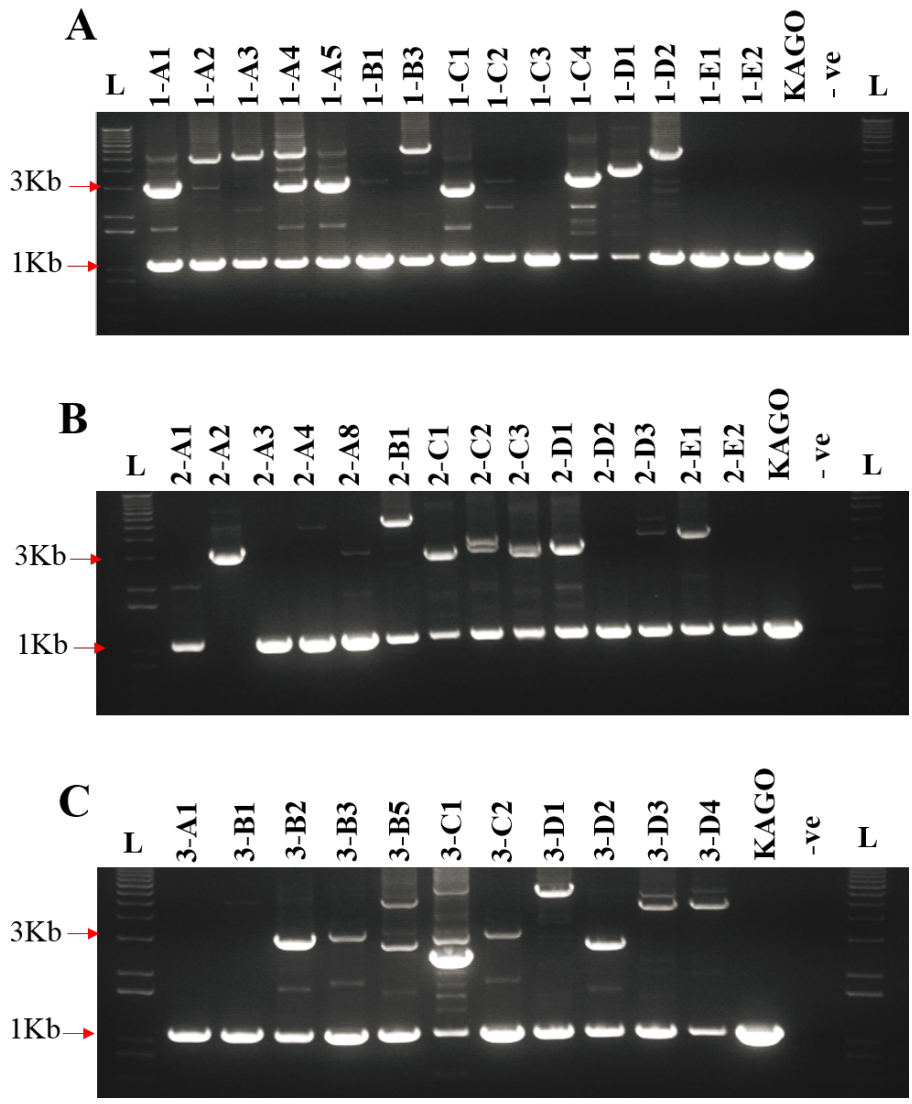


Figure 5.5: **Genotyping analysis of CRISPR/Cas9-mediated knock-in at intron 1 of *CNBP* in DM2 cells in Transfection Experiment 1.** Target-specific PCR amplification of the colonies from **A)** replicate 1, **B)** replicate 2 and **C)** replicate 3. L: 1Kb ladder.

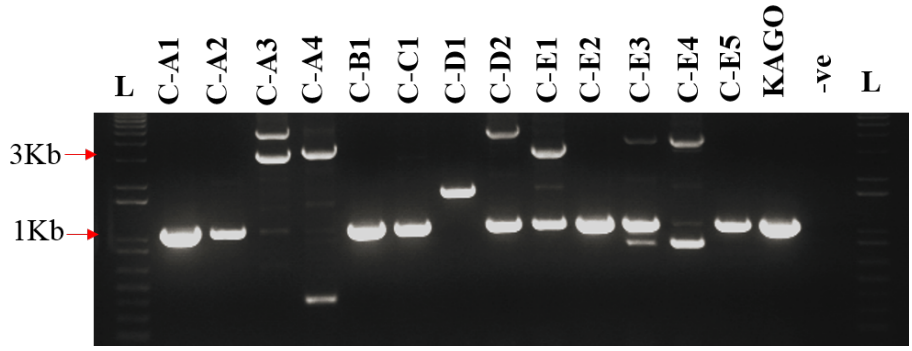


Figure 5.6: **Genotyping analysis of CRISPR/Cas9-mediated knock-in at intron 1 of *CNBP* in DM2 cells transfected without the repair template in Transfection Experiment 1.** Target-specific PCR amplification of the colonies from cells transfected just with the pCas9_gRNA_ZNF9 plasmid and the hygromycin linear molecule. L: 1Kb ladder.

Mostly all of the colonies have a 1 Kb band, which corresponds at least to one of the genomic alleles. Many of them also present an additional band which presumably corresponds to the other allele. A band of a size smaller than 1 Kb means that a deletion has happened in that allele. As the size of the inducible promoter is around 2.1 Kb, the 3 Kb band present in some of the clones suggested that the inducible promoter had been inserted into the target site. Upper bands of different size, such as 5 Kb may correspond to concatemers of the inducible promoter, or such as 2 Kb could correspond to insertions of part of the inducible promoter fragment.

From the 52 colonies screened (3-C1 was not screened as many bands were amplified by PCR from this sample), 39 of them contained an upper band. Seven of the colonies transfected without the repair template (pZNF9_HDR) also showed a 3 Kb band, which was unexpected. In order to discover what was integrated in the target site, the DNA in all the bands from the gels was sequenced. Sequencing of the DNA from the 1 Kb bands revealed different types of mutations in the target site of 36 of the clones. One of the clones contained a big insertion of an unknown fragment (C-D1). 9 clones contained only integration of the hygromycin fragment into one of the alleles, whereas the other allele remained unchanged, and the remaining 6 clones were not mutated by the Cas9. Figure 5.7 shows the type of insertions and deletions found when the DNA from

the 1 Kb band was sequenced.

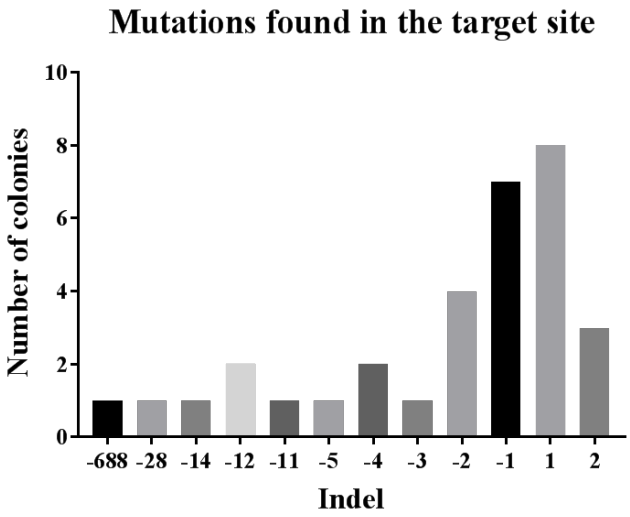


Figure 5.7: **Characterisation of the mutations created after the Cas9 cleavage in intron 1 of the *CNBP* gene in Transfection Experiment 1.** Type of mutations found after sequencing the 1Kb bands previously amplified with target-specific primers. The size of deletion or inclusion is indicated by the numbers on the X axis. The Y axis represents the number of colonies that contained a specific mutation.

The sequence from the upper bands surprisingly didn't correspond to the inducible promoter in any of the samples. Fragments were sequenced from both sides, using the forward and the reverse primers to cover a greater length of sequence. In all colonies the hygromycin linear marker had been integrated into the target site. Part of the 5' sequence of the hygromycin linear marker was found when the forward primer was used for the sequencing, and part of the 3' sequence of the hygromycin was found when the reverse primer was used for the sequencing. The size of the hygromycin linear marker is 2.1 Kb, which is very similar to the size of the inducible promoter and therefore they cannot be distinguished just based on the size of the PCR product. Additionally, rearrangements appear to have happened during the integration process, as upper bands of different sizes are present, which correspond to parts of the hygromycin linear marker or even an unknown sequence corresponding to part of a plasmid when it was analysed by a BLAST search. Table 5.4 gives details on the mutations found in each of the colonies. In this table, colonies in which the Cas9 cleaved both

alleles show two different types of mutation in the target site, whereas colonies with a "-" symbol were not cleaved by the Cas9 in either allele and the target site remains intact.

Table 5.4: **List of the colonies screened and the mutations present in each of them in the Transfection Experiment 1.** Deletions and other changes were noted in the DNA extracted from the 1 Kb fragment, whereas the upper bands correspond to the hygromycin linear molecule integrated in the target site. Clones with a "-" symbol had no mutations in the target site in either of the alleles. Clones with "Sequence change" correspond to clones with the same number of nucleotides compared to control, but with different nucleotides. The number between brackets represent how many nucleotides were changed. Allele 1 refers to the 1 Kb PCR products, whereas Allele 2 refers to the 3 Kb PCR products.

Clone	Type of mutation		Clone	Type of mutation	
	Allele 1	Allele 2		Allele 1	Allele 2
1-A1	12-bp deletion	Hygromycin integration	**2-D3**	Sequence change (3bp)	Hygromycin integration
1-A2	2-bp deletion	Hygromycin integration	**2-E1**	1-bp insertion	Hygromycin integration
1-A3	Hygromycin integration	-	**2-E2**	Sequence change (4bp)	-
1-A4	1-bp deletion	Hygromycin integration	**3-A1**	-	-
1-A5	12-bp deletion	Hygromycin integration	**3-B1**	5-bp deletion	Hygromycin integration
1-B1	1-bp insertion	Hygromycin integration	**3-B2**	Sequence change (1bp)	Hygromycin integration
1-B3	11-bp deletion	Hygromycin integration	**3-B3**	1-bp deletion	Hygromycin integration
1-C1	1-bp deletion	Hygromycin integration	**3-B5**	Hygromycin integration	-
1-C2	1-bp insertion	Hygromycin integration	**3-C2**	1-bp deletion	Hygromycin integration
1-C3	28-bp deletion	-	**3-D1**	4-bp deletion	Hygromycin integration
1-C4	Hygromycin integration	-	**3-D2**	1-bp insertion	Hygromycin integration
1-D1	2-bp insertion	Hygromycin integration	**3-D3**	2-bp insertion	Hygromycin integration
1-D2	1-bp deletion	Hygromycin integration	**3-D4**	2-bp insertion	Hygromycin integration
1-E1	Sequence change (2bp)	-	**C-A1**	-	-
1-E2	1-bp insertion	-	**C-A2**	1-bp insertion	-
2-A1	1-bp deletion	Hygromycin integration	**C-A3**	Hygromycin integration	-
2-A2	Hygromycin integration	-	**C-A4**	688-bp deletion	Hygromycin integration
2-A3	1-bp insertion	-	**C-B1**	-	-
2-A4	Hygromycin integration	-	**C-C1**	-	-
2-A8	Hygromycin integration	-	**C-D1**	Integration of unknown sequence	-
2-B1	14-bp deletion	Hygromycin integration	**C-D2**	1-bp insertion	Hygromycin integration
2-C1	2-bp deletion	Hygromycin integration	**C-E1**	Hygromycin integration	-
2-C2	4-bp deletion	Hygromycin integration	**C-E2**	-	-
2-C3	2-bp deletion	Hygromycin integration	**C-E3**	Sequence change (3bp)	Hygromycin integration
2-D1	1-bp deletion	Hygromycin integration	**C-E4**	Hygromycin integration	-
2-D2	-	-	**C-E5**	3-bp deletion	-

Table 5.5: **Summary of the number of Hygromycin-resistant colonies and alleles mutated after CRISPR/Cas9-genome editing of the *CNBP* gene in Transfection Experiment 1.**

	No cleavage	Mutations	Fragment insertion	Total
Colonies	6	7	39	52
Alleles	29	36	39	104

Table 5.6: **Efficiency of the Cas9 cleavage and recombination event in Transfection Experiment 1.**

	Alleles	% efficiency
Cas9 cleavage efficiency	75/104	72.11%
Recombination efficiency	0/39	0%

As can be seen in tables 5.5 and 5.6, technically the CRISPR/Cas9 system worked, and with a very high efficiency of cleavage of the cells that were hygromycin resistant. In theory *CNBP* should have been inactivated in the colonies generated. However as that was not the aim of this project, experiments to analyse *CNBP* transcript and protein levels were not performed. Efforts were made to get the desired clone.

As the 2.1 Kb hygromycin linear marker had been integrated by NHEJ-mediated knock in at such high efficiency in the target site after the cleavage of the Cas9, I re-designed the transfection plan. For the new round of transfection, the inducible promoter cassette was amplified from the pTet-One vector as a 2.1 Kb linear fragment and used to transfect new KAGO Telo cells.

5.4.5 Detection of mutant transcripts by FISH in the CRISPR-generated clones

As none of the colonies generated contained the inducible promoter in the target site, induction with doxycycline could not be performed. To assess the impact of CRISPR-genome editing on the *CNBP* gene and thus mutant transcript, *in situ*

hybridization was performed on some of the colonies from this first transfection experiment, and number of foci were quantified and compared to control cells. It was hypothesized that mutations in the *CNBP* gene could cause gene disruption and therefore the number of foci would be reduced. Figures 5.8 and 5.9 show the number of foci per cell. Statistical data was obtained after performing an one-way ANOVA and the degree of significant difference with control is shown above each bar.

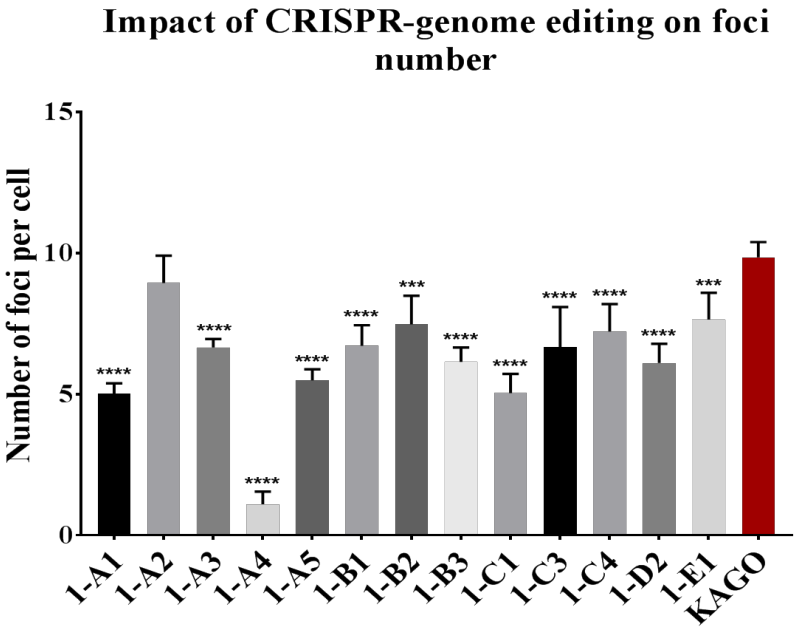


Figure 5.8: **Number of foci per cell in colonies from replicate 1, engineered with CRISPR/Cas9.** *In situ* hybridization was performed on colonies from the replicate 1 from the first round of transfection and the number of foci per cell in each of them is plotted in the graph. ***p=0.001-0.0001; ****p<0.0001.

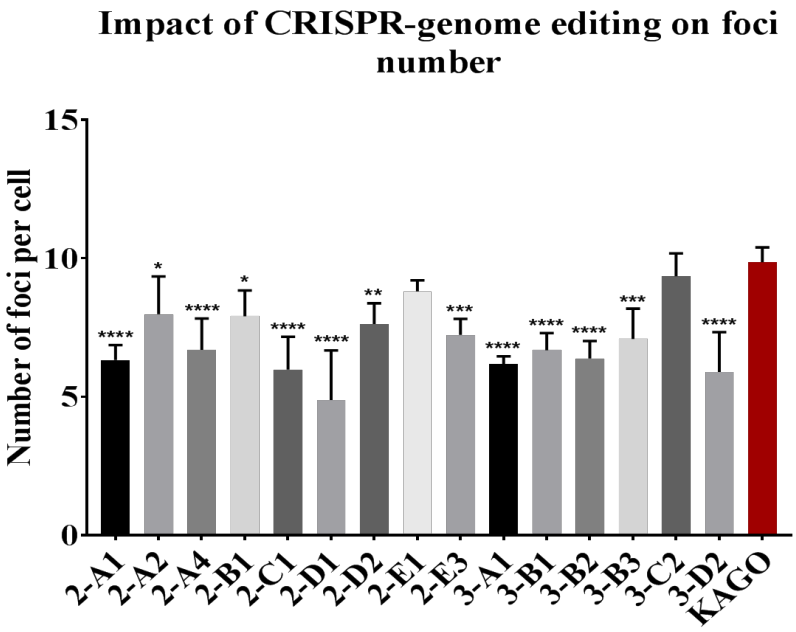


Figure 5.9: **Number of foci per cell in colonies from replicates 2 and 3, engineered with CRISPR/Cas9.** *In situ* hybridization was performed on colonies from the replicates 2 and 3 from the first round of transfection and the number of foci per cell in each of them is plotted in the graph. *p=0.05-0.01; **p=0.01-0.001; ***p=0.001-0.0001; ****p<0.0001.

Except for colonies 1-A2, 2-E1 and 3-C2, all the others showed a statistically significant reduction in the number of foci per cell when compared to control cells (KAGO). Three colonies in particular showed the biggest reduction of the number of foci, 1-A1, 1-A4 and 2-D1, with 1-A4 having virtually no foci in the nucleus of the cells, as can be seen in Figure 5.10.

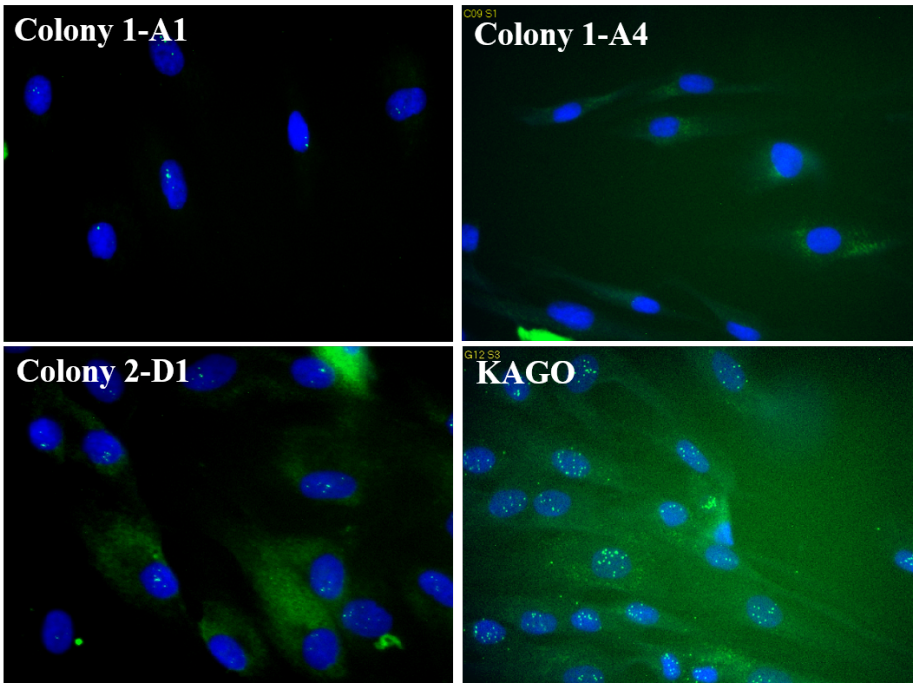


Figure 5.10: **Microscopy images of colonies 1-A1, 1-A4 and 2-D1 after performing *in situ* hybridization.** The nucleus of the cells is shown in blue and the small green spots correspond to foci.

The hygromycin linear marker was integrated into the target site in one of the alleles of these three colonies. Moreover, the three of them also contain mutations in the other allele. 1-A1 has a 12-bp deletion whereas 1-A4 and 2-D1 both contain a 1-bp deletion. Amongst the other colonies, even the ones in which hygromycin was not integrated in the target site and only contain insertions or deletions in one of the alleles, they also have a reduced number of foci. Colonies in which the number of foci was not reduced are likely to contain the mutations in the wild-type allele, and thus the transcription of the mutant allele is not compromised and the foci number is maintained. If they contain the hygromycin linear fragment integrated in the target site and also mutations in the other allele, the hygromycin marker may be integrated into the wild type allele, and the mutations in the mutant allele may not be compromising transcription of the repeats.

5.5 Transfection Experiment 2

5.5.1 Transfection of the cells

Based on the results from the Transfection Experiment 1, a second round of transfection was performed. The pZNF9_HDR plasmid wasn't transfected into the cells as a whole circular plasmid. Instead, a 2.1Kb fragment comprising the inducible promoter cassette was amplified from the plasmid and transfected into KAGO Telo P22 cells as a linear fragment, without homology arms. The linear inducible promoter was co-transfected with the pCas9_gRNA_ZNF9 plasmid and the Hygromycin linear molecule. This transfection was expected to produce colonies with the inducible promoter integrated into intron 1 of *CNBP* by NHEJ-mediated knock in, and the hygromycin marker integrated somewhere else in the genome of the cells to be used as a selectable marker. The negative control consisted of cells transfected with the pCas9_gRNA_ZNF9 plasmid and the linear inducible promoter and as a positive control for transfection efficiency, pmaxGFP was used to transfect the cells. DNA transfected, concentrations and number of replicates are shown in Table 5.7.

Table 5.7: **Concentrations of the reagents transfected into the cells in Transfection Experiment 2.**

pCas9_gRNA_ZNF9 + Inducible Promoter + Hygromycin	**Concentration**	**Number of replicates**
pCas9_gRNA_ZNF9 [1µg/µl]	3 µg	3
Inducible Promoter [0.5µg/µl]	750 ng	
Hygromycin linear molecule [50ng/µl]	75 ng	

pCas9_gRNA_ZNF9 + Inducible Promoter + Hygromycin	**Concentration**	**Number of replicates**
pCas9_gRNA_ZNF9 [1µg/µl]	6 µg	3
Inducible Promoter [0.5µg/µl]	1.5 µg	
Hygromycin linear molecule [50ng/µl]	100 ng	

Negative Control	**Concentration**	**Number of replicates**
pCas9_gRNA_ZNF9 [1µg/µl]	4 µg	1
Inducible Promoter [0.5µg/µl]	660 ng	

pmaxGFP	**Concentration**	**Number of replicates**
pmaxGFP [0.5µg/µl]	2 µg	1

5.5.2 Analysis of transfection efficiency by GFP detection

As for Transfection experiment 1, 48 hours post-transfection cells transfected with pmaxGFP were seeded onto a 96-well plate and the next day the expression of GFP was analysed using a fluorescence microscope, whereas the rest of the cells were subcultured on petri dishes. Figure 5.11 shows images of the cells with the nuclei stained in blue and GFP expression in green. Similar to the previous transfection, in this case the efficiency of transfection was around 40% for pmaxGFP, so successful transfection should be expected with the other plasmids.

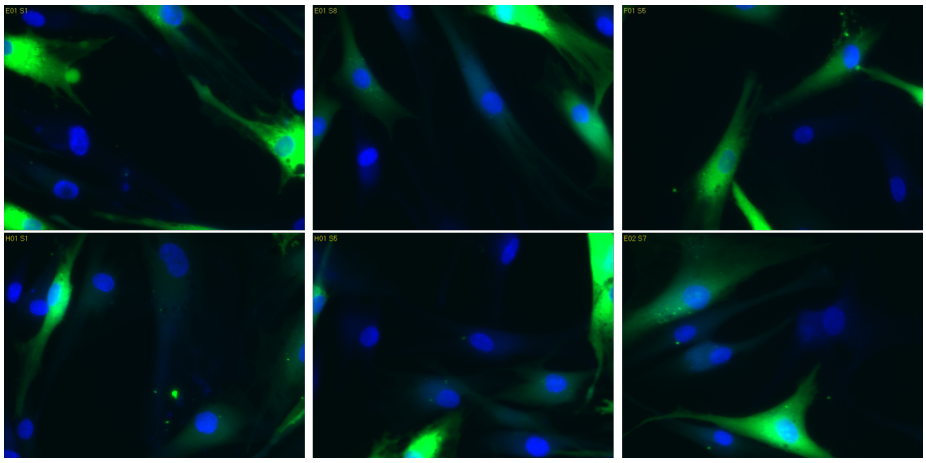


Figure 5.11: **Expression of GFP in DM2 cells transfected with pmaxGFP in Transfection Experiment 2.** Fluorescence images of cells transfected with pmaxGFP, showing the nuclei stained in blue and GFP expression in green.

5.5.3 Expansion of the clones

All steps post-transfection until the pick up of the colonies were the same as described for the previous transfection. In this case, 33 colonies were selected and kept growing. However, some of them died a few days later and just 21 colonies were screened. It was not surprising the cell death because KAGO Telo cells are more sensitive than DM1 cells, and for the cells to form colonies, conditioned media had to be used every time.

Table 5.8: **Number of colonies picked up from each transfection event in Transfection Experiment 2.**

Samples	Number of colonies
pCas9_gRNA_ZNF9 (3µg) + Inducible Promoter + Hygromycin	16
pCas9_gRNA_ZNF9 (6µg) + Inducible Promoter + Hygromycin	5
Total	**21**

Cells transfected with 3 μg of pCas9_gRNA_ZNF9 produced three times more colonies than cells transfected with 6 μg of the same plasmid. Higher cell death was also observed in cells transfected with the highest amount of plasmid.

Big amounts of exogenous DNA can be toxic to the cells, and for KAGO cells, more than 6 μg of exogenous DNA are not suitable for the cells. This has to be considered for future transfections.

5.5.4 Determining of the efficiency of genome editing in Transfection experiment 2

From the second transfection just 21 colonies resistant to hygromycin survived for the screening process. The target site was amplified with primers that were 600 bp apart. If the linear inducible promoter had been integrated, a 2.7 Kb band would be expected. Figure 5.12 shows the pattern of bands obtained for each colony. The same labelling guidelines as for the first transfection were used for these new colonies. Colonies from the 3 replicates in which the cells were transfected with 3 μg of Cas9 were labelled as 1-, 2- and 3-. Colonies from the 3 replicates in which the cells were transfected with 6 μg of Cas9 were labelled as 4-, 5- and 6-. Then each of the petri dishes received a letter and colonies within a same petri dish were numbered.

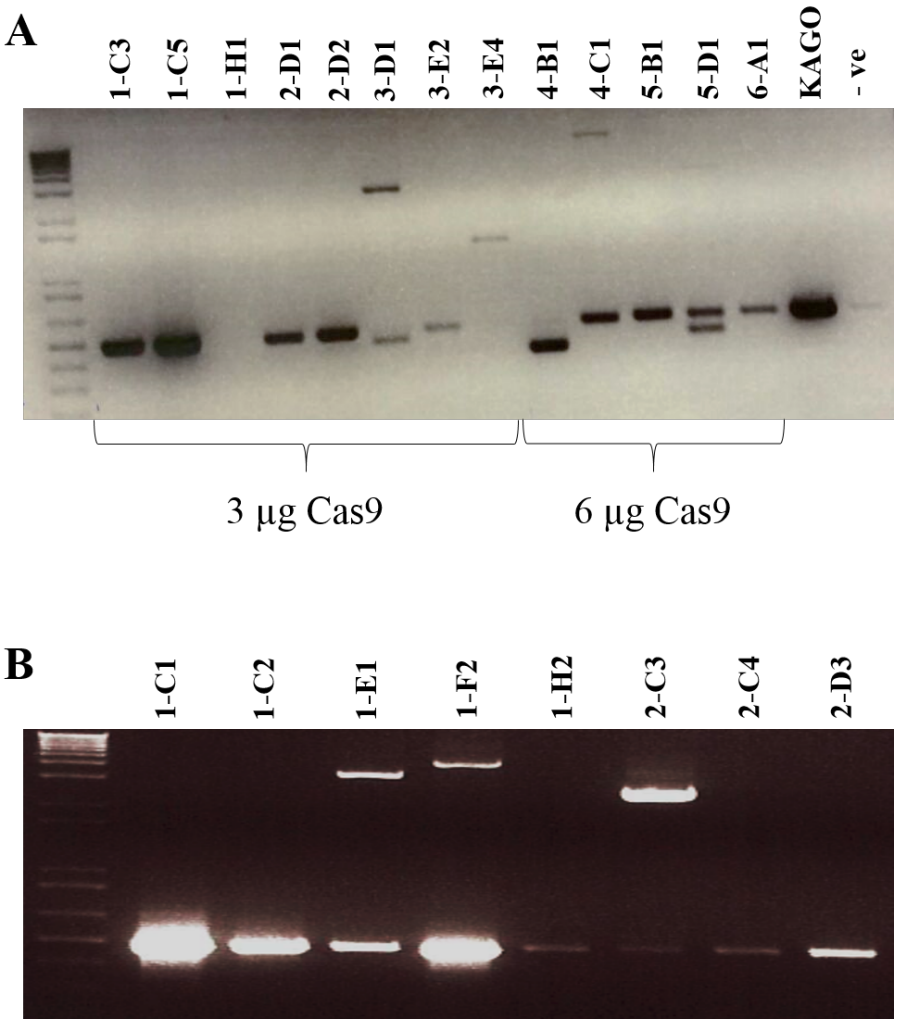


Figure 5.12: **Genotyping analysis of CRISPR/Cas9-mediated knock-in at intron 1 of *CNBP* in DM2 cells in Transfection Experiment 2.** Target-specific PCR amplification of the colonies from the second round of transfections. **A)** Colonies from both transfections, with lower and higher amounts of plasmids. **B)** Colonies transfected with 3 μg of Cas9. L: 1 Kb ladder.

As the hygromycin linear marker and the linear inducible promoter have the same size, it was necessary to sequence the DNA in the bands in order to see what had been integrated into the target site. All the bands in the gels were sequenced, and most of the lower bands showed mutations, such as insertions, deletions or base pairs mutated. The type of insertions and deletions found are shown in Figure 5.13, with 1 bp insertion being the most frequent mutation, and Table 5.9 shows details of the mutations found in each of the colonies.

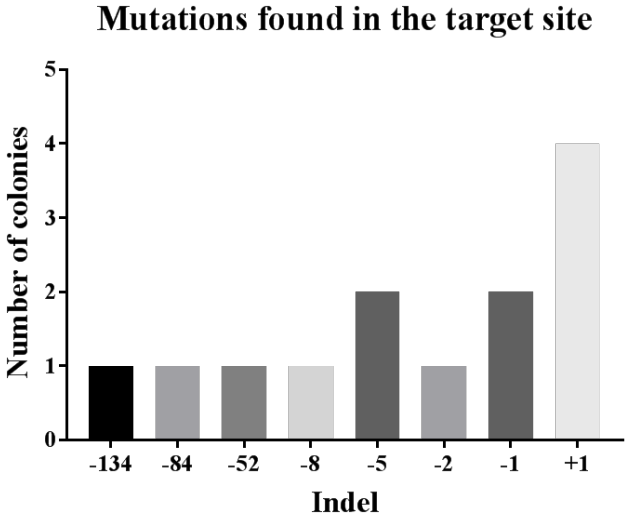


Figure 5.13: **Characterisation of the mutations created after the Cas9 cleavage in intron 1 of the *CNBP* gene in Transfection Experiment 2.** Type of mutations found after sequencing the 600 bp bands previously amplified with target-specific primers.

Table 5.9: **List of the colonies screened and the mutations present in each of them in Transfection Experiment 2.** Deletions and other changes were noted in the DNA extracted from the CRISPR-generated clones. Clones with a "-" symbol had no mutations in the target site in either of the alleles. Clones with "Sequence change" correspond to clones with the same number of nucleotides compared to control, but with different nucleotides. The number between brackets represent how many nucleotides were changed. Allele 1 refers to the 600 bp PCR products, whereas Allele 2 refers to the 2.7 Kb PCR products.

Clone	Type of mutation		Clone	Type of mutation	
	Allele 1	Allele 2		Allele 1	Allele 2
1-C1	Sequence change (7bp)	-	**2-D2**	1-bp insertion	-
1-C2	Sequence change (4bp)	-	**2-D3**	5-bp deletion	-
1-C3	Sequence change (5bp)	-	**3-D1**	52-bp deletion	pTetOne integration
1-C5	1-bp deletion	1-bp insertion	**3-E2**	1-bp insertion	-
1-E1	1-bp deletion	Hygromycin integration	**3-E4**	Sequence change (>10bp)	Hygromycin integration
1-F2	8-bp deletion	Hygromycin integration	**4-B1**	134-bp deletion	-
1-H1	** No PCR product	-	**4-C1**	5-bp deletion	-
1-H2	-	-	**5-B1**	Sequence change (8bp)	-
2-C3	pTetOne integration	-	**5-D1**	84-bp deletion	2-bp deletion
2-C4	-	-	**6-A1**	1-bp insertion	-
2-D1	-	-			

When the sequence from the upper bands was analysed, it was found that in fact 2 colonies contained the inducible promoter in the target site. The other colonies had integrated the hygromycin linear marker, or part of it. However, of the colonies that had integrated the inducible promoter, colony 2-C3 had it in the reverse orientation, and colony 3-D1 had it in the right orientation but the last 50 bp of the fragment were missing, so neither of them was still the desired clone that would allow the control of expression of the mutant transcript. Nonetheless, transfection of the inducible promoter in the form of a linear DNA fragment had indeed increased the efficiency of integration into the target site, so further optimization could help in the generation of the right clone. Furthermore, the efficiency of the CRISPR/Cas9 system was again very high (Table 5.11), similar to what was observed in the first round of transfection.

Table 5.10: **Summary of the number of hygromycin-resistant colonies and alleles mutated after CRISPR/Cas9-genome editing of the *CNBP* gene in Transfection Experiment 2.**

	No PCR amplification	No cleavage	Mutations	Fragment insertion	Total
Colonies	1	3	12	5	21
Alleles	2	18	18	5	42

Table 5.11: **Efficiency of the Cas9 cleavage and pTetOne integration in Transfection Experiment 2.**

	Alleles	% efficiency
Cas9 cleavage efficiency	23/40	57.5%
Recombination efficiency (pTetOne)	2/40	5%

In order to further optimize the integration of the inducible promoter, a third round of transfections using the pZNF9_MMEJ was performed.

5.6 Transfection Experiment 3

5.6.1 Transfection of the cells

Recently published literature (Sakuma et al., 2016; Yao et al., 2017) suggested that precise integration of exogenous DNA could be achieved using Microhomology Mediated End Joining (MMEJ)-donors at increased efficiency. Thus, a MMEJ plasmid (pZNF9_MMEJ) was designed to perform a third round of transfections to integrate the inducible promoter in intron 1 of *CNBP*. KAGO Telo cells (passage 24) were transfected with the pCas9_gRNA_ZNF9 and pZNF9_MMEJ plasmids at two different ratios, 1:1 and 1:3, as well as with the Hygromycin linear marker. DM2 cells transfected with the pCas9_gRNA_ZNF9 and pZNF9_MMEJ plasmids served as negative control during the hygromycin selection process. Table 5.12 shows the concentrations of the plasmids used for transfection and the number of replicates.

Table 5.12: **Concentrations of the reagents transfected into the cells in Transfection Experiment 3.**

pCas9_gRNA_ZNF9 + pZNF9_MMEJ + Hygromycin 1:1	**Concentration**	**Number of replicates**
pCas9_gRNA_ZNF9 [1µg/µl]	2.5 µg	3
pZNF9_MMEJ [1µg/µl]	1.47 µg	
Hygromycin linear molecule [50ng/µl]	75 ng	

pCas9_gRNA_ZNF9 + pZNF9_MMEJ + Hygromycin 1:3	**Concentration**	**Number of replicates**
pCas9_gRNA_ZNF9 [1µg/µl]	2 µg	3
pZNF9_MMEJ [1µg/µl]	3.75 µg	
Hygromycin linear molecule [50ng/µl]	150 ng	

Negative Control	**Concentration**	**Number of replicates**
pCas9_gRNA_ZNF9 [1µg/µl]	2.5 µg	1
pZNF9_MMEJ [1µg/µl]	1.47 µg	

The cells survived the electroporation but after more than 4 weeks in the

petri dishes under Hygromycin selection, none of them formed colonies, even using conditioned media. All the dishes were discarded. Due to time limitations, cells couldn't be transfected again with the pZNF9_MMEJ plasmid, so results using this approach are not available. The screening of Transfection Experiments 1 and 2 suggests that the experiment works. More time was needed to get colonies and screen them to generate the desired clone.

5.7 Summary

In this chapter a suitable target site to integrate an inducible promoter into intron 1 of the *CNBP* gene upstream the CCTG repeats in DM2 cells has been identified and a gRNA with high activity *in vitro* has been designed. A plasmid that contains both the gRNA and the Cas9 protein, pCas9_gRNA_ZNF9, has worked very efficiently in all of the transfections, hence it can be used for future experiments and the concentration needed for each transfection is optimized.

The Cas9 protein, when delivered into the cells in the form of a plasmid, along with the selected gRNA, achieved cleavage efficiencies of around 57 to 72% in the cells that survived the hygromycin selection. Cleavage of the target site promoted the generation of different types of mutations, usually in one of the alleles, and in many samples. Integration of the hygromycin linear molecule was also observed in the other allele. Integration of the inducible promoter in the target site was only achieved when transfected into the cells in the form of a 2.1 Kb linear fragment instead of in the form of a circular plasmid with homology arms for homologous recombination. Even though, the desired clone was not generated, as the linear inducible promoter can be integrated in either orientation by NHEJ, due to the absence of homology arms that would allow integration in the right orientation, and still the integration efficiency was only of around 5%.

In order to optimize the integration of the inducible promoter and achieve a precise integration, pZNF9_MMEJ was engineered and transfected into the cells. However, screening of the colonies was not performed as the DM2 cells didn't form colonies after a long period under hygromycin selection. This DM2

cell line has proven to be more sensitive than others used in the lab and conditioned media was needed every time for them to form colonies. Nonetheless, that didn't assure the formation of the colonies, as some times the transfections had to be repeated because the cells were not forming colonies. Hence, a different DM2 cell line with good ability to form colonies from single cells would be a better model to create the inducible DM2 cell line. Changing the cell line could save a lot of time and the desired clone could be generated, as the gRNA and the Cas9 work very well and the inducible promoter can be integrated in the target site, the only problem has been the low numbers or lack of colonies to be screened.

even though the inducible cell line could not be created, some of the colonies generated in the Transfection Experiment 1 that contain mutations and the hygromycin linear molecule in the target site were screened for foci, and in most of them the number of foci was reduced compared to the control. Some of the colonies with the lowest number of foci were kept in liquid nitrogen for further analysis and use in future experiments, as they are DM2 cells with very few or almost any foci in their nucleus. Presumably, mutations in these colonies compromise mutant transcript expression and therefore repeat expansion transcripts are produced at a lower rate and less foci are present in the nucleus of the cells.

Chapter 6

Final Discussion and Conclusion

Although a great deal of progress has been achieved since the discovery of the gene responsible for myotonic dystrophy, the molecular mechanism underlying the disease is still uncertain. RNA-induced toxicity is the most accepted model of pathogenesis for both types of myotonic dystrophy. This disease mechanism proposes that toxic RNAs, which cannot be exported to the cytoplasm, accumulate in the nucleus of patient's cells altering the function of alternative splicing factors, such as MBNL1 and CELF1, producing the dysregulation of a subset of genes expressed in multiple tissues. This results in a multisystemic condition (reviewed in (Ranum and Cooper, 2006)). The aggregations of mutant RNA with MBNL1 proteins are called foci and they only manifest above a certain threshold of repeat length, nonetheless the molecular mechanism by which the nucleotide repeats result in the formation of foci is not known. DM1 and DM2 repeat expansion tracts form hairpins and secondary structures that are thought to prevent their export to the cytoplasm or their degradation (Davis et al., 1997; Hamshere et al., 1997). It has been shown recently that MBNL1 plays a direct role in retaining the mutant transcript in the nucleus and in the formation of foci (Querido et al., 2011), and also DM1 foci aggregate at the periphery of SC35 speckles, which could mean that the mutant RNA is blocked from entry into speckles and restricted from nuclear export (Holt et al., 2007).

Recent findings suggest that DM pathogenesis may be caused by additional changes besides the spliceopathy, such as changes in gene expression, an-

tisense transcripts, deregulated miRNAs or RAN translation of the mutant transcript, which all appear to contribute to the disease progression. RAN translation is of special interest. According to the original RNA-induced toxicity model of pathogenesis, the mutant RNA is not exported to the cytoplasm, accumulating in the nucleus of the cells and sequestering MBNL1 proteins in both DM1 and DM2 and causing the upregulation of CELF1 in DM1. The discovery of RAN translation requires the mutant RNA to be exported to the cytoplasm in order to be translated into toxic peptides. The mutant transcript is mainly found in the form of foci, nevertheless, a fraction of this transcript remains soluble in DM1 cells (Junghans, 2009). Furthermore, nuclear RNA aggregates are dynamic, constantly forming and disaggregating, and MBNL1 has been directly involved in the aggregation process (Querido et al., 2011). A sequestration failure model has been proposed to explain the export of mutant transcripts into the cytoplasm and therefore the presence of toxic RAN proteins. This model considers 2 possibilities: (1) in cells expressing high levels of mutant RNA, RBPs are saturated and the free mutant RNA in the nucleus is exported to the cytoplasm; (2) in cells expressing low levels of RBPs proteins, the excess of mutant RNA will be exported out of the nucleus (Zu et al., 2017).

RAN proteins have been detected in different repeat-expansion disorders, including SCA8, FXTAS, ALS/FTD, DM1 and DM2 (Zu et al., 2011, 2017). RAN translation is repeat length dependent. In the case of DM1, polyGln proteins can only be detected with 50 or more CTG repeats, whereas in the case of DM2, sense polyLPAC and antisense polyQAGR were detected with repeats larger than 85 CCTGs (Zu et al., 2011, 2017). Large repeats have been seen to produce higher accumulation of RAN proteins, nevertheless, much is unknown about RAN translation in DM. For example, how much of the transcript is exported to the cytoplasm. We don't know how many transcripts are required to produce a significant amount of toxic RAN proteins, or if the transcripts have a flanking sequence that could be creating a bigger protein. In DM2 it was demonstrated that the intronic repeats are spliced out from the *CNBP* transcript and they do not contain flanking sequence (Margolis et al., 2006), however in DM1 the repeat tract contains flanking sequence from the 3'UTR region of *DMPK*. Nev-

ertheless, it is not known how much sequence is retained in the mutant transcript, and how it is being translated.

The presence of RAN proteins suggest that at least some of the mutant transcripts are being exported to the cytoplasm. However, the proportion of exported transcripts must be low and the functional significance of these small repetitive sequence proteins is unclear. Furthermore, the repeats may be affecting the cell in additional ways to the production of RAN proteins. A recent study found that CAG triplet repeats are degraded into short RNAs that act as small interfering RNAs and attack cancer cells, mainly by attacking complementary sequences in the ORF of certain genes (Murmann et al., 2018). In the case of Huntingon's disease there is an inverse correlation between the length of the CAG repeats in the huntingtin gene and cancer incidence (Turner et al., 2013; Coarelli et al., 2017). Nonetheless, myotonic dystrophy patients present a higher risk of developing cancer rather than a reduced risk (Das et al., 2012; Win et al., 2012). In DM1, the presence of mutant transcripts as siRNAs could be causing other negative effects in the cells by attacking the ORF of other genes. Important work is needed to clarify the effect of these short transcripts and to study whether they are also found in DM2 or not.

Other major questions still remain unanswered, such as: what triggers the aggregation of the mutant transcript into foci? What is blocking the export to the cytoplasm of the mutant transcript? How much MBNL1 is sequestered into foci? Is the amount of MBNL1 sequestered enough to significantly contribute to the disease? How does the mutant transcript trigger the hyperphosphorylation of CELF1? How much transcript is exported to the cytoplasm? Do RAN proteins significantly contribute to the disease pathogenesis, and what amount of proteins are needed to be detected and have an effect in the cells? Many studies strive to answer these questions and others that we have not considered about yet, but some of them will be difficult to assess, so in the meantime, every little discovery and advance about the disease is of great benefit.

As RNA foci are important hallmarks of myotonic dystrophy which play an important role in disease pathogenesis, foci have been the main focus of the

work in this thesis. The experiments performed in this work were conducted to achieve the aims proposed at the beginning of this thesis:

- To study the relationship of expansion nuclear foci with other components of the cell, such as CDK12 protein or SC35 speckles.
- To test a series of kinase inhibitors in DM2 cells that have already proven to be effective for foci reduction in DM1 cells, and to study additional effects of the compounds in both DM1 and DM2 cells.
- To unravel the mechanism of action of the kinase inhibitors tested by which they eliminate foci in DM1 and DM2 cells.
- To generate a CDK12 knockdown DM1 cell line using the CRISPR/Cas9 system in order to study the role of CDK12 in myotonic dystrophy and in foci maintenance.
- To generate a DM2 cell line with stable integration of an inducible promoter to control the expression of *CNBP* using the CRISPR/Cas9 system.

Each of these points is considered further below in the next sections.

6.1 The localization of foci in DM1 and DM2 cells

The accumulation of the mutant CTG/CCTG repeat expanded transcript in the nucleus of DM cells and their aggregation with MBNL1 proteins and others to form foci is believed to be the main cause of myotonic dystrophy. The full composition of foci is not known yet, although there are studies focused on the analysis of the mutant transcripts (Wojciechowska and Krzyzosiak, 2011). RNA foci have been seen to associate with different structures, and co-localization of foci with different proteins has been demonstrated. One of the major findings was the association of foci with SC35 speckles in DM1 cells, but not in DM2 cells (Holt et al., 2007). Experiments in this thesis to confirm these results in the cell lines used matched the results obtained by Holt *et al.* (Figure 3.1). The difference in the localization of foci in the nucleus of DM1 and DM2 cells could be one

of the causes of the differences between the two types of the disease. It is also interesting that foci are occasionally detectable in the cytoplasm of DM2 cells, whereas in our DM1 cell lines, foci are virtually always nuclear. The association of foci and SC35 speckles in DM1 could be causing a lower export to the cytoplasm of the CTG repeat expansion, whereas in DM2 foci do not associate with SC35 speckles and therefore some of the transcripts may be exported to the cytoplasm more easily. The different behaviour could also be due to the fact that in DM1 the CTG repeat expansion is located in the 3'UTR region of the *DMPK* gene whereas in DM2 the CCTG repeat expansion is located in an intron (Brook et al., 1992; Mahadevan et al., 1992; Liquori et al., 2001), therefore the DMPK expanded transcripts may be processed down a different pathway compared to CNBP expanded intronic fragments. Furthermore, in DM2 the nuclear foci lack sequences that flank the repeat expansion tracts (Margolis et al., 2006), whereas this is not the case for the DM1 repeat expansions.

SC35 speckles are interchromatin granules that act as storage centres for many proteins involved in pre-mRNA processing. Spector *et al.* identified 146 known proteins in these nuclear speckles and many others yet uncharacterised (Spector and Lamond, 2011). One of these proteins is CDK12, which is known to co-localize with SC35 speckles (Ko et al., 2001). Based on the results obtained from the co-localization of foci with SC35 speckles in the present thesis and on co-localization of CDK12 with SC35 speckles from the literature, co-localization of foci with CDK12 should be expected in DM1 cells but not in DM2 cells. It is worth mentioning that co-localization places the molecules in broadly the same space within the cell, but does not establish a direct physical association between them, with additional studies such as FRET being necessary to go beyond co-localization.

Figure 3.4 confirms that CDK12 co-localize with foci in DM1 cells. However, direct analysis of the degree of co-localization of foci with CDK12 in DM2 cells could not be established using the same method employed for DM1 cells, due to technical issues. Optimization of the protocol was attempted, focusing mainly in the fixation method for the ICC and the FISH experiments. Use of different reagents to fix and permeabilize the cells (Figure 3.3), such as methanol,

acetone, PFA, acetic acid or Triton X-100 revealed that 4% PFA as a fixative is excellent to detect foci in the nucleus of the cells. Nonetheless, it appears to affect either the epitope for the CDK12 antibody or the antibody itself. The best fixative to detect CDK12 were the alcohols, especially 50:50 Methanol:Acetone. Methanol and Acetone are solvents that precipitate and coagulate large protein molecules, and CDK12 is indeed a large protein with a molecular weight of 164 KDa (Ko et al., 2001). However, this fixative method (in my hands) affects the RNA, as less foci are detected in the nucleus when ICC is combined with FISH, around 1-3 foci in DM1 cells and none in DM2 cells, compared to when FISH is performed on its own, around 4-5 foci in DM1 cells and 8-10 foci in DM2 cells. Performance of ICC in the first place followed by FISH allowed me to successfully detect CDK12 spots and very few foci in DM1 cells, but not in DM2 cells. Further optimization of the protocol is needed in order to detect both CDK12 and foci in DM2 cells.

The degree of co-localization of SC35 speckles with CDK12 was used to infer the relationship of foci with CDK12 in DM2 cells. Comparison of this co-localization in non-DM, DM1 and DM2 cells showed statistically significant differences between the cell lines, with a higher degree of co-localization in DM cells compared to non-DM cells (Figures 3.5 and 3.7). When the number and area of the SC35 and CDK12 spots were analysed and compared, there was no evident difference between the cell lines, nonetheless, there is a slight increase in the number of SC35 speckles and in the area of CDK12 spots in DM1 and DM2 cells, which could account for the higher co-localization. Finally, results on DM2 cells and previous results suggest that in this cell line foci do not colocalise with CDK12.

Co-localization of foci with these proteins was studied here in just one DM1 cell line and one DM2 cell line as those were the ones that were going to be used for all the experiments throughout the thesis. Nonetheless, conduction of the same co-localization experiments in additional DM1 and DM2 cell lines could help us clarify if the observed localization of foci regarding SC35 speckles and CDK12 and the difference observed between DM1 and DM2, is a general phenomenon that occurs in all DM1 and DM2 cell lines or if it is specific of the

two cell lines used here.

6.2 Effect of kinase inhibitors on DM cells

There is currently no treatment for myotonic dystrophy, but significant efforts are being made to develop a therapy. Many molecules are under investigation for potential beneficial effects, in order to become a suitable drug to alleviate some of the symptoms of the disease. In previous work from Ketley *et al.* (unpublished), a subset of kinase inhibitors reduced the number of foci in the nucleus of DM1 cells. CDK12 was identified as the main target of these compounds. One hypothesis to explain the elimination of foci by inhibition of CDK12 is that CDK12 is required to maintain the accumulation of mutant transcripts via its presence in nuclear foci, and thus, inhibition of CDK12 would cause the dispersion of foci. An alternative hypothesis involves transcriptional inhibition. CDK12 phosphorylates the CTD of the RNA polymerase II and it is involved in the elongation phase of transcription (Ko et al., 2001; Bartkowiak et al., 2010). If CDK12 maintains the association of mutant transcripts by its presence in foci, treatment of DM2 cells with these kinase inhibitors should not eliminate foci, as foci do not co-localize with CDK12 in these cells. DM2 cells treated with kinase inhibitors showed an elimination of almost all foci at the highest concentrations of 4 of the compounds: Dinaciclib, SNS-032, AT7519 and PHA-767491 (Figure 3.9). Furthermore, the most potent compounds, Dinaciclib and SNS-032, were not toxic for the cells (Figure 3.11). These results suggest that elimination of foci is likely to be caused by the transcriptional inhibition of the *DMPK*/*CNBP* genes.

Further study of the effect of Dinaciclib and SNS-032 in DM1 and DM2 cells revealed a reduction in the number of SC35 speckles, but not in control cells (Figures 3.12 and 3.13), which suggest that the presence of the repeats is required for the compounds to have an effect on these speckles. It is not clear though whether SC35 speckles are reduced as a consequence of inactivating CDK12, with which it colocalizes, or because CDK12 inhibition causes a certain degree of global transcription inhibition that affects transcription and production

of SC35. Additional research is needed to clarify this point.

MBNL1 is also affected in DM1 and DM2 cells after treatment with Dinaciclib and SNS-032. The nuclear-cytoplasmic distribution of MBNL1 appears to be altered at the concentration threshold of the kinase inhibitors that eliminate foci in both DM1 and DM2 cells. The redistribution can be appreciated in the images shown in Figures 3.15 and 3.16. Analysis of MBNL1 levels in DM1 and DM2 cells after inhibitor treatment revealed reduced amounts in these cells (Figure 3.17 and 3.18). The reduction was statistically significant in both DM1 and DM2 cells treated with both Dinaciclib and SNS-032. Different possible mechanisms could explain the effect of CDK12 inhibition on MBNL1. The main role of CDK12 is to phosphorylate the CTD of RNA polymerase II, however, this protein is a serine/threonine kinase that phosphorylates other targets in the cell. Serine/threonine phosphorylation is one of the most frequently occurring post-translational mechanisms to control protein localization. Phosphorylation can cause an increase or reduction of nuclear import/export, altering the subcellular localization of the targeted protein (Bauer et al., 2015). Thus, MBNL1 could be a target of CDK12, such that inhibition of CDK12 could alter the phosphorylation rates of MBNL1 causing a redistribution of the protein in the cell. Furthermore, the reduced levels of MBNL1 in the cells after treatment with kinase inhibitors could be due to inhibition of CDK12 affecting transcription or translation, or even transcript or protein degradation through an unknown mechanism. This can not be ruled out.

The effect of the kinase inhibitors on MBNL1 levels and distribution, apparently didn't affect the alternative splicing of a subset of genes analysed (Figure 3.24), as treatment of DM2 cells with Dinaciclib didn't restore the alternative splicing of any of the genes. Furthermore, alternative splicing of PHKA1 appeared to have been worsened by Dinaciclib, as the pattern of splicing of this gene was shifted towards a DM-specific pattern of splicing, as reported in the literature. It is possible that downregulation of MBNL1 worsens the alternative splicing and the kinase inhibitors partially rescue it, producing a compensation effect that looks like no effect is being produced on alternative splicing. Clarification of the mechanism through which CDK12 inhibition affects MBNL1 is

needed, and could help us understand why the alternative splicing appears to be unchanged after treatment with Dinaciclib at a concentration above the required to eliminate foci.

Taken together, the results obtained in Chapter 3 show that inhibition of CDK12 affects a subset of genes and proteins involved in myotonic dystrophy, and suggest that foci are eliminated from the nucleus by the kinase inhibitors because inhibition of CDK12 stops the transcription of *DMPK* and *CNBP*. Thus, the repeats are not being transcribed and there is no mutant RNA to form foci. Other genes not investigated in this work are also likely to be downregulated after the inhibition of CDK12.

Further work is required to assess the extent of the transcriptional inhibition caused by the compounds. However, CDK12 appears to be a novel target to develop a therapy. In fact, some of the original kinase inhibitors have been modified in an ongoing project in our lab to improve their selectivity towards CDK12 and the best four ones that eliminated foci in DM1 cells were also tested in DM2 cells, producing similar results (Figure 3.25). More work is needed to convert these molecules into drugs, and their effects need to be tested in mice and humans before being commercialized, but some of them hold potential.

6.3 CRISPR-genome engineering of *CDK12* in DM1 cells

One concern about the use of chemical molecules as drugs are the toxic and off-target effects. The kinase inhibitors that were used in this work to treat DM1 and DM2 cells target CDK12, but they are not 100% selective. In order to determine whether the elimination of foci and other effects produced by the inhibitors are due to the inhibition of CDK12 rather than other CDKs, efforts were made to develop a stable *CDK12* knock out cell line. The original aim was to create a DM2 cell line as well to compare the results of genomic versus chemical inhibition of CDK12, but DM2 cells do not form colonies easily so the efforts were focused on the DM1 cell line.

Initially, the pCDK12_HDR plasmid, containing the Neomycin resistance gene flanked by 1 Kb homology arms, was going to be used to knock in this antibiotic resistance gene into exon 1 of *CDK12*, by homologous recombination after the cleavage of the target site by the Cas9. Nonetheless, experiments described in chapter 5 were performed prior to this experiment, and the outcome of those experiments made me re-design the *CDK12* knock out project. Efforts were put to knock out *CDK12* by introducing a hygromycin linear marker into exon 1 of the gene, to disrupt the production of the protein.

A gRNA with high activity *in vitro* and the Cas9 protein, in the form of a plasmid, and along with the hygromycin linear fragment, were transfected into DM1 cells to target exon 1 of *CDK12*. From the colonies that survived the antibiotic selection process, around 45% showed mutations in any of the alleles in the target site, and around 23% of them had integrated the Hygromycin linear marker, a 2.1 Kb linear fragment, into the target site by NHEJ-mediated knock in. The high efficiency of the CRISPR system and the NHEJ-mediated knock in match those reported in the literature, as some of the reasons why CRISPR is being widely used now a days are the ease in the design of the experiment and the high efficiency of the system (Thurtle-Schmidt and Lo, 2018).

Analysis of the effect of the mutations in the CDK12 mRNA and protein levels revealed surprising results (Figures 4.7 and 4.10). Most of the colonies analysed showed increased mRNA levels, with a 2-4-fold change difference. Three of the colonies showed a marked reduction in the level of CDK12 mRNA, compared to control. However, there is no clear relationship between the mutations present in the target site and the changes in mRNA levels, as similar mutations, closely located in the gene, cause either downregulation or upregulation of the mRNA levels (i.e. 3-D4 and 3-E4 in Table 4.4 and Figure 4.7). Furthermore, analysis of the cDNA of the clones revealed no exon skipping events (Figure 4.8) in any of the clones analysed. This was studied as CRISPR has been reported to induce exon skipping as an unexpected outcome (Mou et al., 2017). Nonetheless, surprisingly some of the clones which showed a 3 Kb band when the genomic DNA was amplified (Figure 4.5), didn't present that same band when the cDNA was amplified (Figure 4.8 **B**)). The hygromycin linear fragment, which contains

its own promoter and transcriptional and translational pause sites, may be stopping the transcription from the *CDK12* promoter in that allele. In this case, the template could not be amplified with the reverse primer used.

Analysis of the CDK12 protein levels in some of the colonies showed no correlation between the mRNA and protein levels. Several studies have focused on the analysis of the correlation between mRNA and protein levels, mostly in yeast but also in bacteria and mammalian cells. Results from these studies show that mRNA and protein levels do not follow a normal distribution, and in most of the cases the correlation is weak (Gygi et al., 1999; Griffin et al., 2002; Greenbaum et al., 2003; Tian et al., 2004). Different reasons can explain the apparent poor correlation, such as post-transcriptional parameters, post-translational modifications, or experimental error. The physical properties of the transcript, regulatory proteins, siRNAs, the half-life of the newly created protein, etc, can influence in the mRNA or protein levels. In the case of the colonies that have integrated the hygromycin marker, the promoter or the stop codon of this fragment is probably also influencing the levels of mRNA.

Results from the protein levels revealed that CDK12 had not been successfully knocked down, as expected by the introduction of different mutations into the target site. Different explanations are possible here: (1) the cell has more than 2 copies of the gene, and therefore the colonies screened wouldn't be complete knockdowns, even if the mutations identified knockdown one copy of the gene; (2) the mutations introduced in the target site do not produce a change in any amino acid, or there is a substitution for another amino acid of similar properties which does not change the protein; (3) in the case of colonies which have integrated the hygromycin linear marker, the insertion of the fragment was expected to knock out the protein, but in fact, the promoter of the hygromycin gene could be driving the production of CDK12; (4) post-transcriptional or post-translational parameters could be affecting the levels of the protein; (5) translation of the protein could be starting from an ATG downstream the main ATG. From all these possibilities, the last one seems the most plausible, especially after the recent publication of a paper from Makino *et al.*, in which they introduced mutations in the *Gli3* gene using the CRISPR/Cas9 system, but they

reported the presence of the Gli3 protein. The cause of the translation of the protein, despite the introduction of frameshift mutations into the gene, was illegitimate translation (ITL) starting from an ATG downstream the main ATG, and in frame (Makino et al., 2016). In the case of my gene, *CDK12*, there are indeed other three ATG start codons downstream the initial ATG, in frame with this one, which could serve as an starting point for translation. Due to limitations of time I did not pursue further understanding why the protein had not been knocked down. Instead, I analysed the effect of the mutations in the number of foci per cell.

Quantification of the number of foci per cell in the colonies resistant to hygromycin showed a similar number of foci compared to control cells in most of the colonies. Nonetheless, 6 colonies showed a statistically significant increase in the number of foci. These colonies have in common that they have the hygromycin linear fragment integrated into the target site. Some of them also have additional insertions or deletions in the other allele. These results suggest that, even though integration of hygromycin into the other colonies didn't increase the number of foci, in these 6 colonies it is likely to be producing the increase of the mutant transcript. This could be due to the specific location into which the hygromycin marker was integrated, that could be increasing the transcription of the mutant transcript. Another possibility to consider here, though, is that it is possible each clone has a different number of repeats, so this could account for the differences observed in the foci number. However, we know that we see a range of foci numbers in the parental cell line. FISH measures RNA, not DNA, with the probe, and in the laboratory we know that from the most recent repeat length measure for this cell line, the repeat length was reasonably homogeneous. Thus, it is more likely that the difference in numbers of foci in the various clones reflect the extent of repeat expression.

According to Pham *et al.*, integration of selectable marker cassettes (such as the hygromycin linear fragment in this work) in targeted loci can cause unexpected consequences and phenotypes in "knockdown" models, due to the disruption of expression of neighbouring genes (Pham et al., 1996). A huge amount of time would be required to completely study what occurs in each of the CRISPR-

modified clones. However, what can be concluded from the results obtained in this chapter is that even though CDK12 was not properly knocked out, in general, the CRISPR/Cas9 system worked in the DM1 cell line, as lots of colonies were generated. The gRNA used to target exon 1 of *CDK12* could direct Cas9 to the target site with a high efficiency, and the screening of the colonies revealed many different types of mutations in most of the colonies screened. As mutation of *CDK12* exon 1 does not appear to have a big impact in the production of the protein, additional strategies should be used to completely knock out the gene. One of these strategies that could not be performed due to time limitations was to design two gRNAs, upstream and downstream of an exon of the kinase domain of the protein, with the aim to eliminate this exon and disrupt the kinase domain, which in turn should completely inactivate the protein. Production of this disrupted protein could give us a clearer insight of the role of *CDK12* in myotonic dystrophy.

6.3.1 Role of CDK12 in myotonic dystrophy

Even though further work will be needed to completely characterize the effect of mutations introduced into each of the CRISPR-generated colonies, results from this chapter regarding the number of foci, in conjunction with those from the previous chapter, start to give an idea of the functions of CDK12 on hallmarks of the disease. An additional experiment contrary to what was proposed at the beginning of Chapter 4 was performed. Transient overexpression of CDK12 in DM1 and DM2 cells and quantification of the number of foci revealed a dose-response increase in the number of foci, with increasing concentrations transfected of the plasmid that contains the CDK12 cDNA (Figure 4.17). These results suggest that CDK12 is directly involved in the transcription of the mutant transcript in both DM1 and DM2 cells, and its overexpression is very likely to be causing higher levels of CTD phosphorylation for RNA polymerase II, boosting the transcription of a subset of genes. Consequently, inhibition of CDK12 by the kinase inhibitors, and therefore reduced phosphorylation rates of the CTD domain of the RNA polymerase II, may be having an indirect effect on the production of

SC35 and MBNL1, as chemical inhibition of CDK12 caused a decrease in the number of SC35 speckles (Figure 3.12) and in the MBNL1 protein levels (Figures 3.17 and 3.18). What we don't know yet is whether there are other kinases involved in the transcription of these genes that could become a target to develop a therapeutic strategy around.

6.4 Use of CRISPR/Cas9 system to engineer a disease model for myotonic dystrophy

There are currently many things that we don't understand about myotonic dystrophy, especially about the molecular pathomechanism. The availability of accurate models could help increase our understanding of the disease, and in fact there are several different disease models available, which were discussed in the Introduction. However, none of the models completely recapitulates all aspects of the disease. Creating such disease models is not easy as large CTG or CCTG repeats are difficult to clone due to instability and problems that arise during PCR amplification of long repetitive sequences (Ashizawa et al., 1993). Different technologies have been used to develop existing disease models. With the discovery and development of the CRISPR system as a genome engineering tool, it has already been used in the DM field to eliminate repeats from DM cells (van Agtmaal et al., 2017; Provenzano et al., 2017)].

For this thesis, I attempted to use the CRISPR system to engineer an inducible disease model for DM2. The aim was not to clone large repeats into cells, but instead to engineer the *CNBP* gene of DM2-patients fibroblasts to gain control of the expression of the gene (which already contains 3000-4000 CCTG repeats in intron 1), by introducing an inducible promoter taken from the pTetOne vector. The disease models currently available have a much smaller number of repeats, thus developing a disease model with 3000-4000 CCTG repeats would be very useful for further investigation of the pathomechanism. Also, the expression of the mutant transcript could be controlled by modifying the dose of doxycycline.

Initial transfection of DM2 cells with a plasmid containing a gRNA with high activity *in vitro* targeting intron 1 of *CNBP* and the Cas9 protein, a repair template with homology arms flanking the inducible promoter, and a Hygromycin linear marker that confers resistance to Hygromycin-B, led to the generation of several colonies. Characterization of the mutations in these colonies revealed that the Cas9 cleavage efficiency was around 72% in the selected clones and the hygromycin linear marker had been integrated into the target site at an efficiency of around 37%, whereas no homologous recombination had occurred to introduce the inducible promoter present in the repair template (Table 5.4). These results were not unexpected, as efficiency of the NHEJ repair pathway is much higher than homologous recombination (Mao et al., 2008). However, the NHEJ-mediated knock-in of the hygromycin linear marker at the cleavage site was unexpected. At the time these results were obtained, HDR-strategies were the main approach for genome engineering with CRISPR to produce precise modifications of the genome. However, reports about NHEJ-mediated knock-in subsequently started to appear (He et al., 2016; Morita et al., 2017), with reports matching what had happened in the experiments reported in this thesis. These results were considered for the next round of transfections.

Analysis of the effect of mutations on the transcription of the CCTG repeats revealed lower numbers of foci per nucleus in almost all of the colonies screened compared to the paternal KAGO cell line (Figures 5.8 and 5.9). Many of them contain the hygromycin linear marker integrated into the target site, along with other mutations in the other allele, although it is not known which allele has which. Despite both alleles being affected in some cases, none of the mutations completely eliminates foci. Colonies with a similar number of foci compared to control cells may have the hygromycin linear marker integrated in the wild type allele, and the mutations in the mutant allele may not compromise transcription. This has yet to be determined, and as previously said, it is unlikely, but could be possible that colonies differ in repeat size, which could account for the differences observed. This can not be completely ruled out.

A second round of transfection with the plasmid containing the gRNA and Cas9, the hygromycin linear marker, and the inducible promoter as a 2.1Kb

fragment (without homology arms), produced colonies with similar mutations to those obtained in the previous transfection (Table 5.9). Again, the hygromycin linear marker was integrated in some of the samples. Nonetheless, in two of the colonies, the inducible promoter was integrated into the target site. However, NHEJ-mediated knock in can cause indels in the junction sites, or degradation of the ends of the exogenous DNA. This happened in one of the colonies that contained the inducible promoter, in which the last 50 bp of the fragment were missing. The other colony had integrated the inducible promoter in the wrong orientation, due to the absence of homology arms to allow the integration in the right orientation. None of these colonies could be used for the doxycycline induction, as they were not the desired clones. Based on the results obtained after the two rounds of transfections it can be concluded that the efficiency of the Cas9 is quite high when complexed with a gRNA with high activity *in vitro*. Regarding the integration of the exogenous DNA, the efficiency of the NHEJ-mediated knock-in is much higher than homologous recombination. Thus, in order to integrate the inducible promoter, a strategy based on NHEJ instead of transfecting the cells with a donor plasmid would appear to be preferable.

Further attempts to optimize the integration of the inducible promoter led me to employ a repair template containing the inducible promoter flanked by microhomology arms of around 20-25 bp and flanked by the gRNA sequence, including the PAM sequence. The purpose of this repair template is for the Cas9 to cleave this plasmid and release the inducible promoter flanked by the micro-homology arms as a linear fragment, which should then be integrated into the target site in frame. The reason to adopt this strategy was that papers in the literature claimed efficient and precise integration of exogenous DNA into the genome of zebrafish, cells and animals using this system and this type of repair template (Nakade et al., 2014; Hisano et al., 2015; Sakuma et al., 2016). Apparently, the presence of the arms allows homologous recombination of the DNA cassette, with precise modifications and reasonable efficiency, compared to the efficiency of homologous recombination using repair templates with longer homology arms. Nonetheless, efficacy of this approach in DM2 cells could not be established because the cells didn't form colonies after 4 weeks under hy-

gromycin selection.

Overall the CRISPR system has worked quite well, but the desired clone has not been generated. These experiments require the screening of large numbers of colonies, and due to the characteristics of the DM2 cells used, this was not possible to obtain. For future experiments, the pCas9_gRNA_ZNF9 plasmid has been shown to work well, with high cleavage efficiencies. However, it may be beneficial to transfect a different DM2 cell line. Generation of the attempted disease model will be very valuable and could open the door to the generation of disease models for other disorders using a similar approach, as thus far, no one has attempted to gain control of a certain gene by introducing an inducible promoter in the genome using the CRISPR/Cas9 system.

6.5 Final Conclusion

In summary, experiments conducted during this thesis have allowed me to achieve most of the aims proposed at the beginning. Despite the technical issues encountered, the co-localization of foci with CDK12 and SC35 speckles has been directly studied in DM1 cells and inferred in DM2 cells. Investigation of CDK12 inhibitors as potential anti-DM drugs in both DM1 and DM2 cells has led to the production of interesting results regarding different hallmarks of DM. In conjunction with results from studying the localization of foci in the nucleus of the cells, I have proposed a mechanism by which the drugs eliminate foci from DM cells. I've also gained knowledge about the role of *CDK12* in myotonic dystrophy and its influence over certain hallmarks of the disease.

The generation of DM1 and DM2 cell lines using the CRISPR/Cas9 system has not been completely successful, but the CRISPR experiments have worked every time from a technical point of view, as many colonies have been generated with mutations in both *CDK12* exon 1 and intron 1 of *CNBP*. More colonies need to be generated and screened due to the low efficiency of the NHEJ-mediated knock in. All the plasmids and the parameters to generate relevant colonies are now optimised and can be used for future transfections.

The successful elimination of foci by the kinase inhibitors tested here and the discovery of *CDK12* as a potential therapeutic target open the door to the development of CDK12 inhibitors as new molecules for myotonic dystrophy therapy. Furthermore, attempts to create a disease model using the CRISPR system and an inducible promoter could set precedent for the creation of future disease models using this approach.

References

Adereth, Y., Dammai, V., Kose, N., Li, R., and Hsu, T. (2005). RNA-dependent integrin alpha3 protein localization regulated by the muscleblind-like protein mlp1. *Nat Cell Biol*, 7(12):1240–7.

Alwazzan, M., Newman, E., Hamshere, M. G., and Brook, J. D. (1999). Myotonic dystrophy is associated with a reduced level of RNA from the DMWD allele adjacent to the expanded repeat. *Hum Mol Genet*, 8(8):1491–7.

Amack, J. D., Paguio, A. P., and Mahadevan, M. S. (1999). Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model. *Hum Mol Genet*, 8(11):1975–84.

Arambula, J. F., Ramisetty, S. R., Baranger, A. M., and Zimmerman, S. C. (2009). A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. *Proc Natl Acad Sci U S A*, 106(38):16068–73.

Armas, P., Nasif, S., and Calcaterra, N. B. (2008). Cellular nucleic acid binding protein binds G-rich single-stranded nucleic acids and may function as a nucleic acid chaperone. *J Cell Biochem*, 103(3):1013–36.

Arsenault, M. E., Prevost, C., Lescault, A., Laberge, C., Puymirat, J., and Mathieu, J. (2006). Clinical characteristics of myotonic dystrophy type 1 patients with small CTG expansions. *Neurology*, 66(8):1248–50.

Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., van Blitterswijk, M. M., Jansen-West, K., Paul, J. W., r., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2013). Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron*, 77(4):639–46.

Ashizawa, T., Anvret, M., Baiget, M., Barcelo, J. M., Brunner, H., Cobo, A. M., Dallapiccola, B., Fenwick, R. G., J., Grandell, U., Harley, H., and et al. (1994). Characteristics of intergenerational contractions of the CTG repeat in myotonic dystrophy. *Am J Hum Genet*, 54(3):414–23.

Ashizawa, T., Dubel, J. R., and Harati, Y. (1993). Somatic instability of CTG repeat in myotonic dystrophy. *Neurology*, 43(12):2674–8.

Ashizawa, T. and Epstein, H. F. (1991). Ethnic distribution of myotonic dystrophy gene. *Lancet*, 338(8767):642–3.

Auer, T. O., Duroure, K., De Cian, A., Concordet, J. P., and Del Bene, F. (2014). Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res*, 24(1):142–53.

Bachinski, L. L., Udd, B., Meola, G., Sansone, V., Bassez, G., Eymard, B., Thornton, C. A., Moxley, R. T., Harper, P. S., Rogers, M. T., Jurkat-Rott, K., Lehmann-Horn, F., Wieser, T., Gamez, J., Navarro, C., Bottani, A., Kohler, A., Shriver, M. D., Sallinen, R., Wessman, M., Zhang, S., Wright, F. A., and Krahe, R. (2003). Confirmation of the type 2 myotonic dystrophy (CCTG)n expansion mutation in patients with proximal myotonic myopathy/proximal myotonic dystrophy of different European origins: a single shared haplotype indicates an ancestral founder effect. *Am J Hum Genet*, 73(4):835–48.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819):1709–12.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–97.

Bartkowiak, B. and Greenleaf, A. L. (2015). Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex. *J Biol Chem*, 290(3):1786–95.

Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., Adelman, K., Lis, J. T., and Greenleaf, A. L. (2010). CDK12 is a transcrip-

tion elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev*, 24(20):2303–16.

Batra, R., Charizanis, K., and Swanson, M. S. (2010). Partners in crime: bidirectional transcription in unstable microsatellite disease. *Hum Mol Genet*, 19(R1):R77–82.

Batra, R., Nelles, D. A., Pirie, E., Blue, S. M., Marina, R. J., Wang, H., Chaim, I. A., Thomas, J. D., Zhang, N., Nguyen, V., Aigner, S., Markmiller, S., Xia, G., Corbett, K. D., Swanson, M. S., and Yeo, G. W. (2017). Elimination of toxic microsatellite repeat expansion rna by RNA-targeting Cas9. *Cell*, 170(5):899–912 e10.

Bauer, N. C., Doetsch, P. W., and Corbett, A. H. (2015). Mechanisms regulating protein localization. *Traffic*, 16(10):1039–61.

Berul, C. I., Maguire, C. T., Aronovitz, M. J., Greenwood, J., Miller, C., Gehrmann, J., Housman, D., Mendelsohn, M. E., and Reddy, S. (1999). DMPK dosage alterations result in atrioventricular conduction abnormalities in a mouse myotonic dystrophy model. *J Clin Invest*, 103(4):R1–7.

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., Cimermancic, P., Ule, J., and Peterlin, B. M. (2011). The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev*, 25(20):2158–72.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151(Pt 8):2551–61.

Botta, A., Caldarola, S., Vallo, L., Bonifazi, E., Fruci, D., Gullotta, F., Massa, R., Novelli, G., and Loreni, F. (2006). Effect of the [CCTG]n repeat expansion on ZNF9 expression in myotonic dystrophy type ii (DM2). *Biochim Biophys Acta*, 1762(3):329–34.

Botta, A., Rossi, G., Marcaurelio, M., Fontana, L., D'Apice, M. R., Brancati, F., Massa, R., D, G. M., Sangiuolo, F., and Novelli, G. (2017). Identification and

characterization of 5' CCG interruptions in complex DMPK expanded alleles. *Eur J Hum Genet*, 25(2):257–261.

Braida, C., Stefanatos, R. K., Adam, B., Mahajan, N., Smeets, H. J., Niel, F., Goizet, C., Arveiler, B., Koenig, M., Lagier-Tourenne, C., Mandel, J. L., Faber, C. G., de Die-Smulders, C. E., Spaans, F., and Monckton, D. G. (2010). Variant CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients. *Hum Mol Genet*, 19(8):1399–412.

Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., and et al. (1992). 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*, 69(2):385.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321(5891):960–4.

Calcaterra, N. B., Armas, P., Weiner, A. M., and Borgognone, M. (2010). CNBP: a multifunctional nucleic acid chaperone involved in cell death and proliferation control. *IUBMB Life*, 62(10):707–14.

Carango, P., Noble, J. E., Marks, H. G., and Funanage, V. L. (1993). Absence of myotonic dystrophy protein kinase (DMPK) mRNA as a result of a triplet repeat expansion in myotonic dystrophy. *Genomics*, 18(2):340–8.

Cardani, R., Giagnacovo, M., Rossi, G., Renna, L. V., Bugiardini, E., Pizzamiglio, C., Botta, A., and Meola, G. (2014). Progression of muscle histopathology but not of spliceopathy in myotonic dystrophy type 2. *Neuromuscul Disord*, 24(12):1042–53.

Carrasco, M., Canicio, J., Palacin, M., Zorzano, A., and Kaliman, P. (2002). Identification of intracellular signaling pathways that induce my-

otonic dystrophy protein kinase expression during myogenesis. *Endocrinology*, 143(8):3017–25.

Cass, D., Hotchko, R., Barber, P., Jones, K., Gates, D. P., and Berglund, J. A. (2011). The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements. *BMC Mol Biol*, 12:20.

Chakraborty, M., Sellier, C., Ney, M., Pascal, V., Charlet-Berguerand, N., Artero, R., and Llamusi, B. (2018). Daunorubicin reduces MBNL1 sequestration caused by CUG-repeat expansion and rescues cardiac dysfunctions in a Drosophila model of myotonic dystrophy. *Dis Model Mech*, 11(4).

Chaouch, S., Mouly, V., Goyenvalle, A., Vulin, A., Mamchaoui, K., Negroni, E., Di Santo, J., Butler-Browne, G., Torrente, Y., Garcia, L., and Furling, D. (2009). Immortalized skin fibroblasts expressing conditional myod as a renewable and reliable source of converted human muscle cells to assess therapeutic strategies for muscular dystrophies: validation of an exon-skipping approach to restore dystrophin in duchenne muscular dystrophy cells. *Hum Gene Ther*, 20(7):784–90.

Chapman, J. R., Taylor, M. R., and Boulton, S. J. (2012). Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell*, 47(4):497–510.

Charlet, B. N., Savkur, R. S., Singh, G., Philips, A. V., Grice, E. A., and Cooper, T. A. (2002). Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol Cell*, 10(1):45–53.

Chebel, S., Ben Hamda, K., Boughammoura, A., Frih Ayed, M., and Ben Farhat, M. H. (2005). [cardiac involvement in Steinert's myotonic dystrophy]. *Rev Neurol (Paris)*, 161(10):932–9.

Chen, F., Pruett-Miller, S. M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., Collingwood, T. N., Frodin, M., and Davis, G. D. (2011). High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods*, 8(9):753–5.

Chen, G., Masuda, A., Konishi, H., Ohkawara, B., Ito, M., Kinoshita, M., Kiyama, H., Matsuura, T., and Ohno, K. (2016). Phenylbutazone induces expression of MBNL1 and suppresses formation of MBNL1-CUG RNA foci in a mouse model of myotonic dystrophy. *Sci Rep*, 6:25317.

Chen, H. H., Wang, Y. C., and Fann, M. J. (2006). Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation. *Mol Cell Biol*, 26(7):2736–45.

Chen, W., Wang, Y., Abe, Y., Cheney, L., Udd, B., and Li, Y. P. (2007). Haploinsuffciency for Znf9 in Znf9+/- mice is associated with multiorgan abnormalities resembling myotonic dystrophy. *J Mol Biol*, 368(1):8–17.

Chila, R., Guffanti, F., and Damia, G. (2016). Role and therapeutic potential of CDK12 in human cancers. *Cancer Treat Rev*, 50:83–88.

Childs-Disney, J. L., Parkesh, R., Nakamori, M., Thornton, C. A., and Disney, M. D. (2012). Rational design of bioactive, modularly assembled aminoglycosides targeting the RNA that causes myotonic dystrophy type 1. *ACS Chem Biol*, 7(12):1984–93.

Childs-Disney, J. L., Stepniak-Konieczna, E., Tran, T., Yildirim, I., Park, H., Chen, C. Z., Hoskins, J., Southall, N., Marugan, J. J., Patnaik, S., Zheng, W., Austin, C. P., Schatz, G. C., Sobczak, K., Thornton, C. A., and Disney, M. D. (2013). Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. *Nat Commun*, 4:2044.

Chong, J. J., Forte, E., and Harvey, R. P. (2014). Developmental origins and lineage descendants of endogenous adult cardiac progenitor cells. *Stem Cell Res*, 13(3 Pt B):592–614.

Cleary, J. D. and Ranum, L. P. (2013). Repeat-associated non-ATG (RAN) translation in neurological disease. *Hum Mol Genet*, 22(R1):R45–51.

Coarelli, G., Diallo, A., Thion, M. S., Rinaldi, D., Calvas, F., Boukbiza, O. L., Tataru, A., Charles, P., Tranchant, C., Marelli, C., Ewenczyk, C., Tchikviladze, M., Monin, M. L., Carlander, B., Anheim, M., Brice, A., Mochel, F.,

Tezenas du Montcel, S., Humbert, S., and Durr, A. (2017). Low cancer prevalence in polyglutamine expansion diseases. *Neurology*, 88(12):1114–1119.

Cohen, P. (2002). Protein kinases–the major drug targets of the twenty-first century? *Nat Rev Drug Discov*, 1(4):309–15.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121):819–23.

Coonrod, L. A., Nakamori, M., Wang, W., Carrell, S., Hilton, C. L., Bodner, M. J., Siboni, R. B., Docter, A. G., Haley, M. M., Thornton, C. A., and Berglund, J. A. (2013). Reducing levels of toxic RNA with small molecules. *ACS Chem Biol*, 8(11):2528–37.

Cooper, T. A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. *Cell*, 136(4):777–93.

Cristea, S., Freyvert, Y., Santiago, Y., Holmes, M. C., Urnov, F. D., Gregory, P. D., and Cost, G. J. (2013). In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol Bioeng*, 110(3):871–80.

Dansithong, W., Paul, S., Comai, L., and Reddy, S. (2005). MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1. *J Biol Chem*, 280(7):5773–80.

Das, M., Moxley, R. T., r., Hilbert, J. E., Martens, W. B., Letren, L., Greene, M. H., and Gadalla, S. M. (2012). Correlates of tumor development in patients with myotonic dystrophy. *J Neurol*, 259(10):2161–6.

Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H., and Housman, D. E. (1997). Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc Natl Acad Sci U S A*, 94(14):7388–93.

Day, J. W., Ricker, K., Jacobsen, J. F., Rasmussen, L. J., Dick, K. A., Kress, W., Schneider, C., Koch, M. C., Beilman, G. J., Harrison, A. R., Dalton, J. C., and

Ranum, L. P. (2003). Myotonic dystrophy type 2: molecular, diagnostic and clinical spectrum. *Neurology*, 60(4):657–64.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340):602–7.

Douniol, M., Jacquette, A., Cohen, D., Bodeau, N., Rachidi, L., Angeard, N., Cuisset, J. M., Vallee, L., Eymard, B., Plaza, M., Heron, D., and Guile, J. M. (2012). Psychiatric and cognitive phenotype of childhood myotonic dystrophy type 1. *Dev Med Child Neurol*, 54(10):905–11.

Dunn, K. W., Kamocka, M. M., and McDonald, J. H. (2011). A practical guide to evaluating colocalization in biological microscopy. *Am J Physiol Cell Physiol*, 300(4):C723–42.

Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T., and MacKinnon, R. (2002). X-ray structure of a ClC chloride channel at 3.0 A reveals the molecular basis of anion selectivity. *Nature*, 415(6869):287–94.

Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., and Brook, J. D. (2002). 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*, 11(7):805–14.

Fu, Y. H., Friedman, D. L., Richards, S., Pearlman, J. A., Gibbs, R. A., Pizzuti, A., Ashizawa, T., Perryman, M. B., Scarlato, G., Fenwick, R. G., J., and et al. (1993). Decreased expression of myotonin-protein kinase messenger RNA and protein in adult form of myotonic dystrophy. *Science*, 260(5105):235–8.

Fu, Y. H., Pizzuti, A., Fenwick, R. G., J., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., de Jong, P., and et al. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science*, 255(5049):1256–8.

Fugier, C., Klein, A. F., Hammer, C., Vassilopoulos, S., Ivarsson, Y., Toussaint, A., Tosch, V., Vignaud, A., Ferry, A., Messaddeq, N., Kokunai, Y., Tsuburaya, R., de la Grange, P., Dembele, D., Francois, V., Precigout, G., Boulade-Ladame, C., Hummel, M. C., Lopez de Munain, A., Sergeant, N., Laquerriere, A., Thibault, C., Deryckere, F., Auboeuf, D., Garcia, L., Zimmermann, P., Udd, B., Schoser, B., Takahashi, M. P., Nishino, I., Bassez, G., Laporte, J., Furling, D., and Charlet-Berguerand, N. (2011). Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nat Med*, 17(6):720–5.

Furling, D., Doucet, G., Langlois, M. A., Timchenko, L., Belanger, E., Cossette, L., and Puymirat, J. (2003). Viral vector producing antisense RNA restores myotonic dystrophy myoblast functions. *Gene Ther*, 10(9):795–802.

Gambardella, S., Rinaldi, F., Lepore, S. M., Viola, A., Loro, E., Angelini, C., Vergani, L., Novelli, G., and Botta, A. (2010). Overexpression of microRNA-206 in the skeletal muscle from myotonic dystrophy type 1 patients. *J Transl Med*, 8:48.

Garcia-Lopez, A., Llamusi, B., Orzaez, M., Perez-Paya, E., and Artero, R. D. (2011). In vivo discovery of a peptide that prevents CUG-RNA hairpin formation and reverses RNA toxicity in myotonic dystrophy models. *Proc Natl Acad Sci U S A*, 108(29):11866–71.

Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A. H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320):67–71.

Goers, E. S., Purcell, J., Voelker, R. B., Gates, D. P., and Berglund, J. A. (2010). MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res*, 38(7):2467–84.

Gomes-Pereira, M., Foiry, L., Nicole, A., Huguet, A., Junien, C., Munnich, A., and Gourdon, G. (2007). CTG trinucleotide repeat "big jumps": large expansions, small mice. *PLoS Genet*, 3(4):e52.

Gossen, M., Bonin, A. L., and Bujard, H. (1993). Control of gene activity in higher eukaryotic cells by prokaryotic regulatory elements. *Trends Biochem Sci*, 18(12):471–5.

Gossen, M. and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A*, 89(12):5547–51.

Greco, S., Perfetti, A., Fasanaro, P., Cardani, R., Capogrossi, M. C., Meola, G., and Martelli, F. (2012). Deregulated microRNAs in myotonic dystrophy type 2. *PLoS One*, 7(6):e39732.

Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol*, 4(9):117.

Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L., and Aebersold, R. (2002). Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. *Mol Cell Proteomics*, 1(4):323–33.

Groenen, P. J., Wansink, D. G., Coerwinkel, M., van den Broek, W., Jansen, G., and Wieringa, B. (2000). Constitutive and regulated modes of splicing produce six major myotonic dystrophy protein kinase (DMPK) isoforms with distinct properties. *Hum Mol Genet*, 9(4):605–16.

Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*, 19(3):1720–30.

Haft, D. H., Selengut, J., Mongodin, E. F., and Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*, 1(6):e60.

Hale, M. A., Richardson, J. I., Day, R. C., McConnell, O. L., Arboleda, J., Wang, E. T., and Berglund, J. A. (2018). An engineered RNA binding protein with improved splicing regulation. *Nucleic Acids Res*.

Hamshere, M. G., Harley, H., Harper, P., Brook, J. D., and Brookfield, J. F. (1999). Myotonic dystrophy: the correlation of (CTG) repeat length in leucocytes with age at onset is significant only for patients with small expansions. *J Med Genet*, 36(1):59–61.

Hamshere, M. G., Newman, E. E., Alwazzan, M., Athwal, B. S., and Brook, J. D. (1997). Transcriptional abnormality in myotonic dystrophy affects DMPK but not neighboring genes. *Proc Natl Acad Sci U S A*, 94(14):7394–9.

Hao, M., Akrami, K., Wei, K., De Diego, C., Che, N., Ku, J. H., Tidball, J., Graves, M. C., Shieh, P. B., and Chen, F. (2008). Muscleblind-like 2 (Mbnl2) -deficient mice as a model for myotonic dystrophy. *Dev Dyn*, 237(2):403–10.

Harper, P. S. (2001). Myotonic dystrophy third edition. *Major problems in neurology.*, 37:ALL.

He, X., Tan, C., Wang, F., Wang, Y., Zhou, R., Cui, D., You, W., Zhao, H., Ren, J., and Feng, B. (2016). Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res*, 44(9):e85.

Herrendorff, R., Faleschini, M. T., Stiefvater, A., Erne, B., Wiktorowicz, T., Kern, F., Hamburger, M., Potterat, O., Kinter, J., and Sinnreich, M. (2016). Identification of plant-derived alkaloids with therapeutic potential for myotonic dystrophy type i. *J Biol Chem*, 291(33):17165–77.

Hisano, Y., Sakuma, T., Nakade, S., Ohga, R., Ota, S., Okamoto, H., Yamamoto, T., and Kawahara, A. (2015). Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Sci Rep*, 5:8841.

Ho, T. H., Bundman, D., Armstrong, D. L., and Cooper, T. A. (2005). Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy. *Hum Mol Genet*, 14(11):1539–47.

Ho, T. H., Charlet, B. N., Poulos, M. G., Singh, G., Swanson, M. S., and Cooper, T. A. (2004). Muscleblind proteins regulate alternative splicing. *EMBO J*, 23(15):3103–12.

Hofmann-Radvanyi, H., Lavedan, C., Rabes, J. P., Savoy, D., Duros, C., Johnson, K., and Junien, C. (1993). Myotonic dystrophy: absence of CTG enlarged transcript in congenital forms, and low expression of the normal allele. *Hum Mol Genet*, 2(8):1263–6.

Holt, I., Mittal, S., Furling, D., Butler-Browne, G. S., Brook, J. D., and Morris, G. E. (2007). Defective mRNA in myotonic dystrophy accumulates at the periphery of nuclear splicing speckles. *Genes Cells*, 12(9):1035–48.

Hoskins, J. W., Ofori, L. O., Chen, C. Z., Kumar, A., Sobczak, K., Nakamori, M., Southall, N., Patnaik, S., Marugan, J. J., Zheng, W., Austin, C. P., Disney, M. D., Miller, B. L., and Thornton, C. A. (2014). Lomofungin and dilomofungin: inhibitors of MBNL1-CUG RNA binding with distinct cellular effects. *Nucleic Acids Res*, 42(10):6591–602.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., and Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 31(9):827–32.

Huichalaf, C., Schoser, B., Schneider-Gold, C., Jin, B., Sarkar, P., and Timchenko, L. (2009). Reduction of the rate of protein translation in patients with myotonic dystrophy 2. *J Neurosci*, 29(28):9042–9.

IDMC (2000). New nomenclature and dna testing guidelines for myotonic dystrophy type 1 (dm1). the international myotonic dystrophy consortium (idmc). *Neurology*, 54(6):1218–21.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol*, 169(12):5429–33.

Jahromi, A. H., Nguyen, L., Fu, Y., Miller, K. A., Baranger, A. M., and Zimmerman, S. C. (2013). A novel CUG(exp).MBNL1 inhibitor with therapeutic potential for myotonic dystrophy type 1. *ACS Chem Biol*, 8(5):1037–43.

Jansen, G., Groenen, P. J., Bachner, D., Jap, P. H., Coerwinkel, M., Oerlemans, F., van den Broek, W., Gohlsch, B., Pette, D., Plomp, J. J., Molenaar, P. C., Nederhoff, M. G., van Echteld, C. J., Dekker, M., Berns, A., Hameister, H., and Wieringa, B. (1996). Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. *Nat Genet*, 13(3):316–24.

Jansen, R., Embden, J. D., Gaastra, W., and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, 43(6):1565–75.

Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., and Liu, Q. (2005). Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. *Genes Dev*, 19(14):1674–9.

Jiang, H., Mankodi, A., Swanson, M. S., Moxley, R. T., and Thornton, C. A. (2004). Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum Mol Genet*, 13(24):3079–88.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816–21.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *Elife*, 2:e00471.

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., and Doudna, J. A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 343(6176):1247997.

Junghans, R. P. (2009). Dystrophia myotonia: why focus on foci? *Eur J Hum Genet*, 17(5):543–53.

Kanadia, R. N., Johnstone, K. A., Mankodi, A., Lungu, C., Thornton, C. A., Esson, D., Timmers, A. M., Hauswirth, W. W., and Swanson, M. S.

(2003). A muscleblind knockout model for myotonic dystrophy. *Science*, 302(5652):1978–80.

Kanadia, R. N., Shin, J., Yuan, Y., Beattie, S. G., Wheeler, T. M., Thornton, C. A., and Swanson, M. S. (2006). Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. *Proc Natl Acad Sci U S A*, 103(31):11748–53.

Ketley, A., Chen, C. Z., Li, X., Arya, S., Robinson, T. E., Granados-Riveron, J., Udosen, I., Morris, G. E., Holt, I., Furling, D., Chaouch, S., Haworth, B., Southall, N., Shinn, P., Zheng, W., Austin, C. P., Hayes, C. J., and Brook, J. D. (2014). 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*, 23(6):1551–62.

Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N., and Ishiura, S. (2004). Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum Mol Genet*, 13(5):495–507.

Kino, Y., Washizu, C., Oma, Y., Onishi, H., Nezu, Y., Sasagawa, N., Nukina, N., and Ishiura, S. (2009). MBNL and CELF proteins regulate alternative splicing of the skeletal muscle chloride channel CLCN1. *Nucleic Acids Res*, 37(19):6477–90.

Klesert, T. R., Cho, D. H., Clark, J. I., Maylie, J., Adelman, J., Snider, L., Yuen, E. C., Soriano, P., and Tapscott, S. J. (2000). Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy. *Nat Genet*, 25(1):105–9.

Ko, T. K., Kelly, E., and Pines, J. (2001). CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J Cell Sci*, 114(Pt 14):2591–603.

Kohoutek, J. and Blazek, D. (2012). Cyclin K goes with Cdk12 and Cdk13. *Cell Div*, 7:12.

Korade-Mirnics, Z., Tarleton, J., Servidei, S., Casey, R. R., Gennarelli, M., Pegoraro, E., Angelini, C., and Hoffman, E. P. (1999). Myotonic dystrophy: tissue-specific effect of somatic CTG expansions on allele-specific DMAHP/SIX5 expression. *Hum Mol Genet*, 8(6):1017–23.

Kouranova, E., Forbes, K., Zhao, G., Warren, J., Bartels, A., Wu, Y., and Cui, X. (2016). CRISPRs for optimal targeting: Delivery of CRISPR components as DNA, RNA, and protein into cultured cells and single-cell embryos. *Hum Gene Ther*, 27(6):464–75.

Krahe, R., Ashizawa, T., Abbruzzese, C., Roeder, E., Carango, P., Giacanelli, M., Funanage, V. L., and Siciliano, M. J. (1995). Effect of myotonic dystrophy trinucleotide repeat expansion on DMPK transcription and processing. *Genomics*, 28(1):1–14.

Krishnan, A. V. and Kiernan, M. C. (2006). Axonal function and activity-dependent excitability changes in myotonic dystrophy. *Muscle Nerve*, 33(5):627–36.

Krol, J., Fiszer, A., Mykowska, A., Sobczak, K., de Mezer, M., and Krzyzosiak, W. J. (2007). Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Mol Cell*, 25(4):575–86.

Kuyumcu-Martinez, N. M., Wang, G. S., and Cooper, T. A. (2007). Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Mol Cell*, 28(1):68–78.

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

Lam, L. T., Pham, Y. C., Nguyen, T. M., and Morris, G. E. (2000). Characterization of a monoclonal antibody panel shows that the myotonic dystrophy protein kinase, DMPK, is expressed almost exclusively in muscle and heart. *Hum Mol Genet*, 9(14):2167–73.

Langlois, M. A., Lee, N. S., Rossi, J. J., and Puymirat, J. (2003). Hammerhead ribozyme-mediated destruction of nuclear foci in myotonic dystrophy myoblasts. *Mol Ther*, 7(5 Pt 1):670–80.

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

Lee, J. E. and Cooper, T. A. (2009). Pathogenic mechanisms of myotonic dystrophy. *Biochem Soc Trans*, 37(Pt 6):1281–6.

Lima, W. F., Nichols, J. G., Wu, H., Prakash, T. P., Migawa, M. T., Wyrzykiewicz, T. K., Bhat, B., and Crooke, S. T. (2004). Structural requirements at the catalytic site of the heteroduplex substrate for human RNase H1 catalysis. *J Biol Chem*, 279(35):36317–26.

Lin, X., Miller, J. W., Mankodi, A., Kanadia, R. N., Yuan, Y., Moxley, R. T., Swanson, M. S., and Thornton, C. A. (2006). Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum Mol Genet*, 15(13):2087–97.

Liquori, C. L., Ikeda, Y., Weatherspoon, M., Ricker, K., Schoser, B. G., Dalton, J. C., Day, J. W., and Ranum, L. P. (2003). Myotonic dystrophy type 2: human founder haplotype and evolutionary conservation of the repeat tract. *Am J Hum Genet*, 73(4):849–62.

Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*, 293(5531):864–7.

Llagostera, E., Catalucci, D., Marti, L., Liesa, M., Camps, M., Ciaraldi, T. P., Kondo, R., Reddy, S., Dillmann, W. H., Palacin, M., Zorzano, A., Ruiz-Lozano, P., Gomis, R., and Kaliman, P. (2007). Role of myotonic dystrophy protein kinase (DMPK) in glucose homeostasis and muscle insulin action. *PLoS One*, 2(11):e1134.

Loew, R., Heinz, N., Hampf, M., Bujard, H., and Gossen, M. (2010). Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol*, 10:81.

Lombardo, V. A., Armas, P., Weiner, A. M., and Calcaterra, N. B. (2007). In vitro embryonic developmental phosphorylation of the cellular nucleic acid binding protein by cAMP-dependent protein kinase, and its relevance for biochemical activities. *FEBS J*, 274(2):485–97.

Lusis, A. J., Rajavashisth, T. B., Klisak, I., Heinzmann, C., Mohandas, T., and Sparkes, R. S. (1990). Mapping of the gene for CNBP, a finger protein, to human chromosome 3q13.3-q24. *Genomics*, 8(2):411–4.

Luu, L. M., Nguyen, L., Peng, S., Lee, J., Lee, H. Y., Wong, C. H., Hergenrother, P. J., Chan, H. Y., and Zimmerman, S. C. (2016). A potent inhibitor of protein sequestration by expanded triplet (CUG) repeats that shows phenotypic improvements in a drosophila model of myotonic dystrophy. *ChemMedChem*, 11(13):1428–35.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., and et al. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science*, 255(5049):1253–5.

Mahadevan, M. S., Yadava, R. S., Yu, Q., Balijepalli, S., Frenzel-McCardell, C. D., Bourne, T. D., and Phillips, L. H. (2006). Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy. *Nat Genet*, 38(9):1066–70.

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., and Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, 1:7.

Makino, S., Fukumura, R., and Gondo, Y. (2016). Illegitimate translation causes

unexpected gene expression from on-target out-of-frame alleles created by crispr-cas9. *Sci Rep*, 6:39608.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, 339(6121):823–6.

Malumbres, M., Harlow, E., Hunt, T., Hunter, T., Lahti, J. M., Manning, G., Morgan, D. O., Tsai, L. H., and Wolgemuth, D. J. (2009). Cyclin-dependent kinases: a family portrait. *Nat Cell Biol*, 11(11):1275–6.

Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M., and Thornton, C. A. (2000). Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science*, 289(5485):1769–73.

Mankodi, A., Takahashi, M. P., Jiang, H., Beck, C. L., Bowers, W. J., Moxley, R. T., Cannon, S. C., and Thornton, C. A. (2002). Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol Cell*, 10(1):35–44.

Mankodi, A., Urbinati, C. R., Yuan, Q. P., Moxley, R. T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M. S., and Thornton, C. A. (2001). Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum Mol Genet*, 10(19):2165–70.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science*, 298(5600):1912–34.

Mao, Z., Bozzella, M., Seluanov, A., and Gorbunova, V. (2008). Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)*, 7(10):1765–71.

Margolis, J. M., Schoser, B. G., Moseley, M. L., Day, J. W., and Ranum, L. P. (2006). DM2 intronic expansions: evidence for CCUG accumulation without flanking sequence or effects on ZNF9 mRNA processing or protein expression. *Hum Mol Genet*, 15(11):1808–15.

Mateos-Aierdi, A. J., Goicoechea, M., Aiastui, A., Fernandez-Torron, R., Garcia-Puga, M., Matheu, A., and Lopez de Munain, A. (2015). Muscle wasting in myotonic dystrophies: a model of premature aging. *Front Aging Neurosci*, 7:125.

Matynia, A., Ng, C. H., Dansithong, W., Chiang, A., Silva, A. J., and Reddy, S. (2010). Muscleblind1, but not Dmpk or Six5, contributes to a complex phenotype of muscular and motivational deficits in mouse models of myotonic dystrophy. *PLoS One*, 5(3):e9857.

Maurage, C. A., Udd, B., Ruchoux, M. M., Vermersch, P., Kalimo, H., Krahe, R., Delacourte, A., and Sergeant, N. (2005). Similar brain tau pathology in DM2/PROMM and DM1/Steinert disease. *Neurology*, 65(10):1636–8.

McVey, M. and Lee, S. E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet*, 24(11):529–38.

Meola, G. (2000). Clinical and genetic heterogeneity in myotonic dystrophies. *Muscle Nerve*, 23(12):1789–99.

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

Michalowski, S., Miller, J. W., Urbinati, C. R., Paliouras, M., Swanson, M. S., and Griffith, J. (1999). Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucleic Acids Res*, 27(17):3534–42.

Miller, J. W., Urbinati, C. R., Teng-Umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., and Swanson, M. S. (2000). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J*, 19(17):4439–48.

Mjelle, R., Hegre, S. A., Aas, P. A., Slupphaug, G., Drablos, F., Saetrom, P.,

and Krokan, H. E. (2015). Cell cycle regulation of human DNA repair and chromatin remodeling genes. *DNA Repair (Amst)*, 30:53–67.

Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*, 60(2):174–82.

Mojica, F. J., Diez-Villasenor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol*, 36(1):244–6.

Morales, F., Couto, J. M., Higham, C. F., Hogg, G., Cuenca, P., Braida, C., Wilson, R. H., Adam, B., del Valle, G., Brian, R., Sittenfeld, M., Ashizawa, T., Wilcox, A., Wilcox, D. E., and Monckton, D. G. (2012). Somatic instability of the expanded ctg triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. *Hum Mol Genet*, 21(16):3558–67.

Mori, K., Weng, S. M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H. A., Cruts, M., Van Broeckhoven, C., Haass, C., and Edbauer, D. (2013). The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science*, 339(6125):1335–8.

Morita, H., Taimatsu, K., Yanagi, K., and Kawahara, A. (2017). Exogenous gene integration mediated by genome editing technologies in zebrafish. *Bioengineered*, 8(3):287–295.

Mou, H., Smith, J. L., Peng, L., Yin, H., Moore, J., Zhang, X. O., Song, C. Q., Sheel, A., Wu, Q., Ozata, D. M., Li, Y., Anderson, D. G., Emerson, C. P., Sontheimer, E. J., Moore, M. J., Weng, Z., and Xue, W. (2017). CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol*, 18(1):108.

Mulders, S. A., van den Broek, W. J., Wheeler, T. M., Croes, H. J., van Kuik-Romeijn, P., de Kimpe, S. J., Furling, D., Platenburg, G. J., Gourdon, G.,

Thornton, C. A., Wieringa, B., and Wansink, D. G. (2009). Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. *Proc Natl Acad Sci U S A*, 106(33):13915–20.

Murmann, A. E., Gao, Q. Q., Putzbach, W. E., Patel, M., Bartom, E. T., Law, C. Y., Bridgeman, B., Chen, S., McMahon, K. M., Thaxton, C. S., and Peter, M. E. (2018). Small interfering RNAs based on huntingtin trinucleotide repeats are highly toxic to cancer cells. *EMBO Rep*, 19(3).

Musova, Z., Mazanec, R., Krepelova, A., Ehler, E., Vales, J., Jaklova, R., Prochazka, T., Koukal, P., Marikova, T., Kraus, J., Havlovicova, M., and Sedlacek, Z. (2009). Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene. *Am J Med Genet A*, 149A(7):1365–74.

Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., Daimon, T., Sezutsu, H., Yamamoto, T., Sakuma, T., and Suzuki, K. T. (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat Commun*, 5:5560.

Nakamori, M., Taylor, K., Mochizuki, H., Sobczak, K., and Takahashi, M. P. (2016). Oral administration of erythromycin decreases RNA toxicity in myotonic dystrophy. *Ann Clin Transl Neurol*, 3(1):42–54.

Nakatani, K., Hagihara, S., Goto, Y., Kobori, A., Hagihara, M., Hayashi, G., Kyo, M., Nomura, M., Mishima, M., and Kojima, C. (2005). Small-molecule ligand induces nucleotide flipping in (CAG)n trinucleotide repeats. *Nat Chem Biol*, 1(1):39–43.

Nguyen, L., Lee, J., Wong, C. H., and Zimmerman, S. C. (2014). Small molecules that target the toxic RNA in myotonic dystrophy type 2. *ChemMedChem*, 9(11):2455–62.

Novelli, G., Gennarelli, M., Zelano, G., Pizzuti, A., Fattorini, C., Caskey, C. T., and Dallapiccola, B. (1993). Failure in detecting mRNA transcripts from the mutated allele in myotonic dystrophy muscle. *Biochem Mol Biol Int*, 29(2):291–7.

Orengo, J. P., Chambon, P., Metzger, D., Mosier, D. R., Snipes, G. J., and Cooper, T. A. (2008). Expanded CTG repeats within the DMPK 3' UTR causes severe skeletal muscle wasting in an inducible mouse model for myotonic dystrophy. *Proc Natl Acad Sci U S A*, 105(7):2646–51.

Orlando, S. J., Santiago, Y., DeKelver, R. C., Freyvert, Y., Boydston, E. A., Moehle, E. A., Choi, V. M., Gopalan, S. M., Lou, J. F., Li, J., Miller, J. C., Holmes, M. C., Gregory, P. D., Urnov, F. D., and Cost, G. J. (2010). Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res*, 38(15):e152.

Otten, A. D. and Tapscott, S. J. (1995). Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc Natl Acad Sci U S A*, 92(12):5465–9.

Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006). The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation*, 74(2-3):65–80.

Pellizzoni, L., Lotti, F., Maras, B., and Pierandrei-Amaldi, P. (1997). Cellular nucleic acid binding protein binds a conserved region of the 5' UTR of Xenopus laevis ribosomal protein mRNAs. *J Mol Biol*, 267(2):264–75.

Perbellini, R., Greco, S., Sarra-Ferraris, G., Cardani, R., Capogrossi, M. C., Meola, G., and Martelli, F. (2011). Dysregulation and cellular mislocalization of specific miRNAs in myotonic dystrophy type 1. *Neuromuscul Disord*, 21(2):81–8.

Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V., Lee, Y. L., Guschin, D. Y., Rupniewski, I., Waite, A. J., Carpenito, C., Carroll, R. G., Orange, J. S., Urnov, F. D., Rebar, E. J., Ando, D., Gregory, P. D., Riley, J. L., Holmes, M. C., and June, C. H. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol*, 26(7):808–16.

Pettersson, O. J., Aagaard, L., Jensen, T. G., and Damgaard, C. K. (2015).

Molecular mechanisms in DM1 - a focus on foci. *Nucleic Acids Res*, 43(4):2433–41.

Pham, C. T., MacIvor, D. M., Hug, B. A., Heusel, J. W., and Ley, T. J. (1996). Long-range disruption of gene expression by a selectable marker cassette. *Proc Natl Acad Sci U S A*, 93(23):13090–5.

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

Phylactou, L. A. (2004). Repair of myotonic dystrophy protein kinase (DMPK) transcripts by trans-splicing ribozymes. *Methods Mol Biol*, 252:373–83.

Phylactou, L. A., Darrah, C., and Wood, M. J. (1998). Ribozyme-mediated trans-splicing of a trinucleotide repeat. *Nat Genet*, 18(4):378–81.

Pinto, B. S., Saxena, T., Oliveira, R., Mendez-Gomez, H. R., Cleary, J. D., Denes, L. T., McConnell, O., Arboleda, J., Xia, G., Swanson, M. S., and Wang, E. T. (2017). Impeding transcription of expanded microsatellite repeats by deactivated Cas9. *Mol Cell*, 68(3):479–490 e5.

Poulos, M. G., Batra, R., Li, M., Yuan, Y., Zhang, C., Darnell, R. B., and Swanson, M. S. (2013). Progressive impairment of muscle regeneration in muscleblind-like 3 isoform knockout mice. *Hum Mol Genet*, 22(17):3547–58.

Provenzano, C., Cappella, M., Valaperta, R., Cardani, R., Meola, G., Martelli, F., Cardinali, B., and Falcone, G. (2017). CRISPR/Cas9-mediated deletion of CTG expansions recovers normal phenotype in myogenic cells derived from myotonic dystrophy 1 patients. *Mol Ther Nucleic Acids*, 9:337–348.

Purcell, J., Oddo, J. C., Wang, E. T., and Berglund, J. A. (2012). Combinatorial mutagenesis of MBNL1 zinc fingers elucidates distinct classes of regulatory events. *Mol Cell Biol*, 32(20):4155–67.

Puymirat, J., Giguere, Y., Mathieu, J., and Bouchard, J. P. (2009). Intergenerational contraction of the CTG repeats in 2 families with myotonic dystrophy type 1. *Neurology*, 73(24):2126–7.

Querido, E., Gallardo, F., Beaudoin, M., Menard, C., and Chartrand, P. (2011). Stochastic and reversible aggregation of mRNA with expanded CUG-triplet repeats. *J Cell Sci*, 124(Pt 10):1703–14.

Raheem, O., Olufemi, S. E., Bachinski, L. L., Vihola, A., Sirito, M., Holmlund-Hampf, J., Haapasalo, H., Li, Y. P., Udd, B., and Krahe, R. (2010). Mutant (CCTG)n expansion causes abnormal expression of zinc finger protein 9 (ZNF9) in myotonic dystrophy type 2. *Am J Pathol*, 177(6):3025–36.

Rajavashisth, T. B., Taylor, A. K., Andalibi, A., Svenson, K. L., and Lusis, A. J. (1989). Identification of a zinc finger protein that binds to the sterol regulatory element. *Science*, 245(4918):640–3.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, 8(11):2281–308.

Ranum, L. P. and Cooper, T. A. (2006). RNA-mediated neuromuscular disorders. *Annu Rev Neurosci*, 29:259–77.

Rau, F., Freyermuth, F., Fugier, C., Villemin, J. P., Fischer, M. C., Jost, B., Dembele, D., Gourdon, G., Nicole, A., Duboc, D., Wahbi, K., Day, J. W., Fujimura, H., Takahashi, M. P., Auboeuf, D., Dreumont, N., Furling, D., and Charlet-Berguerand, N. (2011). Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat Struct Mol Biol*, 18(7):840–5.

Reddy, S., Smith, D. B., Rich, M. M., Leferovich, J. M., Reilly, P., Davis, B. M., Tran, K., Rayburn, H., Bronson, R., Cros, D., Balice-Gordon, R. J., and Housman, D. (1996). Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. *Nat Genet*, 13(3):325–35.

Redman, J. B., Fenwick, R. G., J., Fu, Y. H., Pizzuti, A., and Caskey, C. T. (1993). Relationship between parental trinucleotide GCT repeat length and severity of myotonic dystrophy in offspring. *JAMA*, 269(15):1960–5.

Reeks, J., Naismith, J. H., and White, M. F. (2013). CRISPR interference: a structural perspective. *Biochem J*, 453(2):155–66.

Ricker, K., Koch, M. C., Lehmann-Horn, F., Pongratz, D., Otto, M., Heine, R., and Moxley, R. T., r. (1994). Proximal myotonic myopathy: a new dominant disorder with myotonia, muscle weakness, and cataracts. *Neurology*, 44(8):1448–52.

Rzuczek, S. G., Colgan, L. A., Nakai, Y., Cameron, M. D., Furling, D., Yasuda, R., and Disney, M. D. (2017). Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat Chem Biol*, 13(2):188–193.

Rzuczek, S. G., Southern, M. R., and Disney, M. D. (2015). Studying a drug-like, RNA-focused small molecule library identifies compounds that inhibit RNA toxicity in myotonic dystrophy. *ACS Chem Biol*, 10(12):2706–15.

Sabouri, L. A., Mahadevan, M. S., Narang, M., Lee, D. S., Surh, L. C., and Korneluk, R. G. (1993). Effect of the myotonic dystrophy (DM) mutation on mRNA levels of the DM gene. *Nat Genet*, 4(3):233–8.

Sabourin, L. A., Tamai, K., Narang, M. A., and Korneluk, R. G. (1997). Overexpression of 3'-untranslated region of the myotonic dystrophy kinase cDNA inhibits myoblast differentiation in vitro. *J Biol Chem*, 272(47):29626–35.

Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K. T., and Yamamoto, T. (2016). MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat Protoc*, 11(1):118–33.

Saleh-Gohari, N. and Helleday, T. (2004). Conservative homologous recombination preferentially repairs DNA double-strand breaks in the s phase of the cell cycle in human cells. *Nucleic Acids Res*, 32(12):3683–8.

Salisbury, E., Schoser, B., Schneider-Gold, C., Wang, G. L., Huichalaf, C., Jin, B., Sirito, M., Sarkar, P., Krahe, R., Timchenko, N. A., and Timchenko, L. T. (2009). Expression of RNA CCUG repeats dysregulates translation and degradation of proteins in myotonic dystrophy 2 patients. *Am J Pathol*, 175(2):748–62.

Sammons, M. A., Antons, A. K., Bendjennat, M., Udd, B., Krahe, R., and Link, A. J. (2010). ZNF9 activation of IRES-mediated translation of the hu-

man ODC mRNA is decreased in myotonic dystrophy type 2. *PLoS One*, 5(2):e9301.

Sansone, V. A., Brigonzi, E., Schoser, B., Villani, S., Gaeta, M., De Ambroggi, G., Bandera, F., De Ambroggi, L., and Meola, G. (2013). The frequency and severity of cardiac involvement in myotonic dystrophy type 2 (DM2): long-term outcomes. *Int J Cardiol*, 168(2):1147–53.

Santoro, M., Masciullo, M., Pietrobono, R., Conte, G., Modoni, A., Bianchi, M. L., Rizzo, V., Pomponi, M. G., Tasca, G., Neri, G., and Silvestri, G. (2013). Molecular, clinical, and muscle studies in myotonic dystrophy type 1 (DM1) associated with novel variant CCG expansions. *J Neurol*, 260(5):1245–57.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res*, 39(21):9275–82.

Sarkar, P. S., Appukuttan, B., Han, J., Ito, Y., Ai, C., Tsai, W., Chai, Y., Stout, J. T., and Reddy, S. (2000). Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts. *Nat Genet*, 25(1):110–4.

Sashital, D. G., Wiedenheft, B., and Doudna, J. A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell*, 46(5):606–15.

Sauer, B. and Henderson, N. (1988). Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci U S A*, 85(14):5166–70.

Sauer, B. and Henderson, N. (1990). Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol*, 2(5):441–9.

Savkur, R. S., Philips, A. V., and Cooper, T. A. (2001). Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat Genet*, 29(1):40–7.

Schonig, K., Weber, T., Frommig, A., Wendler, L., Pesold, B., Djandji, D., Bujard, H., and Bartsch, D. (2012). Conditional gene expression systems in the transgenic rat brain. *BMC Biol*, 10:77.

Schoser, B. and Timchenko, L. (2010). Myotonic dystrophies 1 and 2: complex diseases with complex mechanisms. *Curr Genomics*, 11(2):77–90.

Sen, S., Talukdar, I., and Webster, N. J. (2009). SRp20 and CUG-BP1 modulate insulin receptor exon 11 alternative splicing. *Mol Cell Biol*, 29(3):871–80.

Seznec, H., Lia-Baldini, A. S., Duros, C., Fouquet, C., Lacroix, C., Hofmann-Radvanyi, H., Junien, C., and Gourdon, G. (2000). Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum Mol Genet*, 9(8):1185–94.

Siboni, R. B., Nakamori, M., Wagner, S. D., Struck, A. J., Coonrod, L. A., Harriott, S. A., Cass, D. M., Tanner, M. K., and Berglund, J. A. (2015). Actinomycin D specifically reduces expanded CUG repeat RNA in myotonic dystrophy models. *Cell Rep*, 13(11):2386–2394.

Sioud, M. (2004). Ribozyme- and siRNA-mediated mRNA degradation: a general introduction. *Methods Mol Biol*, 252:1–8.

Soifer, H. S., Rossi, J. J., and Saetrom, P. (2007). MicroRNAs in disease and potential therapeutic applications. *Mol Ther*, 15(12):2070–9.

Spector, D. L. and Lamond, A. I. (2011). Nuclear speckles. *Cold Spring Harb Perspect Biol*, 3(2).

Stein, H. and Hausen, P. (1969). Enzyme from calf thymus degrading the RNA moiety of DNA-RNA hybrids: effect on DNA-dependent RNA polymerase. *Science*, 166(3903):393–5.

Steinberg, H. and Wagner, A. (2008). [hans Steinert: 100 years of myotonic dystrophy]. *Nervenarzt*, 79(8):961–2, 965–70.

Steyaert, J., de Die-Smulders, C., Fryns, J. P., Goossens, E., and Willekens, D. (2000). Behavioral phenotype in childhood type of dystrophia myotonica. *Am J Med Genet*, 96(6):888–9.

Steyaert, J., Umans, S., Willekens, D., Legius, E., Pijkels, E., de Die-Smulders, C., Van den Berghe, H., and Fryns, J. P. (1997). A study of the cognitive and psychological profile in 16 children with congenital or juvenile myotonic dystrophy. *Clin Genet*, 52(3):135–41.

Storbeck, C. J., Drmanic, S., Daniel, K., Waring, J. D., Jirik, F. R., Parry, D. J., Ahmed, N., Sabourin, L. A., Ikeda, J. E., and Korneluk, R. G. (2004). Inhibition of myogenesis in transgenic mice expressing the human DMPK 3'-UTR. *Hum Mol Genet*, 13(6):589–600.

Suominen, T., Bachinski, L. L., Auvinen, S., Hackman, P., Baggerly, K. A., Angelini, C., Peltonen, L., Krahe, R., and Udd, B. (2011). Population frequency of myotonic dystrophy: higher than expected frequency of myotonic dystrophy type 2 (DM2) mutation in Finland. *Eur J Hum Genet*, 19(7):776–82.

Suominen, T., Schoser, B., Raheem, O., Auvinen, S., Walter, M., Krahe, R., Lochmuller, H., Kress, W., and Udd, B. (2008). High frequency of cosegregating CLCN1 mutations among myotonic dystrophy type 2 patients from Finland and Germany. *J Neurol*, 255(11):1731–6.

Swarts, D. C., Mosterd, C., van Passel, M. W., and Brouns, S. J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS One*, 7(4):e35888.

Symington, L. S. and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. *Annu Rev Genet*, 45:247–71.

Taleei, R. and Nikjoo, H. (2013). Biochemical dsb-repair model for mammalian cells in g1 and early s phases of the cell cycle. *Mutat Res*, 756(1-2):206–12.

Taneja, K. L., McCurrach, M., Schalling, M., Housman, D., and Singer, R. H. (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J Cell Biol*, 128(6):995–1002.

Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Huttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proc Natl Acad Sci U S A*, 99(11):7536–41.

Tang, Z. Z., Yarotskyy, V., Wei, L., Sobczak, K., Nakamori, M., Eichinger, K., Moxley, R. T., Dirksen, R. T., and Thornton, C. A. (2012). Muscle weakness in myotonic dystrophy associated with misregulated splicing and altered gating of Ca(v)1.1 calcium channel. *Hum Mol Genet*, 21(6):1312–24.

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

Thornton, C. A., Griggs, R. C., and Moxley, R. T., r. (1994). Myotonic dystrophy with no trinucleotide repeat expansion. *Ann Neurol*, 35(3):269–72.

Thornton, C. A., Wymer, J. P., Simmons, Z., McClain, C., and Moxley, R. T., r. (1997). Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat Genet*, 16(4):407–9.

Thurtle-Schmidt, D. M. and Lo, T. W. (2018). Molecular biology at the cutting edge: A review on CRISPR/Cas9 gene editing for undergraduates. *Biochem Mol Biol Educ*, 46(2):195–205.

Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J. P., Gunter, B., Linseley, P. S., Stoughton, R. B., Aebersold, R., Collins, S. J., Hanlon, W. A., and Hood, L. E. (2004). Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Mol Cell Proteomics*, 3(10):960–9.

Timchenko, L. T., Miller, J. W., Timchenko, N. A., DeVore, D. R., Datar, K. V., Lin, L., Roberts, R., Caskey, C. T., and Swanson, M. S. (1996). Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res*, 24(22):4407–14.

Timchenko, N. A., Iakova, P., Cai, Z. J., Smith, J. R., and Timchenko, L. T. (2001). Molecular basis for impaired muscle differentiation in myotonic dystrophy. *Mol Cell Biol*, 21(20):6927–38.

Timchenko, N. A., Patel, R., Iakova, P., Cai, Z. J., Quan, L., and Timchenko, L. T. (2004). Overexpression of CUG triplet repeat-binding protein, CUGBP1, in mice inhibits myogenesis. *J Biol Chem*, 279(13):13129–39.

Timchenko, N. A., Wang, G. L., and Timchenko, L. T. (2005). RNA CUG-binding protein 1 increases translation of 20-kDa isoform of CCAAT/enhancer-binding protein beta by interacting with the alpha and beta subunits of eukaryotic initiation translation factor 2. *J Biol Chem*, 280(21):20549–57.

Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A. J., Chen, K. C., Scaglione, K. M., Basrur, V., Elenitoba-Johnson, K., Vonsattel, J. P., Louis, E. D., Sutton, M. A., Taylor, J. P., Mills, R. E., Charlet-Berguerand, N., and Paulson, H. L. (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron*, 78(3):440–55.

Tran, H., Gourrier, N., Lemercier-Neuillet, C., Dhaenens, C. M., Vautrin, A., Fernandez-Gomez, F. J., Arandel, L., Carpentier, C., Obriot, H., Eddarkaoui, S., Delattre, L., Van Brussels, E., Holt, I., Morris, G. E., Sablonniere, B., Buee, L., Charlet-Berguerand, N., Schraen-Maschke, S., Furling, D., Behm-Ansmant, I., Branlant, C., Caillet-Boudin, M. L., and Sergeant, N. (2011). Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of Muscleblind-like-1 isoforms. *J Biol Chem*, 286(18):16435–46.

Tsilfidis, C., MacKenzie, A. E., Mettler, G., Barcelo, J., and Korneluk, R. G. (1992). Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. *Nat Genet*, 1(3):192–5.

Turner, C. and Hilton-Jones, D. (2010). The myotonic dystrophies: diagnosis and management. *J Neurol Neurosurg Psychiatry*, 81(4):358–67.

Turner, M. R., Goldacre, R., and Goldacre, M. J. (2013). Reduced cancer incidence in Huntington's disease: record linkage study clue to an evolutionary trade-off? *Clin Genet*, 83(6):588–90.

van Agtmaal, E. L., Andre, L. M., Willemse, M., Cumming, S. A., van Kessel, I. D. G., van den Broek, W., Gourdon, G., Furling, D., Mouly, V., Monckton, D. G., Wansink, D. G., and Wieringa, B. (2017). CRISPR/Cas9-induced (CTGCAG)n repeat instability in the myotonic dystrophy type 1 locus: Implications for therapeutic genome editing. *Mol Ther*, 25(1):24–43.

van der Oost, J., Westra, E. R., Jackson, R. N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol*, 12(7):479–92.

van Gent, D. C., Hoeijmakers, J. H., and Kanaar, R. (2001). Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet*, 2(3):196–206.

Vlasova, I. A., Tahoe, N. M., Fan, D., Larsson, O., Rattenbacher, B., Sternjohn, J. R., Vasdewani, J., Karypis, G., Reilly, C. S., Bitterman, P. B., and Bohjanen, P. R. (2008). Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1. *Mol Cell*, 29(2):263–70.

Wang, E. T., Cody, N. A., Jog, S., Biancolella, M., Wang, T. T., Treacy, D. J., Luo, S., Schroth, G. P., Housman, D. E., Reddy, S., Lecuyer, E., and Burge, C. B. (2012). Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell*, 150(4):710–24.

Wang, G. S., Kearney, D. L., De Biasi, M., Taffet, G., and Cooper, T. A. (2007). Elevation of RNA-binding protein CUGBP1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy. *J Clin Invest*, 117(10):2802–11.

Wang, G. S., Kuyumcu-Martinez, M. N., Sarma, S., Mathur, N., Wehrens, X. H., and Cooper, T. A. (2009). PKC inhibition ameliorates the cardiac phenotype in a mouse model of myotonic dystrophy type 1. *J Clin Invest*, 119(12):3797–806.

Wang, J., Pegoraro, E., Menegazzo, E., Gennarelli, M., Hoop, R. C., Angelini, C., and Hoffman, E. P. (1995). Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation. *Hum Mol Genet*, 4(4):599–606.

Wang, L., Li, F., Dang, L., Liang, C., Wang, C., He, B., Liu, J., Li, D., Wu, X., Xu, X., Lu, A., and Zhang, G. (2016). In vivo delivery systems for therapeutic genome editing. *Int J Mol Sci*, 17(5).

Wang, Y. H., Amirhaeri, S., Kang, S., Wells, R. D., and Griffith, J. D. (1994). Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science*, 265(5172):669–71.

Ward, A. J., Rimer, M., Killian, J. M., Dowling, J. J., and Cooper, T. A. (2010). CUGBP1 overexpression in mouse skeletal muscle reproduces features of myotonic dystrophy type 1. *Hum Mol Genet*, 19(18):3614–22.

Warf, M. B., Diegel, J. V., von Hippel, P. H., and Berglund, J. A. (2009a). The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc Natl Acad Sci U S A*, 106(23):9203–8.

Warf, M. B., Nakamori, M., Matthys, C. M., Thornton, C. A., and Berglund, J. A. (2009b). Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc Natl Acad Sci U S A*, 106(44):18551–6.

Westra, E. R., Swarts, D. C., Staals, R. H., Jore, M. M., Brouns, S. J., and van der Oost, J. (2012). The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet*, 46:311–39.

Wheeler, T. M., Leger, A. J., Pandey, S. K., MacLeod, A. R., Nakamori, M., Cheng, S. H., Wentworth, B. M., Bennett, C. F., and Thornton, C. A. (2012). Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature*, 488(7409):111–5.

Wheeler, T. M., Lueck, J. D., Swanson, M. S., Dirksen, R. T., and Thornton, C. A. (2007). Correction of ClC-1 splicing eliminates chloride channelopathy and myotonia in mouse models of myotonic dystrophy. *J Clin Invest*, 117(12):3952–7.

Wheeler, T. M., Sobczak, K., Lueck, J. D., Osborne, R. J., Lin, X., Dirksen, R. T., and Thornton, C. A. (2009). Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science*, 325(5938):336–9.

White, P. J., Anastasopoulos, F., Pouton, C. W., and Boyd, B. J. (2009). Overcoming biological barriers to in vivo efficacy of antisense oligonucleotides. *Expert Rev Mol Med*, 11:e10.

Win, A. K., Perattur, P. G., Pulido, J. S., Pulido, C. M., and Lindor, N. M. (2012). Increased cancer risks in myotonic dystrophy. *Mayo Clin Proc*, 87(2):130–5.

Winchester, C. L., Ferrier, R. K., Sermoni, A., Clark, B. J., and Johnson, K. J. (1999). Characterization of the expression of DMPK and SIX5 in the human eye and implications for pathogenesis in myotonic dystrophy. *Hum Mol Genet*, 8(3):481–92.

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

Wong, C. H., Nguyen, L., Peh, J., Luu, L. M., Sanchez, J. S., Richardson, S. L., Tuccinardi, T., Tsoi, H., Chan, W. Y., Chan, H. Y., Baranger, A. M., Hergenrother, P. J., and Zimmerman, S. C. (2014). Targeting toxic RNAs that cause myotonic dystrophy type 1 (DM1) with a bisamidinium inhibitor. *J Am Chem Soc*, 136(17):6355–61.

Yao, X., Wang, X., Liu, J., Hu, X., Shi, L., Shen, X., Ying, W., Sun, X., Wang, X., Huang, P., and Yang, H. (2017). CRISPR/Cas9 - mediated precise targeted integration in vivo using a double cut donor with short homology arms. *EBioMedicine*, 20:19–26.

Yosef, I., Goren, M. G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res*, 40(12):5569–76.

Yuan, Y., Compton, S. A., Sobczak, K., Stenberg, M. G., Thornton, C. A., Griffith, J. D., and Swanson, M. S. (2007). Muscleblind-like 1 interacts with

RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res*, 35(16):5474–86.

Zeesman, S., Carson, N., and Whelan, D. T. (2002). Paternal transmission of the congenital form of myotonic dystrophy type 1: a new case and review of the literature. *Am J Med Genet*, 107(3):222–6.

Zhang, F., Bodycombe, N. E., Haskell, K. M., Sun, Y. L., Wang, E. T., Morris, C. A., Jones, L. H., Wood, L. D., and Pletcher, M. T. (2017). A flow cytometry-based screen identifies MBNL1 modulators that rescue splicing defects in myotonic dystrophy type i. *Hum Mol Genet*, 26(16):3056–3068.

Zhou, X., Vink, M., Klaver, B., Berkhout, B., and Das, A. T. (2006). Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Ther*, 13(19):1382–90.

Zisman, A., Peroni, O. D., Abel, E. D., Michael, M. D., Mauvais-Jarvis, F., Lowell, B. B., Wojtaszewski, J. F., Hirshman, M. F., Virkamaki, A., Goodyear, L. J., Kahn, C. R., and Kahn, B. B. (2000). Targeted disruption of the glucose transporter 4 selectively in muscle causes insulin resistance and glucose intolerance. *Nat Med*, 6(8):924–8.

Zu, T., Cleary, J. D., Liu, Y., Banez-Coronel, M., Bubenik, J. L., Ayhan, F., Ashizawa, T., Xia, G., Clark, H. B., Yachnis, A. T., Swanson, M. S., and Ranum, L. P. W. (2017). RAN translation regulated by muscleblind proteins in myotonic dystrophy type 2. *Neuron*, 95(6):1292–1305 e5.

Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., Margolis, J., Peterson, M., Markowski, T. W., Ingram, M. A., Nan, Z., Forster, C., Low, W. C., Schoser, B., Somia, N. V., Clark, H. B., Schmechel, S., Bitterman, P. B., Gourdon, G., Swanson, M. S., Moseley, M., and Ranum, L. P. (2011). Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A*, 108(1):260–5.

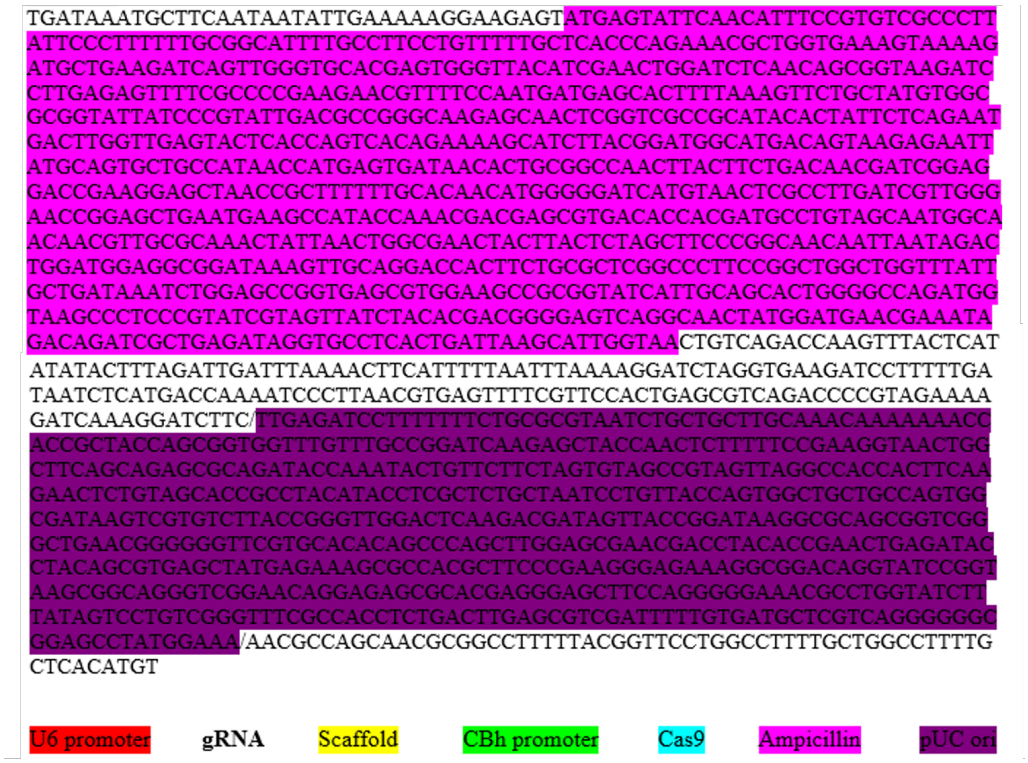
Appendices

Appendix 1 - Plasmids sequence

pCas9_gRNA_CDK12, pCas9_gRNA_ZNF9 and pScramble_gRNA have the same backbone sequence except for the gRNA sequence, represented by the "X" in bold in the next sequence:

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATT
GGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATT
TCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGA
AAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG**XXXXXXXXXXXXXXX**
XXXXXXXGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA
GTGGCACCGAGTCGGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG
TTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGT
AAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAG
GGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAG
TGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTG
CCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC
CATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATT
TTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCG
CCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCC
AATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAA
AAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCG
CCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGAC
GGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAAGGGTTTAAGGGATGGTTGGTTGGT
GGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCACTTTTTTTCAGGTTGGACCGGTGC
CACCATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATG
ACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAA
GAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGT
ACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAA
CCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCG
CCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAG
ATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAA
GAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACC
CCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATC
TATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCC
GACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGA
AAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCA
GACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTG
ATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAA
ACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACC
AGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCAC
CACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGAT

TTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGT
TCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTG
AACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCA
CCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACC
GGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGA
AACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGA
AGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACC
TGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAG
CTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGA
AAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAG
GACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAAC
GCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGA
GGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGA
TCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGG
CGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGT
CCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGA
TCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGAT
AGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGAC
AGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCG
AAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGC
GGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACAC
CCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACC
AGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGA
AGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAA
CGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGC
TGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGAT
AAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGAT
CCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGA
TCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGA
TCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAA
AAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGAT
GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCA
TGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAG
ACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGT
GCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAG
AGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAG
AAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAA
GGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGC
AGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCT
GATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCT
CTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTAC
CTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGT
GGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGA
TCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATC
AGAGAGGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTC
AAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCAC
CCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCG
ACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGTAAGAATTCCTAGA
GCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGC
ATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT
TGGGAAGAGAATAGCAGGCATGCTGGGGAGCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCA
CTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCT
TTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGGGCGCCTGATGCGGTAT
TTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGTCAAAGCAACCATAGTACGCGCCCT
GTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGC
GCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCA
AGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAA
ACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACG
TTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACTCTATCTCGG
GCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGTCTATTGGTTAAAAAATGAGCTGATTTA
ACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTATGGTGCACTCTCAGTAC
AATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTG
ACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTG
TCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTT
ATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCG
CGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCC



The gRNA sequence present in each plasmid is:

- pCas9_gRNA_CDK12: GCTTGGTATCGAAGCACAAG
- pCas9_gRNA_ZNF9: GATTGCATACATAGTACCTA
- pScramble_gRNA: GCACTACCAGAGCTAACTCA

Appendix 2 - Genomic position of the gRNAs targeting *CDK12*

CDK12 exon 1 sequence showing the position of all the gRNAs designed and tested to knockdown *CDK12*.

