University of Glasgow
Institute of Biomedical and Life Sciences
Division of Molecular Genetics

Understanding the pathogenesis of myotonic dystrophy type 1

Christine Haworth

Thesis submitted for the degree of Doctor of Philosophy

February 2008

Abstract

In 1992, the expansion of a CTG repeat within the 3' UTR of the *DMPK* gene was identified as the cause of myotonic dystrophy type 1 (DM1), the most common form of adult onset muscular dystrophy affecting one in 8,000 Caucasians. The array-length determines the severity and age of onset of the disease. This expansion, once transcribed, sequesters a developmental splicing regulator MBNL1 in the form of nuclear foci. The MBNL1 antagonist, CUG-BP1, also a developmental splicing regulator, is raised by some unknown mechanism in DM1 patients. This alteration between the dynamic balance of these two proteins results in the missplicing of genes, which are thought to contribute profoundly to the multisystemic pathology of DM1. The disease is primarily RNA mediated, but how the mutant transcript brings about such pleiotropic effects is not yet clear. Recent research indicates a more direct involvement of CUG-BP1, and other potential factors may be important. To identify the full range of targets and the pathogenic consequences, we sought to mimic the pathogenesis of myotonic dystrophy type 1 with temporal and spatial control: Temporal to reproduce the developmental pathogenesis of the congenital form, and spatial to isolate tissue specific pathology. To do this, we attempted to use the Cre-*lox* system for the conditional expression of an EGFP reporter-linked expanded CUG repeat RNA in the mouse. Expression of the transgene was controlled by Cre excision of a transcriptional stop, placed upstream of the EGFP-expanded repeat open reading frame. The transgenes were constructed and tested successfully, and a normal length repeat transgenic line was established. Unfortunately generation of the expanded repeat line was not successful. The constructs were used to generate cell-culture models of DM1, in both human and murine cells, which mimicked the nuclear foci formation and MBNL1 co-localisation seen in patient cells. Expression of exogenous MBNL1/GFP fusion protein in this model resulted in an increase in the size of foci, indicating that MBNL1 protein is limiting within the cell, and may possibly play a protective role. The murine DM1 cell-culture model was used to investigate the effects of expanded CUG repeat expression on splicing within the transcriptome. The differential effect between 5 and 250 repeat RNA expression using Affymetrix whole transcript and exon arrays was compared. Using whole genome arrays, 6 genes were down-regulated and 128 up-regulated. With exon arrays, 58 genes showed alternative exon usage. Six genes were selected for further bioinformatics analysis: *MtmR4*, which has possible neuromuscular involvement; *Kcnk4*, *Narg1*, *Ttyh1* and *Bptf*, potentially related to brain development; and *Cacna1c*, a promising candidate for heart conductance defects and sudden death.

Table of Contents

1 Introduction **15**

1.1 The nature of myotonic dystrophy 15
1.1.1 Pathology 15
1.1.2 Unstable tandem repeat diseases 16
1.1.3 The myotonic dystrophy mutation 17
1.1.4 Somatic variation and anticipation 20
1.1.5 Pathogenesis 22
1.1.5.1 DMPK haploinsufficiency 22
1.1.5.2 Disruption of chromatin structure 23
1.1.5.3 Gain of function 25
1.1.6 Myotonic dystrophy type 2 28
1.1.7 Alternative splicing: Dynamic control by CUG-BP1 and MBNL1 30

1.2 RNA processing disorders 31
1.2.1 Ocularpharyngeal muscular dystrophy 31
1.2.2 Spinal muscular atrophy 32
1.2.3 Fascioscapulohumeral muscular dystrophy 33
1.2.4 Fragile X syndrome and fragile X-associated tremor/ataxia 344
1.2.5 Huntington disease-like 2 35
1.2.6 Friedreich ataxia 36
1.2.7 Spinocerebellar ataxias 8, 10 and 12 37

1.3 Transgenic mouse models of myotonic dystrophy 38
1.3.1 Gene deficiency and over-expression models 38
1.3.2 Expressed transgenes 38

1.4 Project outline 43

2 Materials and methods **44**

2.1 Materials 44
2.1.1 Tissue culture disposable materials 44
2.1.2 Size markers 44
2.1.3 Constructs 44
2.1.4 Kits 45
2.1.5 Microscopy and photography 45
2.1.6 Membranes and miscellaneous 46
2.1.7 Cell-lines and bacterial hosts 46
2.1.8 Antibodies 47
2.1.9 Equipment 47
2.1.10 Solutions 48
2.1.10.1 General 48
2.1.10.2 Bacterial culture 52
2.1.10.3 Cell culture 52
2.1.10.3.1 General 52
2.1.10.3.2 ES cells 53

2.2 DNA methods 54
2.2.1 Preparation of plasmid DNA 54
2.2.2 Purification of DNA 54
2.2.2.1 Transfection into cultured cells 54
2.2.2.2 Pronuclear injection 54

2.2.3 DNA extraction from cultured cells 54
2.2.4 Preparation of mouse tail lysates 55
2.2.5 Determination of DNA concentration 55
2.2.6 Oligonucleotides and PCR 56
2.2.6.1 Primer pairs used for construct sequences 56
2.2.6.2 PCR Primers and conditions 57
2.2.6.3 Small pool PCR 58
2.2.7 Gel electrophoresis 59
2.2.8 Southern transfer 59
2.2.9 Preparation of radiolabelled DNA 59
2.2.10 Southern hybridisation 60
2.2.11 Cloning techniques 60
2.2.11.1 Endonuclease restriction of DNA 60
2.2.11.2 Filling in of recessed 3' ends 60
2.2.11.3 Ligation 61
2.2.11.4 Transformation of bacterial hosts 61
2.2.11.5 Culture of bacteria 61
2.2.11.6 Glycerol stocks 61
2.2.12 Transfection of DNA into cells 61
2.2.12.1 Transient 61
2.2.12.2 Stable 62
2.2.12.3 Liposome based 62
2.2.12.4 Electroporation 63

2.3 RNA methods 63
2.3.1 Isolation of RNA 63
2.3.2 Determination of RNA quality and concentration 64
2.3.3 Complementary DNA synthesis 64
2.3.4 Fluorescent *in situ* hybridisation 64
2.3.5 Double labelling with ICC and FISH 65
2.3.6 Microarray analysis 65

2.4 Protein methods 65
2.4.1 Isolation of protein 65
2.4.2 Determination of protein concentration 66
2.4.3 Immunodetection of protein 66

2.5 Cell culture methods 67
2.5.1 Feeding cultured cells 67
2.5.2 Subculturing cells 67
2.5.3 Thawing frozen cells 67
2.5.4 Freezing live cells for storage 67
2.5.5 Determination of cell concentration 68
2.5.6 Establishment of primary cell-lines 68
2.5.6.1 Transgenic tail cell-lines 68
2.5.6.2 MEF feeder cells 68
2.5.6.2.1 Preparation 68
2.5.6.2.2 Growth arrest 69

2.6 Pronuclear injection 69
2.6.1 Mouse strains 70

3 Design and generation of a murine model of DM1 pathogenesis 71

3.1 Synopsis 71

3.2 Introduction 71

3.2.1 Conditional expression 73
3.2.2 Transgene structure 76
3.3 Transgene assembly 78
3.4 Component function 79
3.4.1 CMV promoter, EGFP and loxP interference 79
3.4.2 Neomycin 80
3.4.3 Thymidine kinase 81
3.4.4 Transcriptional stop, loxP and Cre excision 82
3.4.5 The repeats 83
3.4.5.1 Further attempts to clone 800 repeats 86
3.4.5.2 Propagation of repeat constructs 90
3.5 The mouse model 92
3.5.1 Targeted integration 92
3.5.2 Random integration 97
3.5.3 Genotyping 98
3.6 The Cre-lox mechanism ex vivo 101
3.7 Discussion 106
3.7.1 Repeats 107
3.7.2 Random integration 108
3.7.3 Cre-lox mechanism and activated fluorescence 110

4 Design and characterisation of a cell culture model of DM1 pathogenesis 112

4.1 Synopsis 112
4.2 Introduction 113
4.3 Validation of an inducible model 114
4.3.1 Cre, Lox and foci 114
4.3.2 Stable cell-lines 117
4.3.3 The constitutively expressed transgene 118
4.3.4 EGFP and RNA foci 120
4.3.5 RNA binding proteins 122
4.3.5.1 CUG-BP1 123
4.3.5.2 MBNL 125
4.3.5.3 Co-localisation between MBNL and foci 127
4.4 Discussion 132
4.4.1 Cell-lines 133
4.4.2 MBNL 134

5 The effects of expanded CUG-repeat expression on mRNA steady state levels and splicing patterns 136

5.1 Synopsis 136
5.2 Background 136
5.3 Microarray analysis 138
5.3.1 Parameter selection 139
5.3.1.1 Dynamics of foci formation 139
5.3.1.2 Fluorescent activated cell sorting (FACS) 145
5.3.1.2.1Cell density 145
5.3.1.2.2Gates and purity 146

5.3.1.2.3Cell viability and RNA integrity 148
5.3.2 Transcript expression array 152
5.3.2.1 Sample preparation for whole transcript arrays 152
5.3.2.2 Results of analysis 153
5.3.3 Exon expression array 172
5.3.3.1 Sample preparation for exon arrays 173
5.3.3.2 Results of analysis 174
5.3.3.2.1MtmR4 178
5.3.3.2.2Kcnk4 180
5.3.3.2.3Narg1 182
5.3.3.2.4Ttyh1 183
5.3.3.2.5Bptf 185
5.3.3.2.6Cacna1c 187
5.4 Expression of known genes 190
5.5 Discussion 193

6 Discussion **197**
6.1 DM1 as an RNA processing disorder 197
6.2 Other RNA processing disorders 198

Dedication

To Mum and Dad:

"Although my nature like a butterfly is to sip the nectar from many flowers, here is the proof I can also gorge on a single bloom."

Acknowledgements

I would like to say a massive thank you to my supervisor, Darren Monckton, for his endless support, guidance and enthusiasm, which was highly contagious, and his patience in reading the (not too) many drafts of this thesis.

I would also like to thank my assessors, Sheila Graham and particularly Peggy Shelbourne for her encouragement during the troughs; John for keeping up the genotyping during my absences; Jing and Pawel at the MBSU; Catherine for sharing the trials and tribulations of ES cell work, and Lynn for showing us how.

A big hug goes to Kevin my lovely husband, who extended the time taken on this project exponentially by giving me two more things to worry about –Adam and Annie (also lovely). Thank you for taking them away for the occasional afternoon, and also for the bucket-loads of emotional support.

Thanks to my friends and colleagues on level 5, past and present –especially Mario (life was never the same after you left), Graham B., Graham H., Colm, Saadia, Yvonne, Claudia, Berit, Nicola, Meera, Asantha, Fernando, Alison, Jon and John, who made the days fly by with their support and good humour.

This wouldn't have been possible without funding – so thank you IBLS, the Muscular Dystrophy Campaign and The Wellcome Trust.

Abbreviations

λ	Wavelength
[α-^{32}P]dCTP	α-^{32}P-labelled 2'-deoxycitidine-5'-triphosphate
°C	Degrees Celsius
AMCA	Aminomethylcoumarin
TRE	Tetracycline responsive element
ANOVA	Analysis of variance
ZNF9	Zinc finger 9
AMPH	Amphiphysin
ANT	Adenine nucleotide transclocator
APP	Amyloid precursor protein
APS	Ammonium persulphate
ARLD	Autosomal recessive limb-girdle dystrophy
b	Base
bp	Base pair
Bptf	Bromodomain PHD finger transcription factor
BSA	Bovine serum albumin
Cacna1c	Calcium channel alpha 1c
CAG	Trinucleotide of cytosine, adenosine and guanine
CCTG	Tetranucleotide of cytosine, cytosine, thymine and guanine
cDNA	Complementary deoxyribonucleic acid
CELF	CUG-BP and ETR3-like factor
CGG	Trinucleotide of cytosine, guanine and guanine
Ci	Curie
Clcn1	Chloride channel 1
CMV	Cytomegalovirus
CNBP	Cellular nucleic acid binding protein
CNS	Central nervous system
Cre	Cre recombinase
CTG	Trinucleotide of cytosine, thymine and guanine
CUG	Trinucleotide of cytosine, uracil and guanine
CUG-BP	CUG-binding protein
Cy3	Cyanine 3
Cy5	Cyanine 5
DAPI	4'-6-Diamidino-2-phenylindole

dATP	2’-Deoxyadenosine-5’-triphosphate
dCTP	2’-Deoxycytidine-5’-triphosphate
DEPC	Diethylpyrocarbonate
dGTP	2’-Deoxyguanosine-5’-triphosphate
DM	Myotonic dystrophy, *dystrophia myotonica*
DMEM	Dulbecco’s modified Eagle medium
DMPK	Dystrophia myotonica protein kinase
DMSO	Dimethylsulphoxide
DMWD	Dystrophia myotonica-containing WD repeat motif
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotidetriphosphate
DRPLA	Dentatorubral pallidoluysian atrophy
DTT	Dithiothreitol
dTTP	2’-Deoxythymidine-5’-triphosphate
E	Embryonic
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetracetic acid
EGFP	Enhanced green fluorescent protein
Elav	Embryonic lethal abnormal vision
ERDA1	Expanded repeat domain CAG/CTG 1
ES cells	Embryonic stem cells
EST	Expressed sequence tag
EtBr	Ethidium bromide
ETR	Elav-type RNA-binding protein
EXP	Triplet repeat expansion proteins
FBS	Foetal bovine serum
FITC	Fluoresceine isothiocyanate isomer I
FMR	Fragile X mental retardation
FMRP	Fragile X mental retardation protein
FRAXA	Fragile X syndrome site A
FRAXE	Fragile X syndrome site E
FRDA	Friedreich ataxia
FRG	FSHD candidate region gene
FSHD	Fascioscapularhumeral muscular dystrophy
g	Gravity acceleration
g	Gram

GAA	Trinucleotide of guanine, adenine and adenine
HD	Huntington disease
HDL	Huntington disease like
HNPCC	Hereditary non-polyposis colorectal cancer
HRP	Horseradish peroxidase
ICC	Immunocytochemistry
InsR	Insulin receptor
IPTG	Isopropylthyo-ß-D-galactoside
JPH3	Junctophilin-3
k	Kilo (10^3)
Kb	Kilo base
KcnK4	Potassium channel K4
KLHL1	Kelch-like 1
l	Litre
LB	Luria Bertani
Lox	Sequence specific Cre binding site
m	Milli (10^{-3})
M	Molar
MAPT	Microtubule associated protein tau
MBNL	Muscleblind like
MCS	Multiple cloning site
MLH	MutL homologue
MMR	Mismatch repair
MOPS	3-(*N*-Morpholino)-propanesulphonic
mRNA	Messenger ribonucleic acid
miRNA	Micro RNA
miRNP	Micro RNA associated ribonuclear protein
MSH	MutS homologue
MtmR4	Myotubularin related 4
mUSF	Mouse upstream stimulatory factor
MW	Molecular weight
N	Nucleotide of adenine, cytosine, guanine or thymidine.
n	Nano (10^{-9})
Narg1	NMDA receptor regulated 1
neo	Neomycin
nt	Nucleotide

OD	Optical density
CpG	Cytosine-phosphate-Guanine
OPMD	Occulopharyngeal muscular dystrophy
ORF	Open reading frame
p	Pico (10^{-12})
PABP	PolyA binding protein
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PKR	Double-stranded RNA-activated protein kinase
PE	Phycoerythrin
PROMM	Proximal myotonic myopathy
PVDF	Polyvinylidene difluoride
RNA	Ribonucleic acid
RNAi	RNA interference
rRNA	Ribosomal ribonucleic acid
RSHL	Radial spokehead-like
RT-PCR	Reverse transcribed PCR
RyR	Ryanodine receptor
SBMA	Spinal and bulbar muscular atrophy
SCA	Spinocerebellar ataxia
SDS	Sodium dodecyl sulphate
SEF	SL3-3 enhancer factor 2
Serca	Sarco/Endoplasmic Reticulum Ca2+-ATPase
SiRNA	Short interfering RNA
SIX5	Sine oculis related homeobox 5
SP-PCR	Small pool polymerase chain reaction
SV40	Simian virus 40
TEMED	*NNN'N'*-Tetramethylethylenediamine
Tg	Transgene
Tnnt	Troponin T
Tnnt2	Cardiac troponin T
Tris	Tris(hydroxymethyl)amino methane
Ttyh1	Tweety homologue 1
U	Unit
UTR	Untranslated region

UV	Ultraviolet
v	Volume
V	Volt
w	Weight
WT	Wild type
X-gal	5-Bromo-4-chloro-3-indolyl-ß-D-galactoside
µ	Micro (10^{-6})

The research reported in this thesis is my own original work, except where otherwise stated, and has not been submitted for any other degree.

Christine Haworth

February 2008

1 Introduction

The birth of a first grandchild is usually an uplifting and joyous event. Often, in families with the myotonic dystrophy type 1 mutation, it is devastating. Grandparents and parents are frequently asymptomatic, unaware that they have the mutation until a congenitally affected grandchild is born.

1.1 The nature of myotonic dystrophy

Myotonic Dystrophy Type 1 (DM1) is the most common form of adult muscular dystrophy with an occurrence of 1 in 8,000 individuals worldwide. The symptoms are pleiotropic, not purely affecting muscle, manifesting clinically as myotonia; progressive muscle weakness and wasting; cardiac conduction defects; cataracts; insulin resistance; premature frontal balding and testicular atrophy in males; reduced fertility in females; and in the more severe form, mental handicap and respiratory distress (Harper, 2001). The severity of symptoms is extremely variable ranging from asymptomatic or only mildly affected adults in old age, to severely affected neonates. Genetically the disease shows autosomal dominance, and anticipation, whereby the severity of a disease is greater, and the age of onset becomes earlier through successive generations of the affected family.

1.1.1 Pathology

DM1 is often diagnosed late in the progression of the disease probably because of the tolerant, uncomplaining nature of the patients, a symptom in itself. Facial weakness including ptosis and wasting of the jaw muscles is a constant and characteristic feature, which can be severe in children with the congenital form of the disease, affecting speech and swallowing. Neck and distal limb weakness is also apparent. The definitive clinical diagnosis, which excludes other forms of muscular dystrophy including fascio-scapulo humeral dystrophy (FSHD), autosomal recessive limb-girdle dystrophy (ARLD) and Becker dystrophy, is of myotonia (36% of patients) in conjunction with progressive muscle wasting. Histologically, muscle defects have been well characterised. Patient muscle tissue shows centralised nuclei; poor fibre packing; ringed fibres, predominance of type 1 fibres and type 1I fibre atrophy (Borg *et al.*, 1987; Tohgi *et al.*, 1997; Vihola *et al.*, 2003). In the brain, cell loss, neuronal inclusion bodies and tau-associated pre-senile neurofibriliary tangles are evident (Kiuchi *et al.*, 1991; Sergeant *et al.*, 2001).

DM1 is not purely a muscular disorder and the patient may present with other seemingly unrelated systemic indications such as cardiac conduction defects –particularly heart block and atrial arrhythmias; cardiomyopathy; aspiration pneumonia; alveolar hypoventilation (breathing difficulties leading to lethargy and daytime sleepiness); minor sensory loss of the peripheral nerves; mild mental retardation –severe in the congenital form; hypersomnia; testicular atrophy; diabetes; cataract; retinal degeneration; ocular hypotonia; skeletal deformities of the jaw and palate; cranial hyperostosis; air sinus enlargement; premature balding and calcifying epithelioma (stony tumour of the face or arms). Mental retardation and skeletal deformities are more pronounced in childhood cases and include talipes (deformities of the foot) and scoliosis (lateral spinal curvature). The causes of death are primarily cardiac and respiratory related, sudden death from cardiac arrhythmia accounting for 29% of deaths (Harper, 2001).

1.1.2 Unstable tandem repeat diseases

Myotonic dystrophy belongs to a growing group of inherited human disorders associated with the expansion of microsatellite repeats, which were until the 1990's thought to be commonplace benign stretches of DNA. The repeats are usually stable at around 30 repeats, but arrays over 40 repeats tend to be unstable (Harper, 2001). Since then, the cause of several diseases has been attributed to the aberrant expansion of these regions, most of which are neurological such as the spinocerebellar ataxias; Huntington disease; spinobulbar muscular atrophy; fragile X syndrome and Fredrich ataxia (Wells *et al.*, 2006). Most of the repeat regions within these genes are highly polymorphic throughout the general population where normal and affected individuals fall into ranges, rather than having a definitive cut-off point. There are defining features shared amongst microsatellite repeat disorders. Repeat expansions show both somatic and germline instability, expanding throughout the lifetime of the individual, which contributes to the pathology, since longer repeat lengths result in more severe symptoms. Genetic anticipation is also observed, whereby the longer the inherited tract length, the earlier the age of onset of the disease and the more severe the clinical presentation (Wells *et al.*, 2006).

Not all CTG repeat tracts cause disease. The *SEF2-1* gene on chromosome 18q21.1 (Breschel *et al.*, 1997) for instance contains a heritable intronic expanding CTG repeat, and the *ERDA1* locus 17q213, a polymorphic CTG repeat array (Schalling *et al.*, 1993; Nakamoto *et al.*, 1997) which lack pathogenic effect. Of those expansions that are symptomatic there are two subclasses of disease, defined by the location of the repeat region relative to the gene. The first contains the repeats within the reading frame of the

gene leading to long stretches of the same amino acid, usually glutamine (in-frame CAG repeats), incorporated into the translated protein. The second subclass comprises those tracts situated within non-coding regions of the gene such as the promoter region; 5'UTR; intron or 3' UTR (Figure 1). In this class, except for those within the promoter region, which are not transcribed, expanded repeats are transcribed but not translated. Myotonic dystrophy falls into the second category in that the expanded RNA is transcribed but not translated (Brook *et al.*, 1992). Additionally, the mutation is situated within the promoter region of a flanking gene where it is not transcribed (Boucher *et al.*, 1995).

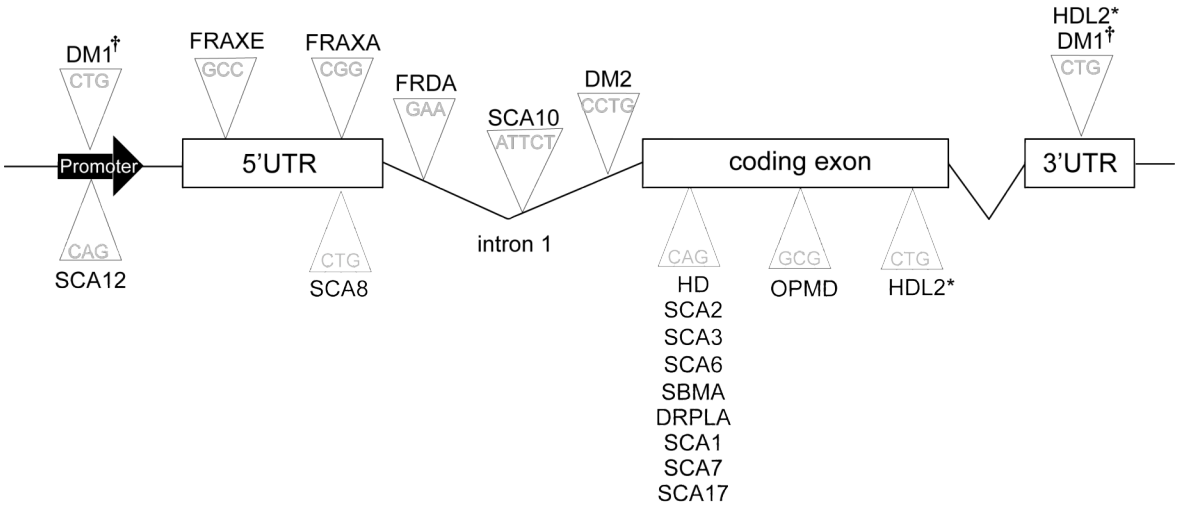


Figure 1 The relative positions of disease associated expanded repeats within a fictitious gene. Schematic diagram. Diseases are classed depending whether they are within coding or untranslated regions. * HDL2 isoforms generated by alternative splicing. † The DM1 mutation is positioned within the 3'UTR of *DMPK* and the promoter region of *SIX5*. Abbreviations: FRAXA, Fragile X Syndrome (Fu *et al.*, 1991); FRAXE, Fragile XE Mental retardation (Chakrabarti *et al.*, 1996; Gu *et al.*, 1996; Gecz *et al.*, 1997); FRDA, Friedreich ataxia (Campuzano *et al.*, 1996); SCA , Spinocerebellar ataxia (Gatchel *et al.*, 2005); HD, Huntington disease (Group, 1993); DM1, Myotonic dystrophy type 1 (Aslanidis *et al.*, 1992; Brook *et al.*, 1992; Buxton *et al.*, 1992; Harley *et al.*, 1992); SBMA, Spinobulbar muscular atrophy (La Spada *et al.*, 1991); DRPLA, Dentatorubralpallidoluysian atrophy (Ikeuchi *et al.*, 1995) HDL2, Huntington disease-like 2 (Margolis *et al.*, 2001; Holmes *et al.*, 2001).

1.1.3 The myotonic dystrophy mutation

In molecular terms, DM1 is associated with the expansion of a CTG repeat in the 3' UTR of the *DM protein kinase* (*DMPK*) gene, and within the promoter region of the flanking gene *SIX5,* which maps to chromosome 19q13.3 (Buxton *et al.*, 1992; Boucher *et al.*, 1995), a gene-dense region of the genome (Figure 2). A correlation exists between the length of the repeat tract, and the age of onset and severity of the disease. Each *DMPK* allele in unaffected individuals has between 5 and 37 CTG repeats, with tracts upwards of

50 repeats becoming disease-associated. Expansions can be divided into three groups (mild; moderate and congenital) depending on the severity of the disease, which also relates to the age of onset. Mild or asymptomatic (late onset) cases have 50 to 100 repeats. Myotonia and weakness are rare in these patients, the most common finding being cataracts. Moderate (adult onset) cases have 100 to 500 repeats. These patients are likely to suffer one or several features of DM1 including myotonia; progressive muscle weakness – becoming debilitating in the fifth and sixth decades; cardiac conduction defects; respiratory failure; cataracts and gonadal and endocrine abnormalities. The symptoms tend to be more acute with increased repeat length. In severe (congenital) cases repeat lengths from 500 up to ~3000 are detected (Harley *et al.*, 1993). Hypotonia is present from birth, adult onset symptoms are also present from early childhood, and there is much more involvement of the central nervous system –congenital patients being mentally retarded, with an average IQ of 66 (Harper, 2001).

The DMPK protein itself belongs to the family of serine-threonine kinases and is expressed highly in skeletal muscle and heart and in lower levels in smooth muscle (Lam *et al.*, 2000). Using antibodies, forms of DMPK protein between 42 to 84 kDa have been identified, expressed primarily in muscle, heart and brain, and at low level ubiquitously (van der Ven *et al.*, 1993; Maeda *et al.*, 1995; Salvatori *et al.*, 1997; Shimokawa *et al.*, 1997; Pham *et al.*, 1998; Beffy *et al.*, 2005). There is some controversy over the full extent of DMPK expression since different sized proteins have been found in different tissues using different antibodies. Using a phage display generated monoclonal antibody panel to DMPK epitopes, Lam *et al.* (2000) reported exclusive expression of the 80KDa form, and only in skeletal muscle, smooth muscle and heart. They suggested that the previously observed size and tissue distribution using first generation antibodies, was probably caused by cross reaction between DMPK related proteins such as the 72KDa myotonic dystrophy related cdc-42 binding kinase (MRCK) (Leung *et al.*, 1998). However, the second-generation antibody specificity could be limited by antigen conformational artefacts of the phage display technique, resulting in some epitopes avoiding recognition, and so the controversy continues.

The *DMPK* gene has 15 exons. Initially, many alternatively spliced *DMPK* transcripts were identified in mouse and humans (Jansen *et al.*, 1992). However further analysis using existing cDNA and EST expression data for both species (Jansen *et al.*, 1992; Fu *et al.*, 1993; Mahadevan *et al.*, 1993; Shaw *et al.*, 1993), and sequencing of these libraries suggested that not all transcripts were equally prevalent between tissue types and splicing

in mouse may be more complex than in humans, (Groenen *et al.*, 2000). Using transgenic mice over-expressing *hDMPK*, Groenen *et al*. made a detailed comparison of the splicing behaviour of human and mouse *DMPK* genes. They showed that most isoforms were expressed in many tissues including heart, skeletal muscle, liver and brain, but the proportions differed, with predominant isoforms specific to heart; skeletal muscle, brain and smooth muscle tissue. Several additional splicing events were specific to mouse. Using RT-PCR, six major alternatively spliced *DMPK* isoforms have been identified in the human and the mouse (Brook *et al.*, 1992; Jansen *et al.*, 1992; Groenen *et al.*, 2000), and a seventh minor isoform in DM1, which includes a novel terminal exon resulting in the excision of the expanded repeat tract (Tiscornia *et al.*, 2000). This form of the transcript has been shown to freely exit the nucleus in DM1 patients, whereas forms with expanded repeats are retained. All isoforms include an N-terminal leucine rich domain –which possibly regulates DMPK activity (in Wells *et al.*, 2006, chapter 5), a serine-threonine kinase domain and a coiled-coil region putatively involved in multimerisation and substrate binding (van Herpen *et al.*, 2005). Variation occurs via alternative splicing between exons 12 and 15 producing 3 different C-termini involved in isoform specific subcellular localisation to either the endoplasmic reticulum, mitochondrial membrane (van Herpen *et al.*, 2005) or cytosol (Wansink *et al.*, 2003). For each of these 3' variations there is a "VSGGG" motif "with" and "without" version, formed from an extension of exon 8 during splicing. The seventh minor isoform, translated from the repeat-free transcript also includes a "VSGGG" motif, and a unique C-terminus which may direct localisation elsewhere within the cell. The function of the "VSGGG" motif is not yet known, but is thought to be a target for conformational regulation (Groenen *et al.*, 2000).

The biological function of DMPK has been widely studied since the discovery of its involvement in DM1, and considerable progress has been made with the identification of activators, substrates and functional domains within the protein. This information allows speculation and hypothesis about the biological significance of the *DMPK* gene product within the cell, but the full explanation has not yet been elucidated. Kinases are known to regulate the majority of cellular pathways, especially those involved in signal transduction. The DMPK protein like other kinases regulates substrate activity and is regulated itself, by phosphorylation. There is evidence that DMPK is involved in ion homeostasis and aspects of actin cytoskeleton remodelling (Jin *et al.*, 2000), specifically myotube differentiation (Beffy *et al.*, 2005), via Rac-1 and Raf-1 signalling (Shimizu *et al.*, 2000). *In vitro*, substrates of DMPK include CUG-BP1, a developmental splicing regulator strongly implicated in DM1 pathogenesis, whose concentration and activity is increased in the

nucleus in DM1 patients (Timchenko *et al.*, 2001). Phosphorylation of CUG-BP1 results in a decreased nuclear concentration of this form in DM1 patients (Roberts *et al.*, 1997), so DMPK may contribute to the regulation of cellular CUG-BP1 distribution. However, CUG-BP1 has not been confirmed as a *DMPK* substrate *in vivo*.

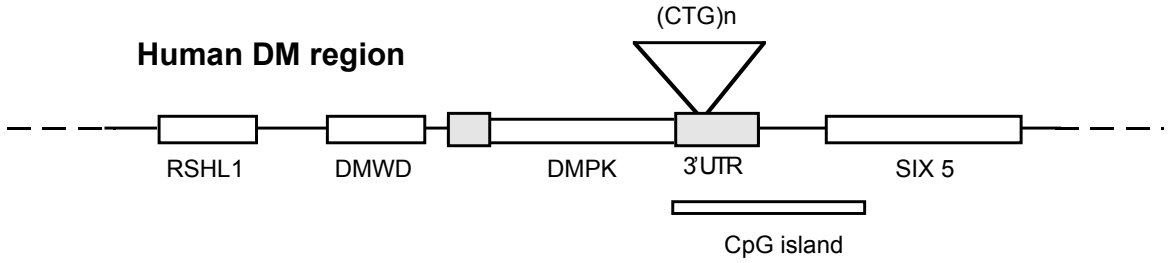


Figure 2 Schematic diagram depicting the gene-dense region around the myotonic dystrophy protein kinase gene (*DMPK*). Abbreviations: RSHL1, Radial Spoke Head Like protein; DMWD, Myotonic Dystrophy WD gene; UTR, Untranslated region; SIX, sine oculis-like.

1.1.4 Somatic variation and anticipation

The somatic variation of expanded repeats was originally uncovered by the diffuse signals observed on Southern blot analysis of repeat region restriction fragments (Buxton *et al.*, 1992; Wong *et al.*, 1995). The use of small pool PCR techniques later revealed these fragments to comprise a heterogeneous mosaic of variable repeat lengths which tend to be expansion biased, age dependent and tissue specific (Monckton *et al.*, 1995; Wong *et al.*, 1995). Instability of DM1 CTG repeats has been reported in a wide range of human tissues, including peripheral blood lymphocytes, liver, pancreas, kidney, brain and heart (Lavedan *et al.*, 1993; Kinoshita *et al.*, 1996; Martorell *et al.*, 1998). Furthermore, instability in muscle is high compared to other tissues such as blood (Thornton *et al.*, 1994), which could account for the major muscle involvement of the myotonic dystrophy phenotype. Germline instability is clearly responsible for much of the phenotypic variation and intergenerational effects (Harper, 2001). In sperm, the repeat array is also unstable and expansion biased, concordant with anticipation, but the overall expansion rate is not as high as in the soma. This may be due to contractions which have been observed in, and are restricted to the male germline (Shelbourne *et al.*, 1992; Monckton *et al.*, 1995; Martorell *et al.*, 2000), or selection against the transmission of large expansions (Jansen *et al.*, 1994). Consequently in congenital cases, the mutant allele is usually inherited from an affected mother (Lavedan *et al.*, 1993).

The fact that mutant *DMPK* alleles are unstable and expand throughout the lifetime of the individual both somatically and within the germline supports the observed genetic anticipation, whereby subsequent generations of an affected family suffer earlier disease onset and increasingly severe symptoms. Alleles with repeat arrays at the longer end of the 'normal' range have been termed 'pre-mutation' and people possessing these lengths are at high risk of having affected family members within a limited number of generations (Martorell *et al.*, 2001).

In vivo data suggests the actual mechanism of expansion is independent of cell division. In mice, as in humans, there is no obvious correlation between cell division rates and the tissue specificity of somatic mosaicism *in vitro* (Gomes-Pereira *et al.*, 2001 and unpublished data) and *in vivo*. Indeed, expansions continue to accumulate in post-mitotic tissues such as brain, arguing against a major role for cell division in repeat expansion (Lia *et al.*, 1998; Fortune *et al.*, 2000; Kennedy *et al.*, 2000). Mouse models indicate repeat instability is dependent on the mismatch repair process rather than replication. In a DM1 context, Pms2 (a MutL homologue) null mice, demonstrated increased somatic instability (Gomes-Pereira *et al.*, 2004). Also in DM1 mice, expansion was found to be dependent on Msh3 (a mutS homologue) (Foiry *et al.*, 2006). In Huntington transgenic mice, instability was found to be dependent on Msh2 (a MutS homologue), both somatically and within the germline (Manley *et al.*, 1999; Kovtun *et al.*, 2001). MSH2 was also required for germline contractions in sperm in DM1 transgenic mice expressing >300 repeats (Savouret *et al.*, 2004). Research suggests then that major components of mismatch repair including MSH2; MSH3 and PMS2 are required to bring about repeat expansion.

The two principal aspects of DM1 –instability of the repeat expansion and pathogenesis of the disease, are currently under intense academic scrutiny. Since the severity and the age of onset in myotonic dystrophy relate to the length of the expansion, the rate of somatic instability has strong implications for the pathogenesis and progression of the disease. Longer repeats elicit more severe symptoms, which appear earlier probably by advancing the pathogenic response (Gomes-Pereira *et al.*, 2006 for review). This may produce different phenotypic effects depending on the developmental stage of the affected individual, accounting for the congenital form. Research into mechanisms in these areas could lead to treatment of the condition by either impeding expansion from an early age; actual reduction in repeat length, or modification of the downstream effects caused by expression of the expanded arrays.

1.1.5 Pathogenesis

The genetic mutation was identified in the mid 1990s (Aslanidis *et al.*, 1992; Brook *et al.*, 1992; Buxton *et al.*, 1992; Harley *et al.*, 1992), and yet our knowledge of how the CTG expansion leads to the DM1 phenotype is still incomplete. Current research is focussed in three main areas: Haploinsufficiency of DMPK; Chromatin disruption and RNA gain of function.

1.1.5.1 DMPK haploinsufficiency

Once the myotonic dystrophy mutation was uncovered, it was expected that expansions of CUG repeats would affect transcription or translation of DMPK leading to haploinsufficiency. Messenger RNA levels were assessed by researchers and in some cases found to be reduced in adult (Fu *et al.*, 1993; Hofmann-Radvanyi *et al.*, 1993; Eriksson *et al.*, 1999) as well as unaffected in adult (Inukai *et al.*, 2000), and raised (Sabouri *et al.*, 1993) and reduced (Hofmann-Radvanyi *et al.*, 1993) in congenital tissues. Conflicting reports as to whether mRNA transcripts were increased or decreased may have stemmed from the different methods and position of PCR primers used in each study, and the different tissue types analysed (Hofmann-Radvanyi *et al.*, 1993). Krahe *et al.* showed that the levels of unprocessed pre-mRNA was equivalent between wild type (wt) and disease alleles, and that the overall amount of processed polyA+ *DMPK* transcript (wt and mutant) was equivalent between unaffected and affected individuals, but that processed (polyA+ and spliced) levels of the disease allele were reduced compared to wild type (Krahe *et al.*, 1995). Around the same time, analysis of the intracellular localisation of transcripts had shown mutant *DMPK* RNA to become aggregated and retained within the nucleus in discrete foci in patient tissues (Taneja *et al.*, 1995). Later, this was also shown in myoblast culture expressing CUG repeat tracts (Davis *et al.*, 1997), and that differentiation of the myoblasts was inhibited by expression of the CUG expansions (Amack *et al.*, 1999). This lead researchers to postulate that aberrant RNA processing of the *DMPK* gene leading to haploinsufficiency of myotonic dystrophy protein kinase (DMPK), could be the cause of disease development. This hinted that the method of RNA extraction used by researchers – whether nuclear RNA in addition to cytoplasmic RNA had been recovered, was vital in the interpretation of previously published results. It has already been mentioned that six major cell-type specific alternatively-spliced isoforms of *DMPK* have been detected to date (Groenen *et al.*, 2000). In addition, Tiscornia *et al.* reported a novel isoform of *DMPK* lacking repeats not retained in the nucleus in DM1 cells. This would result in imbalances in relative levels of cytoplasmic *DMPK* mRNA isoforms (Tiscornia *et al.*, 2000)

corroborating the theory that the position of the PCR primers –whether they were within alternatively spliced exons, and the type of tissue analysed could have been important.

Using antiserum directed against DMPK, levels of the protein were reduced in patient tissue and cells, adding credence to the haploinsufficiency hypothesis (Fu *et al.*, 1993; Koga *et al.*, 1994). However a more recent publication (Narang *et al.*, 2000) reports that the antisera used in these experiments was not specific for DMPK –cross-reacting with other proteins, and that using a DMPK-specific antibody, levels are moderately raised in adult DM1 tissues but in congenital cases a slight decline is apparent. In contrast, Salvatori *et al.* showed a decrease of DMPK to 50% normal levels in 16 adult DM1 skeletal muscle samples. The lowest concentration of DMPK protein correlated to those samples containing the least number of type 1 skeletal muscle fibres, but did not correlate with repeat length (Salvatori *et al.*, 2005).

Aside from the conflicting reports of *DMPK* mRNA and DMPK protein levels to date, there are other anomalies to the haploinsufficiency hypothesis. Mouse models homozygous for the loss of *Dmpk* have been extensively analysed phenotypically and pathologically and so far develop only a mild myopathy (Jansen *et al.*, 1996) and cardiac conduction defects (Berul *et al.*, 1999). In isolated cardiac myocytes, abnormal contractile activity and calcium cycling (Pall *et al.*, 2003), and abnormal sodium channel gating (Lee *et al.*, 2003) was apparent. Homozygous loss of *DMPK* in mice has therefore failed to reproduce many of the multisystemic effects observed in DM1, such as myotonia; progressive muscle wasting; diabetes and cataracts. In addition, individuals homozygous for the DM1 mutation have been documented, and diploid production of mutant *DMPK* mRNA did not adversely affect the clinical presentation in these patients (Martorell *et al.*, 1996). It has to be noted however that in this particular study, the length of repeats, are not the same with 1000/60; 61/38 and 51/120 respectively which makes assessing “homozygotes” difficult. If the effect of the repeats were additive, then the gain would not be expected to increase disease severity into the next category. Nevertheless, these observations, and the fact that no other DM1 cases have been identified arising from a point mutation or a deletion within the coding region, support the view that the complete clinical pathology of DM1 does not result from simple loss of function alone.

1.1.5.2 Disruption of chromatin structure

The *DMPK* gene is located within a gene dense region of 19.3q (Figure 2). Expression of the mutant allele has been associated with the loss of a nearby DNase-I hypersensitive site

(Otten *et al.*, 1995), and therefore could affect transcription of surrounding genes. Using electron microscopy and competitive nucleosome reconstitution, blocks of repeats n=75 and n=100 have been shown to form unusually stable nucleosomes *in vitro* which could reasonably be expected to profoundly alter local chromatin structure (Wang *et al.*, 1995). More recently Flippova *et al.* showed that CTG repeats at the DM1 locus are a component of a CTCF-dependent insulator element, and that repeat expansion results in conversion of the region to heterochromatin (Filippova *et al.*, 2001). Cho *et al.* confirmed this and also showed that an antisense transcript emanating from the adjacent SIX5 regulatory region extending into the insulator element is produced, and converted into 21 nucleotide (nt) fragments. CTCF restricted the extent of the antisense RNA at the wt DM1 locus and constrained the methylation to the nucleosome associated with the CTG repeat, whereas the expanded allele in congenital DM1 was associated with loss of CTCF binding, spread of heterochromatin, and regional CpG methylation (Cho *et al.*, 2005). The presence of 21 nucleotide repeats and associated histone methylation would suggest a role in gene silencing, which would be expected to have a profound effect on *DMPK* expression in congenital DM when CTCF insulation failed.

Of the genes surrounding the site of the DM1 mutation, *SIX5* (formerly known as DM locus-associated homeodomain protein, *DMAHP*) has been the most extensively studied. SIX5 contains a Six domain and a homeodomain (Boucher *et al.*, 1995) as first characterised in *sine oculis*, a homeobox gene essential for the development of the eye in *Drosophila melanogaster* (Cheyette *et al.*, 1994). The 5' UTR and promoter region of *SIX5* is associated with a CpG island at the 3' end of *DMPK*, extends over 3.5 Kb and is interrupted by the DM1 CTG repeat (Figure 2). RT-PCR analysis shows that *SIX5* is expressed in a number of human tissues, including skeletal muscle, heart and brain (Boucher *et al.*, 1995), and in DM1, levels have been shown to be reduced (Thornton *et al.*, 1997). As a result, *SIX5* was selected as a candidate gene for DM1 pathogenesis due to haploinsufficiency. However, mice homozygous for loss of *SIX5* expression had no apparent abnormalities of skeletal muscle function, but did develop lenticular opacities at a higher rate than controls (Klesert *et al.*, 2000). Sarkar *et al.* demonstrated that heterozygous loss of *SIX5* in mice was sufficient to produce cataracts (Sarkar *et al.*, 2000). SIX5 deficiency could contribute to the cataract phenotype in myotonic dystrophy providing evidence of multigenic involvement. The cataracts lacked the characteristic posterior positioning and red-green iridescence seen in patients (Ranum *et al.*, 2004) however, but this may be due to the differences between the murine and the human eye.

Of the other genes surrounding the DM1 locus (Figure 2), *DMWD* is normally expressed highly in brain and testes (Jansen *et al.*, 1992)–tissues affected in myotonic dystrophy. Reduced expression is evident in the cytoplasm of DM1 skeletal muscle (Eriksson *et al.*, 1999; Westerlaken *et al.*, 2003). *Radial Spokehead-Like* gene, (*RSHL1*) has been implicated in sperm motility (Eriksson *et al.*, 2001). Both fertility and brain function are affected in DM1, but little is known about the function of these genes, so how their disruption would affect the phenotype is not yet clear.

1.1.5.3 Gain of function

The evidence in support of a gain of function of the mutant allele arises from its dominance and from a broad correlation between the length of the repeat and the severity of the disease. The first experimental evidence for a dominant effect of mutant *DMPK* transcripts on RNA metabolism was postulated in 1995. In DM1 patients, Wang *et al.* found relatively small decreases of DM kinase RNA in the total RNA pool from muscle, but dramatic disease-specific decreases of both the mutant and wt DM kinase RNAs in the poly(A)+ fraction (Wang *et al.*, 1995). They postulated a *trans*-dominant effect of the expanded repeat RNA upon both WT and expanded transcripts. This data didn't hold up however since the levels of allele-specific RNA were not measured directly, but extrapolated by combining two sets of data –the relative levels of each allele based on a restriction site difference, and the relative levels of *DMPK* polyA RNA (WT and mutant) compared to *creatine kinase*. As mentioned earlier, using direct allele-specific measurement, Krahe *et al.* showed that the levels of polyA+ RNA from normal and mutant alleles were equivalent, and that the DM allele pre-mRNA was incorrectly processed at the splicing level. The unstable repeat impaired post-transcriptional pre-mRNA processing (Krahe *et al.*, 1995). The misleading results from Wang *et al.* could be explained then, if the quantitative measurements were based on PCR primer position spanning an intron. Around the same time, Taneja *et al.* discovered foci of CUG transcripts within the nucleus of patient cells (Taneja *et al.*, 1995). Subsequently, attention has focussed on binding interactions, both RNA secondary structure and RNA-protein interactions, and their possible effects *in vivo*.

Timchenko *et al.* isolated two novel forms of a hetero-nuclear binding protein (hnRNP) $(CUG)_n$ triplet-repeat binding protein CUGBP1 and CUGBP2 and proposed that repeat expansion leads to sequestration of these hnRNPs on mutant transcripts (Timchenko *et al.*, 1996). Napierala and Krzyosiak demonstrated that expanded CUG transcripts formed stable duplex hairpin structures *in vitro* (Napierala *et al.*, 1997). The RNA dominant mutation model for DM1 pathogenesis at this time predicted that the expansion mutation

acts at the RNA level, by forming long dsRNAs that sequester certain RNA-binding proteins. CUGBP1 was shown to regulate alternative splicing of the human *cardiac troponin T* (*cTnT*) muscle fibre protein. In the embryo, exon5 of c*TnT* is included in the transcript whereas in the adult it is not. Splicing of *cTnT* was disrupted in DM1 striated muscle and in normal cells expressing transcripts that contained CUG repeats (Philips *et al.*, 1998). CUG-BP1 levels and activity are increased in DM1 cells (Timchenko *et al.*, 2001). Exon-skipping or alternative splicing of the insulin receptor has also been shown to occur when CUGBP1 is over-expressed in normal cells, or when an expanded CUG repeat is expressed in normal cells (Savkur *et al.*, 2001). As mentioned earlier, the symptoms of myotonic dystrophy include cardiac myopathy and insulin resistance. So, altered splicing of genes regulated post-transcriptionally by CUG-BP1 may play a pivotal role in DM1 pathogenesis.

Michalowski *et al*. tested the hairpin model. They demonstrated similar RNA secondary structure and found *in vitro*, that purified CUGBP1 bound to the hairpin base but not to the stem. In DM1 cells they found no change in the intracellular localisation of CUGBP1, and no association with nuclear foci concluding that sequestration of CUGBP1 was unlikely (Michalowski *et al.*, 1999). This is not necessarily what happens *in vivo* however. In the cell, processes are dynamic, occurring within a complex assortment of proteins, substrates, and catalysts. As RNA in transcribed for instance, proteins may bind as the RNA is generated, before it has formed a secondary double-stranded structure. In this form CUG-BP1 could bind. Preliminary research indicates this to be the case in that cytoplasmic RNA foci are associated with CUG-BP1 (Timchenko, unpublished).

Miller *et al.* proposed that DM1 is caused by aberrant recruitment of triplet repeat expansion (EXP) proteins to the mutant DMPK repeat array, since in DM1 cells they accumulated in nuclear foci (Miller *et al.*, 2000). EXP, activated during mammalian myoblast differentiation (Miller *et al.*, 2000), is homologous to the *Drosophila* muscleblind proteins required for terminal differentiation of muscle and photoreceptor cells (Begemann *et al.*, 1997). EXP proteins have since been renamed muscleblind-like (MBNL) of which there are three proteins in the human and mouse, MBNL1; 2 and 3 all of which co-localise with nuclear foci in DM1 and DM2 myoblasts (Fardaei *et al.*, 2002). In humans, the expression pattern varies between tissues, with MBNL3 expression mostly confined to the placenta. MBNL2 expression is equally and abundantly expressed in tissues examined (pancreas; kidney; skeletal muscle; liver; lung; placenta; brain and heart) and MBNL1 is highest in skeletal muscle (Fardaei *et al.*, 2002). In the adult mouse the expression pattern

is similar for MBNL2, low levels of MBNL3, and MBNL1 were expressed most highly in heart (Kanadia *et al.*, 2003).

With the discovery of aberrant splicing of the *CLC-1* voltage-dependent chloride channel in DM1 patient skeletal muscle (Cooper *et al.*, 2001), evidence was growing in favour of a splicing disorder. This channel is responsible for maintaining resting membrane potential (Bretag, 1987) and there is a direct relation of CLC-1 dysfunction to myotonia (Steinmeyer *et al.*, 1991; Gronemeier *et al.*, 1994), and myotonic disorders such as autosomal dominant myotonia congenita (George *et al.*, 1993) and the "myotonic mouse" (Gurnett *et al.*, 1995). It has been reported that CUGBP1 belongs to a family of RNA binding proteins involved in the developmental regulation of alternative splicing (Ladd *et al.*, 2001), but CUG-BP1 did not co-localise with nuclear foci and as such was not a likely candidate for sequestration on the CUG repeat array. Attention therefore focused on the MBNL double-stranded CUG binding proteins. The muscle and eye pathology and the RNA splicing defects of *Clcn1* were recreated in a mouse model by disruption of the *Mbnl1* gene. Myotonia and cataracts were amongst the symptoms, and *Tnnt2* (*cTnT*) and *TnnT3* splicing was also shown to be defective, as found in DM1 skeletal muscle (Kanadia *et al.*, 2003). Further research has shown MBNL1 to developmentally regulate alternative splicing in genes some of which related to DM symptoms; such as the *insulin receptor* and *cardiac troponin T* (Ho *et al.*, 2004; Dansithong *et al.*, 2005), in a dynamic balance between itself and CUG-BP1 (Ladd *et al.*, 2005). MBNL1 also regulates *protein Tau*, although the connection between disruption and DM1 pathology has not yet been resolved (Leroy *et al.*, 2006).

Yet, sequestration of MBNL1 to foci may not be the cause. Ho *et al.* showed that the formation of foci and co-localisation of MBNL1 are separable events. They expressed expanded CAG or CUG repeats in DM1 myoblasts and found that both expansions aggregated into nuclear foci, and both expansions co-localised with MBNL1, but mis-regulated splicing was seen only by expressing CUG repeats (Ho *et al.*, 2005). Foci formation may not be the basis, but merely a marker of pathogenesis: It may even be protective. More recently a reversible mouse model of DM1 has been developed, employing controllable expression of the DMPK 3' UTR with and without expanded CTG repeats. Here the researchers were surprised to find that mice over-expressing a normal length repeat of 5 CUGs reproduced cardinal features of myotonic dystrophy, including myotonia; cardiac conduction abnormalities; histopathology and RNA splicing defects, but in the absence of detectable nuclear inclusions. Increased levels of CUG-BP1 were

detected in skeletal muscle, as seen in individuals with DM1 (Mahadevan *et al.*, 2006). An expanded CUG repeat is known to produce a long single hairpin *in vitro* (Michalowski *et al.*, 1999) and long CUG hairpins have been shown to induce dicer cleavage and RNAi knockdown of DMPK (Krol *et al.*, 2007). Krol *et al* demonstrated that CNG hairpins are also formed in the triplet expansion disorders Huntington disease, and spinocerebellar ataxia type 1 and constitute another class of dicer targets, uncovering a previously unknown mechanism involving dicer-controlled down-regulation of mutant transcripts in the triplet repeat disease process. It is possible that dicer cleavage of CUG hairpins could release increased amounts of short CUG repeats into the cytoplasm increasing CUG-BP1 levels as a result of titration by binding to the short repeat RNAs. In this case foci could be protective up to a point, preventing toxic short repeats from entering the cytoplasm via dicer cleavage, leading to elevated CUGBP1. MBNL1-associated foci were also found to lack pathogenic effect in *Drosophila* (Houseley *et al.*, 2005). These accounts are consistent with the report of foci formation being separable from missplicing (Ho *et al.*, 2005), since cleaved CAG repeats may not alter CUG-BP1 levels. MBNL1 disruption however, leads to muscle; eye and RNA splicing abnormalities characteristic of DM disease (Kanadia *et al.*, 2003), which is inconsistent with a protective role of MBNL1 sequestration. In this model however, symptoms only occur in the homozygote –a total functional knockdown is apparent, which does not truly replicate the disease state since *in vivo* there is a certain amount of association and disassociation of MBNL1 within foci (Ho *et al.*, 2005).

Could dicer CUG hairpin cleavage play a role in congenital DM? It has recently been reported that dicer is required for embryonic skeletal muscle development (O'Rourke J *et al.*, 2007). In the congenital form the repeat tracts are lengthy and could reasonably be expected to deplete dicer reserves within the cell, affecting other miRNA-controlled functions such as myogenesis.

1.1.6 Myotonic dystrophy type 2

As mentioned earlier, not all cases of DM arise from *DMPK* associated CTG repeat expansions. The DM type 1 mutation accounts for 98% of myotonic dystrophy cases (Harper, 1989). Recently, a second mutation – myotonic dystrophy type 2 (DM2), was discovered, accounting for the majority of the remaining 2% of cases, caused by an uninterrupted CCTG expansion within the first intron of the *cellular nucleic acid binding protein* (*CNBP*) gene (Liquori *et al.*, 2001), also known as *ZNF9*. Type 2 symptoms are almost identical to those of type 1. The notable distinctions are the lack of the congenital and childhood onset forms of the disease, and the lack of central nervous system

developmental abnormalities and retardation (Day *et al.*, 2003). In general, symptoms are less pronounced and occur later in life. It is not clear whether anticipation is present, but the length of repeat is much greater, ranging between 75 and 11,000 CCTG repeats (Liquori *et al.*, 2001).

Although expansions in *CNBP* cause DM-like symptoms, the affected gene product –a sterol regulated single stranded nucleic acid binding protein (Rajavashisth *et al.*, 1989) is apparently unrelated to DMPK. The function is currently under intensive scrutinisation. Six of the seven fingers can be substituted into HIV1 nucleocapsid and support viral genomic RNA replication (McGrath *et al.*, 2003). CNBP can bind to IRES (internal ribosome entry site) sequences and may stimulate cap independent translation as part of a ribonucleoprotein complex (Gerbasi *et al.*, 2007). It is active throughout development, mediating neural crest expansion (Weiner *et al.*, 2007) and regulating forebrain formation in the mouse (Shimizu *et al.*, 2003). Mouse models homozygous for CNBP gene disruption are embryonic lethal, confirming an essential role in development. Interestingly, when heterozygous for CNBP gene disruption, mice demonstrate cardinal features of DM such as abnormal muscle histology; myotonic discharges and heart conduction defects, and a reduction in the levels of *Clcn1* expression (Chen *et al.*, 2007), suggesting that CBNP haploinsufficiency plays an important role in DM2 pathogenesis. However in DM2 patients, CNBP levels are unaffected (Botta *et al.*, 2006; Margolis *et al.*, 2006). As a result, these seemingly conflicting reports suggest that even if haploinsufficiency of CNBP does cause a DM-like phenotype, it still may not actually make any contribution in DM2 patients.

The genes surrounding the DM2 mutation (*KIAA1160, Rab 11B, glycoproteinIX, FLJ11631,* and *FLJ12057*) do not bear any obvious similarities to those surrounding *DMPK* (Figure 2). It is therefore unlikely that any disruption of chromatin structure by the repeats, leading to altered regulation of neighbouring genes, would result in diseases with such remarkably similar features. Molecularly, parallels arise on inspection of the expanded RNAs. Both DM1 and DM2 transcripts are retained within the nucleus and co-localise with MBNL1 (Mankodi *et al.*, 2001; Fardaei *et al.*, 2002). Similar splicing defects for the insulin receptor and the chloride channel are also observed between DM1 and DM2 in muscle biopsies (Savkur *et al.*, 2004; Botta *et al.*, 2007), and also for microtubule associated protein Tau (Maurage *et al.*, 2005). It would not be unreasonable then to hypothesise that sequestration of MBNL1, leading to alterations of the dynamics of CUG-BP1 and MBNL1 regulation, is also pivotal to the mis-regulated splicing in DM2.

1.1.7 Alternative splicing: Dynamic control by CUG-BP1 and MBNL1

The *cardiac troponin T* gene is misspliced in myotonic dystrophy, the embryonic exon 5 inappropriately included in the adult. In 1998, Philips *et al.* reported that in the chicken homologue (ccTNT), CUG-BP1 promoted inclusion by binding to muscle specific enhancers (MSE). Four enhancers were found to be essential for enhanced exon inclusion – one upstream and three downstream of the exon. CUG-BP1 was found to specifically bind to a sequence containing CUGCUG, common to the human gene, within enhancers two and four. Using a human *cTNT* minigene system they determined that the CUG sequences were essential for exon inclusion (Philips *et al.*, 1998). Ho *et al.* investigated the involvement of MBNL1 proteins and determined that they also regulated exon 5 alternative splicing in *cTNT*, and additionally, promoted exon 11 exclusion in the insulin receptor gene –another transcript misspliced in DM. In the *cTNT* gene, MBNL1 was found to bind to introns surrounding exon 5 in both chicken and in human. Two MBNL1 binding sites which affected splicing regulation were defined upstream of the human exon 5, a site distinct from the CUG-BP1 binding site. They also noted that in *ccTNT*, MBNL1 bound weakly to MSE1 (upstream) and strongly to MSE4 (downstream) where the CUG-BP1 binding site is located, indicating antagonistic regulation with CUG-BP1 (Ho *et al.*, 2004). CUG-BP1 is down-regulated in adult heart compared to the embryo, concomitant with exon 5 exclusion, and is also down-regulated in other tissues during development (Ladd *et al.*, 2005). This suggests that competition between CUG-BP1 and MBNL1 for binding sites could be a plausible mechanism for developmentally regulated splicing. Recent work demonstrates that as little as 6 bases are sufficient for MBNL1 binding (Warf *et al.*, 2007), via a structured GC-rich stem-loop containing pyrimidine mis-matches, in both normal splicing substrates and pathogenic RNA. Under the electron microscope MBNL1 forms a ring-like structure, which binds the double-stranded RNA stem. It is possible then that CUG and CCUG expansions may trap MBNL1 protein within stacked ring structures along the double-stranded hairpin (Warf *et al.*, 2007; Yuan *et al.*, 2007).

It seems clear that alteration of the dynamic balance between CUG-BP1 and MBNL1 splicing regulators causes inappropriate generation of foetal isoforms during adulthood. Some misspliced genes can be attributed to a symptom in both myotonic dystrophy type 1 and type 2, although the association is not clear for some genes such as *cardiac troponin T,* or *protein Tau*. If the pathogenesis arises from an RNA gain of function, the differences between the two types –notably the lack of the congenital form in DM2– must relate to

temporal or spatial expression patterns of the *DMPK* and *CNBP* genes themselves, to expression of the surrounding genes due to structural effects of the repeat, or to the specific function of DMPK and CNBP, separately or in combination. While each of the hypotheses addresses part of the pathogenesis, it is most likely that a number of complex mechanisms operate in concert, leading to the substantial clinical variability between patients.

1.2 RNA processing disorders

When the DM1 mutation was first discovered, it was expected that expansion of the CTG repeat tract would result in a simple loss of function of the DMPK gene product, perhaps by interference with transcription or translation. As research has evolved, evidence for a gain of function of the toxic RNA has become pivotal in understanding the pathogenesis in DM, to join a growing group of RNA processing disorders. Insights into the disease mechanism of other members of this group could shed light on the mechanism of DM pathology.

1.2.1 Oculopharyngeal muscular dystrophy

Oculopharyngeal muscular dystrophy (OPMD) is a progressive myopathy, which shares some characteristics of myotonic dystrophy, namely ptosis and dysphagia due to muscle weakness in the face and neck. Difficulty in swallowing can become so severe as to result in death by choking, aspiration pneumonia or malnutrition. Limb-girdle muscles are also affected to a lesser extent (Hill *et al.*, 2001). Usually inherited as an autosomal dominant disease, OPMD is caused by the expansion of a GCG trinucleotide repeat within the coding region of the polyA binding protein 2 (*PABP2*) gene, leading to extension of an alanine stretch within the protein. Functionally, PABP2 binds with high affinity to the polyA tail of messenger RNA via a specific RNA binding domain and is involved in its polyadenylation, controlling adenylate addition to approximately 250 nucleotides (Wahle, 1991; Wahle, 1995). PABP2 also has an oligomerisation domain, and expansion of the alanine stretch promotes atypical oligomerisation of the protein to form intranuclear inclusions in muscle cells: The inclusions, considered to be the pathological hallmark of the disease, comprise oligomerised PABP2 protein; mRNA and other RNA binding proteins (Fan *et al.*, 2003), some of which have been identified. It is not clear as to whether the inclusions themselves are pathogenic, but could possibly become generalised PolyA

RNA traps, and their presence leads to apoptosis (Fan *et al.*, 2001). In mouse models there is a close correlation with the presence of inclusions to pathology and that symptoms tend to be more severe with longer repeat lengths (Hino *et al.*, 2004). Interference with aggregate formation reduces toxicity (Bao *et al.*, 2002; Bao *et al.*, 2004), but interference of oligomerisation domains of expanded PABP2 leads to their solublisation and a subsequent increase in toxicity implying inclusions may be protective (Messaed *et al.*, 2007). Inclusions are located in speckles –domains within the nucleus that are enriched in polyA+ RNA, splicing factors and other mRNA processing machinery that are thought to represent mRNA processing centres, and the nucleolus –sites of ribosome assembly. The mutant protein inhibits myogenesis in cell-culture where the expression of several muscle-specific proteins, alpha actinin; muscle creatine kinase; myogenin and myoD is reduced (Wang *et al.*, 2006). It is not clear why pathology is limited to muscle cells –since presumably PABP2 is ubiquitously expressed. Where this disease becomes particularly interesting in the context of DM1, is that CUG-BP1, a splicing regulator implicated in DM1 pathology, plays a role in deadenylation (Paillard *et al.*, 2003; Moraes *et al.*, 2006). Since in DM1 the concentration and activity of CUG-BP1 increases, it follows that deadenylation may increase. If in OPMD adenylation is disrupted, then both mutant mechanisms could result in short polyA tails, which would affect mRNA stability (Lewis *et al.*, 1997; Cougot *et al.*, 2004; Kuhn *et al.*, 2004). Indeed, some characteristic symptoms –progressive myopathy, ptosis and dysphagia, are shared by both disorders.

1.2.2 Spinal muscular atrophy

Spinal muscular atrophy is a spliceopathy rather than a microsatellite repeat disorder. It is the second largest autosomal recessive cause of infant mortality after cystic fibrosis affecting 1:6,000-10,000 live births (Wirth, 2000). It comprises a group of neuromuscular disorders characterised by the selective destruction of α-motor neurons within the spinal cord leading to progressive muscle atrophy in the limbs and trunk. Life expectancy is reduced, and between individuals the disease severity is highly variable ranging from difficulty in standing, to becoming wheel chair bound. Severely affected infants often die from respiratory insufficiency by the age of two.

On chromosome 5q, in normal DNA, there is an inverted duplication of a 500 kb element spanning at least four genes. The disease is caused by deletions of, or mutations in, the telomeric copy of one of these genes –the survival of motor neurone 1 gene (*SMN1*) (Lefebvre *et al.*, 1995). The severity of the disorder is modified by expression of one or more centromeric duplications of the gene (termed *SMN2*), which can partially compensate

for the primary mutation (Gavrilov *et al.*, 1998; Vitali *et al.*, 1999; Harada *et al.*, 2002). Since *SMN1* is ubiquitously expressed, it is thought that *SMN2* expression can satisfactorily rescue the mutation in all tissues except motor neurones, although the reason for this is not clear. The centromeric copies are almost identical to the *SMN1* gene except for a translationally silent C-T transition, which results in predominant exon 7 skipping (Lorson *et al.*, 1999; Monani *et al.*, 1999) creating defective oligomerisation within the resultant protein (Lorson *et al.*, 1998).

In eukaryotes, nuclear pre-mRNA splicing is essential for mRNA biogenesis, carried out by the spliceosome. In terms of function, the SMN1 protein forms part of a large macromolecular complex involved in the assembly of small nuclear ribonucleoproteins (snRNPs) –major components of the spliceosome. Although snRNPs have been shown to self associate from their component parts in an ATP independent manner (Sumpter *et al.*, 1992; Raker *et al.*, 1996; Raker *et al.*, 1999), recent research suggests that SMN1 complex is absolutely required for the efficiency and accuracy of snRNP assembly (Pellizzoni *et al.*, 2002), akin to the role of chaperones in protein folding. The molecular consequences of reduced levels of full-length SMN1 in motor neurones are not known, but may lead to the formation of aberrant snRNPs, which could contribute to the specificity and function of the spliceosome affecting downstream mRNA processing. There has been no association between the location of foci and the spliceosome in DM1 (Houseley *et al.*, 2005), and no other evidence of spliceosomal involvement , therefore it is unlikely to contribute to DM pathogenesis.

1.2.3 Fascioscapulohumeral muscular dystrophy

Fascioscapulohumeral muscular dystrophy is the third most prevalent muscular dystrophy after the dystrophinopathies (including Duchenne and Becker muscular dystrophy) and myotonic dystrophy, affecting 1 in 20,000 of the population. There is a progressive loss of all skeletal muscle, with noticeable weakness usually starting with facial, scapular/back and upper arm muscles (Tyler *et al.*, 1950). High frequency hearing loss and retinal abnormalities are also common associations (Gurwin *et al.*, 1985). With patients ranging from asymptomatic to wheelchair bound, the clinical spectrum is highly variable, as is the age of onset, but most carriers of the mutation are symptomatic by their second decade. Genetically the mutation is autosomal dominant, and is associated with the deletion of a number of 3.2 Kb repeat units termed *D4Z4* on chromosome 4q35 (van Deutekom *et al.*, 1993). The normal population carries 11-100 *D4Z4* repeat units, but in affected individuals repeats are reduced to 1-10, with a rough inverse correlation between the age of onset and

severity of the disease: One to three repeats causes more severe symptoms (Lunt *et al.*, 1995; Tawil *et al.*, 1996). Although the mutation was mapped to the region in 1993, the pathogenesis is not clear. The *D4Z4* repeat units contain an open reading frame, *DUX4*, but expression of this gene has not been established. The mutation however is within a gene-dense region and so the expression of flanking genes has been extensively studied. A repressor complex controlling upstream *FRG1* transcription levels, binds to *D4Z4* repeats (Gabellini *et al.*, 2002), and so a deletion of these repeats could result in de-repression of the target gene, leading to increased expression levels. Recently Gabellini *et al.* generated transgenic mice over-expressing the human *DUX4* flanking genes *FRG2; FRG1* and *ANT1* in skeletal muscle. Those mice expressing *FRG2* and *ANT1* seemed normal, but those expressing *FRG1* developed a muscular dystrophy with features characteristic of human FSHD (Gabellini *et al.*, 2006). Fascinatingly, over-expression of *FRG1* resulted in the missplicing of some genes also targeted in myotonic dystrophy, namely *MTMR1* and *TNNT3,* but not *CLC1*. These results were corroborated in FSHD patient cell-cultures, and in mouse C2C12 muscle cells over expressing *FRG1.* The FRG1 protein is located in the nucleolus, cajal bodies and speckles (van Koningsbruggen *et al.*, 2004), and so is predicted to be involved in RNA processing, perhaps as a component of the spliceosome. Recent research carried out to discover which proteins associated with FRG1, identified RNA biogenesis proteins SMN1 and PABP2 as binding partners, mutations in which are also associated with neuromuscular disorders (spinal muscular atrophy, see 1.2.2 and ocularpharyngeal muscular dystrophy, see 1.2.1) (van Koningsbruggen *et al.*, 2007). Interestingly there is a link between adenylation factor PABP2 and FRG1, which initiated a literature search for any association of CUG-BP1 with FRG-1, but none was found.

1.2.4 Fragile X syndrome and fragile X-associated tremor/ataxia

Fragile X syndrome is the most prevalent heritable single gene form of mental retardation. Symptoms are broad, ranging from severe mental retardation and autism to those with a normal IQ, although these patients usually have some form of learning disability. Females tend to suffer less severe symptoms since the disorder is X-linked recessive. The mutation is caused by the expansion of a CGG triplet repeat within the 5'UTR of the fragile X mental retardation 1 gene (*FMR1*) and in the unaffected population the repeat tract varies between 5 and 44. Arrays greater than 200 are considered full mutation and give rise to fragile X syndrome, but the presence of 50-200 repeats –the premutation range gives rise to the neurodegenerative disorder fragile X-associated tremor/ataxia syndrome (FXTAS) (Hagerman and Hagerman, chapter 10 in Wells *et al.*, 2006). This disease differs from

fragile X syndrome in that patients tend to be affected later in life rather than from childhood, with gait ataxia and intention tremor, and more variably parkinsonism and numbness in the lower extremities. Although the same mutation is causative in both fragile X syndrome and FXTAS, it is thought that pathogenesis follows two different mechanisms. Fragile X syndrome is clearly borne of protein deficiency due to silencing of the gene by methylation (Lim *et al.*, 2005; Ladd *et al.*, 2007), whereas in FXTAS, the gene is transcriptionally active with elevated levels and altered sites of initiation, suggesting a gain of function mechanism. Post mortem, ubiquitinated *FMR1* mRNA-positive nuclear inclusions in cells of the central nervous system are the cardinal feature of FXTAS (Tassone *et al.*, 2004). Other proteins identified from purified inclusions include the RNA binding proteins MBNL1 (instrumental in DM missplicing) and hnRNP A2 as well as heat shock proteins HSP27 and αB-crystallin, and neurofilament proteins –one of which is mutated in Charcot-Marie-Tooth disease of the axonal type (peripheral neuropathy) (De Sandre-Giovannoli *et al.*, 2002).

Recent research indicates that FMR protein effects microRNA-guided translational repression as part of the miRNP assembly by exchange of the complex-associated micro RNA for the target mRNA (Plante *et al.*, 2006; Plante *et al.*, 2006). Micro RNAs (miRNAs) are thought to translationally regulate ~30% of the genes in the human genome. They are borne from extensive processing of non-coding RNA, transcribed from endogenous miRNA genes; first to stem-loop double-stranded primer RNA by the microprocessor complex containing drosher, then to single-stranded miRNA by a complex containing dicer and transactivating response RNA-binding protein (TRBP). This complex associates with AGO2 to form the microribonucleoprotein (miRNP) or RNA induced silencing complex (RISC), which recognises the mRNA target, leading to cleavage if the miRNA is a perfect match, or transcriptional repression if not (Ouellet *et al.*, 2006). It is thought that FMR1 is involved in post-transcriptional gene control. Micro CUG RNAs (21nt) have been detected in the total RNA of DM1 cells, so it is possible that *DMPK* is regulated by dicer (Krol *et al.*, 2007). There is then a possibility that the expanded repeat also disrupts the post-transcriptional control of other genes in DM1 by disruption of the same process.

1.2.5 Huntington disease-like 2

Huntington disease-like 2 (HDL2) is an autosomal dominant progressive disorder almost indistinguishable from Huntington disease (HD). The diseases are characterized by abnormalities of movement; dementia; and psychiatric disturbances, and pathologically,

marked striatal atrophy and moderate cortical atrophy, and intranuclear inclusions. The inclusions in HDL2 are not huntingtin or junctophilin-3 (the affected gene in HDL2) positive however. Death occurs approximately 20 years after onset, which is usually the fourth decade, but longer repeat tracts elicit an earlier onset. Whereas HD is caused by the expansion of a CAG repeat leading to an extended polyglutamine repeat within the Huntingtin protein, the HDL2 CTG repeat is situated within the alternatively-spliced *junctophilin-3* gene (*JPH3*), and manifests variably –in frame to code for polyalanine or polyleucine, or non-coding within the 3'UTR, the repeat region is removed from the full-length transcript (Margolis *et al.*, 2001). How the expansion in HDL2 leads to an almost identical phenotype to HD is unclear, but an RNA gain-of function mechanism seems likely. The HDL2 repeat is transcribed in the CTG orientation producing RNA foci that co-localise with MBNL1. Missplicing is apparent in *amyloid precursor protein* (*APP*) and *microtubule associated protein tau* (*MAPT*) (Rudnicki *et al.*, 2007). If MBNL1 depletion is involved in missplicing as in DM, the differences in target genes may relate to the expression pattern of *JPH3* compared to *DMPK*. *Myotonic dystrophy protein kinase* is present at low levels in brain and high in heart, skeletal muscle and testis (Sarkar *et al.*, 2004). *Junctophilin-3* is expressed at high levels in brain, is modest in testis and minimal elsewhere (Takeshima *et al.*, 2000).

1.2.6 Friedreich ataxia

Friedreich ataxia (FRDA) is a progressive neurodegenerative disorder caused by an unstable GAA repeat expansion within intron 1 of the *frataxin* (*FXN*) gene (Campuzano *et al.*, 1996). Symptoms include gait instability and general clumsiness, sensory loss and muscle weakness, and less often cardiomyopathy, scoliosis and diabetes (Geoffroy *et al.*, 1976). Onset varies between 5 and 25 years of age. Frataxin is a mitochondrial protein and current research suggests that loss of frataxin impairs mitochondrial iron handling and respiratory chain function contributing to increased oxidative stress and cell damage leading to spinal atrophy (Campuzano *et al.*, 1997; Lodi *et al.*, 1999; Bradley *et al.*, 2000; Cavadini *et al.*, 2002). The disease is recessive and point mutations are rare (2-4%), but can cause the disease (Puccio *et al.*, 2000). It is thought that an expansion over 60 repeats allows the formation of "sticky" triplex DNA formation, which inhibits transcription of the gene by direct sequestration of RNA polymerase II (Sakamoto *et al.*, 1999).

1.2.7 Spinocerebellar ataxias 8, 10 and 12

The spinocerebellar ataxias (SCAs) encompass a large group of dominantly inherited progressive disorders in which the cerebellum slowly degenerates. Coordination of movement is generally affected involving gait, speech and hand and eye motor control, which deteriorates over time. Mental function is not affected. Spinocerebellar ataxias are grouped into three types; CAG repeat expansions encoding a polyglutamine stretch; non-coding repeat expansions and non-repeat associated (deletion; missense; nonsense and splicing) mutations. SCA 8, 10 and 12 belong to the non-coding expansion group, although SCA 8 is now also associated with group 1 since recent research shows it is bi-directionally transcribed resulting in a polyglutamine expansion in the *ataxin 8* (*ATXN8*) gene, and a complementary mRNA *ataxin 8 opposite strand* (*ATXN8OS*) containing an untranslated CUG expansion (Moseley *et al.*, 2006). Moseley *et al.* generated a successful transgenic SCA8 mouse model by expressing the human *ATXN8OS* gene with 116 CTG repeats, demonstrating the pathogenicity of the CUG repeats. Little is known about the function of the *ATXN8OS* gene, but in light of these results, an RNA mediated gain of function mechanism is still likely.

The SCA 10 mutation consists of an expanded ATTCT repeat within exon 9 of the *ATXN10* gene, expressed in brain. Little is known about the function ATXN10, but in SCA10 patient-derived cell-lines, transcription levels are normal and the RNA is correctly processed. Mouse models null for the gene do not survive embryogenesis in the homozygotes, and heterozygotes show no phenotype (Wakamiya *et al.*, 2006) indicating that a simple loss or gain of function of the protein is doubtful.

The SCA12 CAG repeat is located within exon 7 of the *PPP2R2B* gene (Holmes *et al.*, 1999), which encodes a brain-specific regulatory subunit Bβ of the ubiquitous PP2A phosphatase (Mayer *et al.*, 1991). This gene is alternatively spliced, the different isoforms determining substrate specificity and subcellular localisation. The disease pathogenesis has not yet been established, but no evidence that SCA12 is a polyglutamine disease has been found. Expansions were not detected in western blots of patient derived cell-lines or in SCA12 brains (Holmes, 2003). Research by Holmes *et al.* suggests that the mutation increases expression levels of the predominant Bβ isoforms (Bβ1), which use the exon 7 alternative promoter, perhaps shifting PP2A target specificity (Holmes, 2000).

1.3 Transgenic mouse models of myotonic dystrophy

Many mouse models have been generated to further understand the progression of disease in myotonic dystrophy. Not all models have been created with pathogenesis in mind -a large amount of research has focussed on the repeat expansion and instability. Published DM1-related mouse model data is compared here from a pathogenesis perspective. The transgenes are summarised in Figure 3.

1.3.1 Gene deficiency and over-expression models

These models have already been discussed in relation to the pathogenesis of DM1 in Pathogenesis 1.1.5, but to recap: Of the gene disruption models, transgenic mice deficient in *DMPK* (*Dm15* or *Dmpk* gene in mice) did not show significant myopathy, but have shown cardiac conduction abnormalities (Reddy *et al.*, 1996; Berul *et al.*, 2000). Mice lacking the adjacent *Six5* gene developed cataracts with either heterozygous or homozygous loss, but of a different type, not posterior subcapsular iridescent cataracts, (Klesert *et al.*, 2000; Sarkar *et al.*, 2000). MBNL1 disruption in contrast, lead to muscle, eye and RNA splicing abnormalities characteristic of DM disease (Kanadia *et al.*, 2003). Transgenic mice over-expressing CUG-BP1 in muscle and heart also replicated DM1-like features in these tissues, including abnormal muscle histology and disrupted splicing (Ho *et al.*, 2005). None of these single gene disruption models mimic the complete multisystemic phenotype of the disease including the congenital form in DM1. However, both the MBNL1 deficiency and CUG-BP1 over-expression models are consistent with a mechanism for DM pathogenesis in which expanded repeats result in alteration in the dynamic balance between MBNL1 and CUG-BP1, and have *trans*-dominant effects on specific pre-mRNA targets.

1.3.2 Expressed transgenes

If the phenotype of myotonic dystrophy results from a gain of function of the mutant *DMPK* transcript, one would expect that the mouse models generated to date would reflect this. Not all models were specifically created to answer pathogenesis hypotheses, so analysis of the models in this way is slightly hampered by the fact that the researchers who created them tended to concentrate on either the mechanism of repeat expansion or the mechanism of pathogenesis, and publish accordingly. Of the mouse models, some transgenes included a promoter:

Gourdon *et al.* 1997: The DM1 region including upstream *DMWD* and downstream *SIX5* was used for this transgenic model. Fifty-five CTG repeats are present within the *DMPK* 3'UTR and all three genes are expressed at wild type levels. Although the repeat tract was found to be unstable, no phenotype was observed. The same research group then produced mice using the same region with 20 repeats in the normal line or 300 repeats for the expanded line (Seznec *et al.*, 2000). The levels of CTG expression were again found to be consistent with wild type levels. There was no phenotype observed with 20 repeats, but with 300 repeats, mice displayed clinical, histological, molecular and electrophysiological abnormalities in skeletal muscle consistent with those observed in DM1 patients. Abnormal tau expression was also observed in the brain. In addition, mice had crossed teeth, perhaps indicating a weakness in the jaw musculature. A proportion of offspring expressing sufficient mRNA exhibited myotonia (Seznec *et al.*, 2001) An abnormal glucose and insulin response was evident in the expanded repeat mice and molecularly, tissue and age-dependent abnormal regulation of *IR* mRNA splicing was found in skeletal muscles, adipose tissue, liver and pancreas (Guiraud-Dogan *et al.*, 2007).

Jansen *et al.* 1996, restricted the DM1 region used for their mice to the last exon of *DMWD* and the *DMPK* gene, which included 20 CTG repeats within the 3'UTR. *DMPK* levels were altered between 0 to 10 fold depending on the line, due to multicopy integration. Offspring suffered high pre and neonatal mortality, as high as 50% in *Tg* positive mice through female transmission, where the females themselves often became sick and occasionally died, and 20% mortality when transmitted through the male germline. Mice developed cardiac hypertrophy, but no DM-like pathology or phenotype was detected. O'Cochlain *et al.* analysed mice expressing 26 copies of the human *DMPK* gene with 11 CTG repeats. The mice exhibited hypertrophic cardiomyopathy with dysrhythmia, myotonic myopathy and hypotension, all distinctive muscle traits of DM1 (O'Cochlain *et al.*, 2004).

In their mouse model, Monckton *et al.* 1998, reduced the region still further, to the *DMPK* 3'UTR that contained 162 repeats. Transcription was directed from the *EF1α* promoter. They reported myotonia in three of four lines at 2 years of age. The mice also suffered ataxia, testicular atrophy and infertility so severe as to lose the lines (Monckton, 1998).

Mankodi *et al.* 2000, included 5 and 250 CTG repeats within the 3'UTR of a gene unrelated to *DMPK*, *human skeletal actin* (*HSA*). Gene expression was limited by the *HSA*

promoter to skeletal muscle. In the 250 CTG repeat, but not the 5 CTG repeat mice myotonia was present from four weeks before muscle pathology was histologically defined, and transcripts were retained in the nucleus within discrete foci.

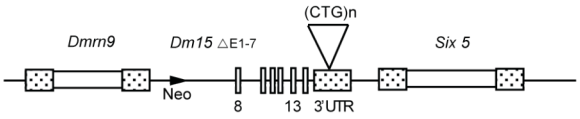
Mankodi *et al.*, 2000, Seznec *et al.* 2000 (Figure 3) and Monckton *et al.* (unpublished, personal communication) have communicated findings of myotonia and other aspects of DM pathology in their mice. Clearly in these examples, mice expressing CTG repeats with expansions of 162 repeats and above exhibit DM-related symptoms, including the hallmark myotonia. None of the other mouse lines presented myotonia, perhaps either because the repeat expansion was insufficient to cause disease within the lifetime of the mouse (Gourdon *et al.*, 1997, 55 repeats), or expression levels were insufficient due to transgene positional effects. The most interesting model here is that of Mankodi *et al.* 2000, since myotonia was reproduced in mice simply by expressing a modified *human skeletal actin* gene containing 250 CTG repeats within the 3' UTR. The repeat sequence comprised a dimerised $(CTG)_{130}$ fragment devoid of flanking sequences from the DM1 region (Mankodi *et al.*, 2000). Over-expression of MBNL1 in the same transgenic line rescued the pathogenic effect: Skeletal muscle from the HSA^{LR} transgenic mouse line was subjected to *in vivo* mediated transduction with a recombinant adeno-associated viral vector over-expressing MBNL1, which rescued the myotonia. This reversal occurred concurrently with restoration of the normal adult-splicing patterns of four pre-mRNAs misspliced in DM muscle (*Clcn1*, *Cypher*, *Serca1* and *Tnnt3*) (Kanadia *et al.*, 2006): Strong evidence for the MBNL1 depletion hypothesis.

Removal of toxic RNA was also shown to alleviate DM-like symptoms in an inducible mouse model of DM1, an important finding indicating that in the future, treatment of DM may become a possibility. Mahadevan *et al*. produced a tet responsive GFP reporter flanked by the *DMPK* 5' UTR, and the *DMPK* 3'UTR containing 5 or 200 CTG repeats. Induction of expression resulted in a major phenotype in the 5 repeat mice a few days after induction, and after nine months without induction due to a leaky promoter. Although foci were seen in the 200 repeat mice, no phenotype was observed probably because expression was too low. Cessation of transcription restored function and splicing anomalies. As with the HSA-(CTG)5/200 model (Mankodi *et al.*, 2000), no DMPK protein was expressed giving clear delineation of the contribution of DMPK 5' and 3'UTR mRNA to DM pathophysiology (Mahadevan *et al.*, 2006).

Transgenic mouse models of myotonic dystrophy

Deficient in DM region genes

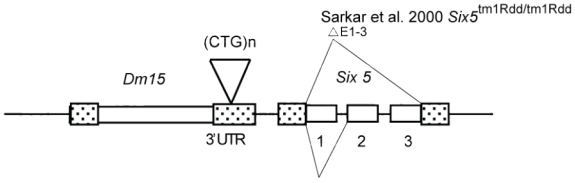
Reddy *et al.* 1996 *Dmpk*$^{\mathbf{tm1Rdd}/tm1Rdd}$



Mouse DM region

Mild skeletal myopathy and wasting.
Ca2+ homeostasis in differentiated cultured myotubes is altered.

Sarkar et al. 2000 *Six5*$^{tm1Rdd/tm1Rdd}$



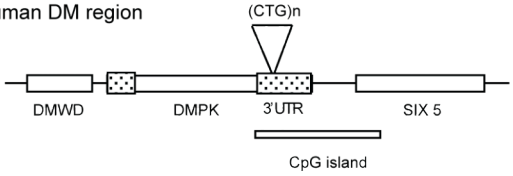
Mouse DM region

Klesert et al. 2000 *Six5*$^{\triangle E1/\triangle E1}$

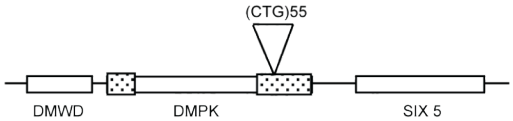
Cataracts.
Sterility and decrease testicular mass.
Muscle histology normal.

Repeat expansions and the human DMPK region

Human DM region

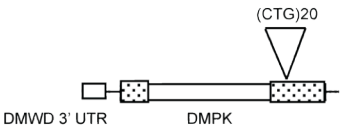


Gourdon et al. 1997 DM 55



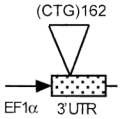
Very small amount of somatic instability.
All three genes expressed at wild type levels.

Jansen et al. 1996 Tg



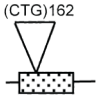
Endogenous DMPK levels altered
0 to 10 fold depending on the line.
High infant mortality distinct from congenital DM.
Cardiac hypertrophy but no DM-like pathology or phenotype.

Monckton et al. 1998 Te 162



Severe infertility and testicular atrophy
Late onset myotonia and ataxia.

Monckton et al. 1997 Dmt 162



Normal phenotype

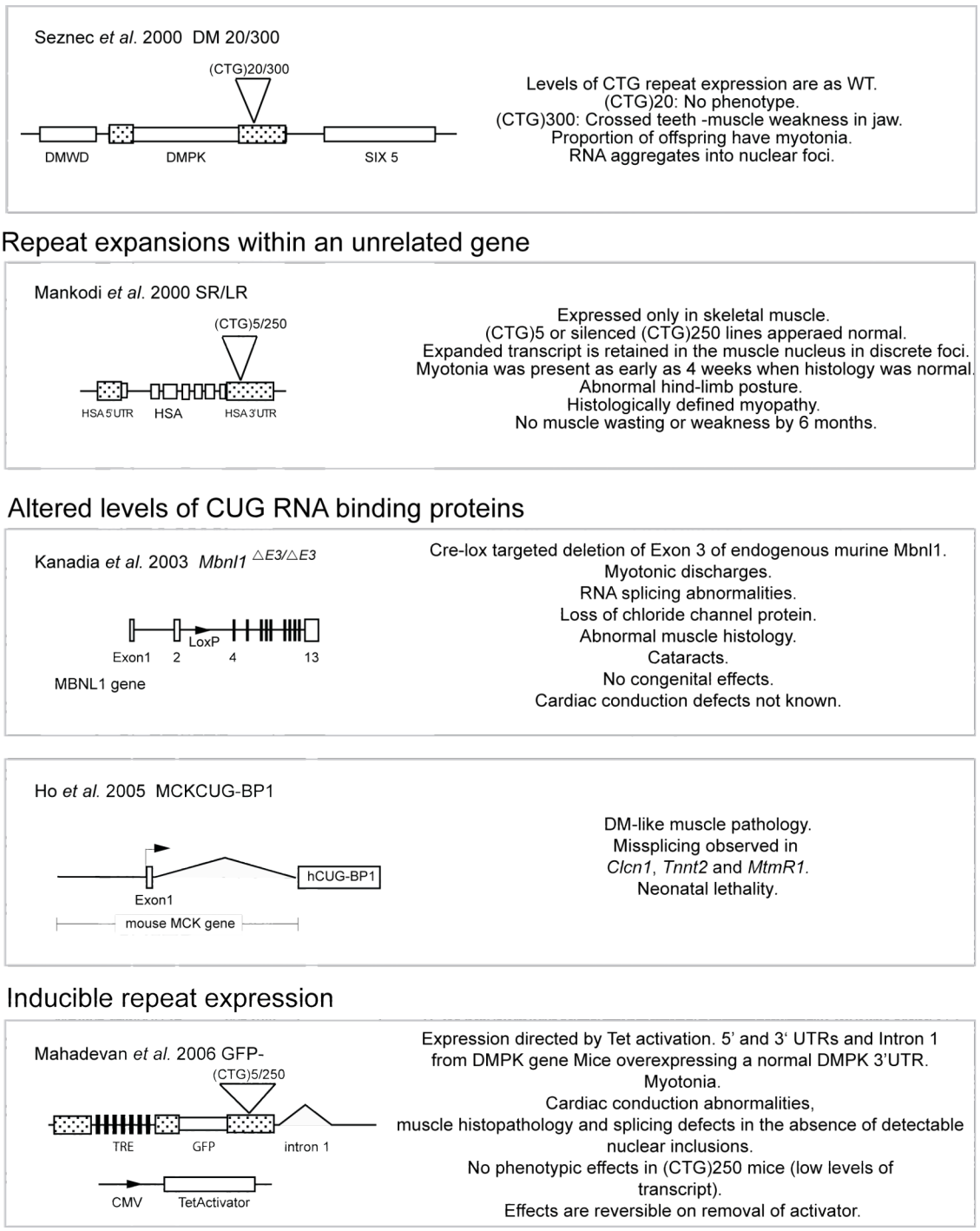


Figure 3 Transgenic mouse models of myotonic dystrophy. Schematic diagram summarising the transgenes used, and a brief description of the phenotype produced. Notably DM like symptoms only arise on the expression of large repeats, the deficiency of MBNL or the over-expression of CUG-BP1. Abbreviations: EF1α , *elongation factor 1*α promoter; HSA, *human skeletal actin*, and see figure 1 legend. N.B. The listed phenotypic effects are not exhaustive, in cases where research groups primarily focus on instability, the phenotype has often not been stated.

1.4 Project outline

Myotonic dystrophy type 1 is a multisystemic disease caused essentially by the dysregulated splicing of MBNL1 and CUG-BP1 pre-mRNA targets, some whose function can be directly related to a clinical feature, such as the chloride channel to myotonia, or the insulin receptor to diabetes. Many symptoms nevertheless have yet resisted explanation: Cataracts, testicular atrophy, heart block, anaesthetic sensitivity and mental retardation for example. The purpose of this thesis is to increase understanding of the gain of function hypothesis –the pathogenesis of an expressed expanded CUG repeat tract, within multiple organ systems using both *in vivo* and *in vitro* approaches. To delineate cause and effect, we attempted to use the Cre-*lox* system for the conditional expression of an expanded CTG repeat in the mouse, to mimic the pathogenesis of myotonic dystrophy type 1 with temporal and spatial control. We successfully generated a 5 repeat transgenic line, but not an expanded repeat line. The constructs generated were validated *in vitro*, and formed the basis of a cell-culture model, which reproduced the pathogenic marker of DM1, MBNL1-associated foci. This culture system was then used to identify further genes possibly misspliced in DM1, by comparing the effects of repeat expression on downstream RNA processing, using differential expression and alternative exon microarray analysis. The pore-forming unit of the cardiac L-type calcium channel (*Cacna1c*) was identified as a candidate gene, mutations in which are known to cause conduction defects identical to those observed in myotonic dystrophy type 1.

2 Materials and methods

2.1 Materials

All chemicals, molecular biology reagents, disposable tissue culture plastics, enzymes and kits used during the duration of this project were obtained from standard suppliers, such as Biorad, Fisher Scientific, Invitrogen, New England Biolabs, Pierce, Promega, Merck Ltd., Qiagen, and Sigma unless otherwise stated.

2.1.1 Tissue culture disposable materials

Tissue culture disposable plastic materials are listed in Table 2.1 below.

Table 2.1. Tissue culture materials and suppliers.

Disposable Materials	Supplier
100 and 35 mm polystyrene tissue culture dishes	Corning
25 tissue culture flasks	Costar and Iwaki
5, 10 and 25 ml pipettes	Corning
6 and 96-well tissue culture microplates	Iwaki
60 x 15 mm polystyrene tissue culture dishes	Corning
75 and 150 cm^2 tissue culture flasks	Iwaki
Cell scrapers	Sigma
Cryo 1°C freezing container	Nalgene
8-well chamber glass slides	Nalge Nunc
15 and 50 ml Falcon Tubes	Corning

2.1.2 Size markers

For DNA, 1Kb+ ladder from Invitrogen was used throughout unless otherwise stated. For protein, Seeblue prestained molecular weight markers (also Invitrogen) were used.

2.1.3 Constructs

The GFP/MBNL fusion construct was kindly donated by Marion Hamshere of the Institute of Genetics, University of Nottingham (Fardaei *et al.*, 2001).

pCre (aka pCAGGS-nls-CRE) was kindly donated by Andras Nagy of the Samuel Lunenfeld Research Institute, Toronto.

2.1.4 Kits

Reagents used throughout this project in kit form are listed in Table 2.

Table 2

Kit	Manufacturer
Qiaprep spin miniprep kit	Qiagen
Endofree maxiprep kit	Qiagen
RNeasy mini kit	Qiagen
DNeasy BLOOD AND TISSUE KIT	Qiagen
Supersignal west pico chemiluminescent substrate kit	Pierce
ECL+ chemiluminescence detection system	Amersham Biosciences
PGEM T-easy vector system	Promega
Ready-to-go DNA labelling beads (-dCTP)	Amersham Biosciences
Nucleon BACC kit	Nucleon
CelLytic NuCLEAR extraction kit	Sigma

2.1.5 Microscopy and photography

A Nikon TNS phase contrast microscope and a Zeiss Axiovert S100 fluorescence microscope were used routinely during the course of this project. Other items of equipment used throughout this project are listed in Table 3.

Table 3

Equipment	Manufacturer
Axiovert 100S Fluorescent microscope	Zeiss
Filter sets:	
EGFP: FITC filter set number 9	Zeiss
Cy3: Rhodamine filter set number 15	Zeiss
DAPI, AMCA and Alexafluor 350: DAPI filter set number 2	Zeiss
Axiophot II with Axiocam digital imaging system	Zeiss
UVP gel documentation system 7500	UVP
EOS 300 SLR	Canon
400ASA colour negative film	Kodak
EOS 350D Digital SLR	Canon

LSM 510 Meta confocal imaging system	Zeiss
X-ray film	Konica and Kodak

2.1.6 Membranes and miscellaneous

Hybond N+ was obtained from Amersham Pharmacia Biotech, Immobilon P PVDF membrane from Millipore, and Na-45 paper from Schleicher and Schulell. Schleicher & Schuell also supplied gel blotting paper. Saran wrap was obtained from Dow.

Snakeskin™ dialysis tubing was obtained from Pierce.

Precast Nupage 4-12% bis-Tris polyacrylamide gels were supplied by Invitrogen.

Calibrite beads used to test FACS recovery modes, were manufactured by BD Biosciences.

2.1.7 Cell-lines and bacterial hosts

Cell-lines and bacterial hosts are listed in Table 4.

Table 4

Cell-line/Strain	Description
Mammalian:	
Cos-7	African green monkey, kidney
HeLa	Human, epithelial, fibroblast
DmtD	Kidney cell-line derived from DM transgenic mouse model Dmt162 (Monckton et al., 1997, see figure 3 for transgene)
3T3	Murine embryonic fibroblast, J2
CGR8.8	ES cell-line derived from strain 129/Ola
DM1 fibroblast	DM1 fibroblast cells lines donated by J.D. Brook, University of Nottingham
Bacterial:	
Top10f'	F´[lacIq Tn10 (TetR)] mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL endA1 nupG
JM109	F′ [traD36 proABlacIqZΔM15] e14-(McrA-) recA1 endA1 gyrA96 thi-1 hsdR17(rK- mK+) supE44 relA1Δ(lac-proAB)
DH5α	F-, φ80dlacZΔM15, Δ(lacZYA-argF)U169, deoR, recA1, endA1, hsdR17(rk-,

	mk+), phoA, supE44, λ-, thi-1, gyrA96, relA1
Stbl2 (for information –not used)	F-, mcrA Δ(mcrBC-hsdRMS-mrr) recA1 endA1 lon gyrA96 thi supE44 relA1 Δ(lac-proAB)

2.1.8 Antibodies

Antibodies used throughout this project are listed in Table 5.

Table 5

Antibody		Supplier
Mouse anti vimentin	530513	Pharmingen
Rabbit anti EXP42 (MBNL1)	Polyclonal	Maurice Swanson*
Mouse anti MBNL1	MCA-MBNL	Universal Biologicals
Mouse anti CUG-BP1	Clone 3B1	Abcam
FITC conjugated rabbit anti GFP	Ab6662-100	Abcam
AMCA conjugated anti-rabbit	111-156-003	Stratech
Alexafluor350 conjugated anti-rabbit	A-21062	Molecular probes
FITC conjugated anti-mouse	Ab50084	Abcam
HRP-conjugated anti mouse	115-035-003	Stratech

*Miller *et al.*, 2000.

2.1.9 Equipment

Equipment used throughout this project is listed in Table 6.

Table 6

Model	Manudacturer
FacsCalibur	BD Biosciences
Genepulser II	Biorad
37ºC Humidified tissue culture incubator	Sanyo
X-cell II gel system	Invitrogen
Novex blot module	Invitrogen
DU 530 UV/Vis spectrophotometer	Beckman

Nanodrop ND-1000 UV/Vis spectrophotometer	Thermo Fisher Scientific
Biometra Uno II thermal cycler	Biometra
Agilent 2100 Bioanalyser	Agilent Technologies

2.1.10 Solutions

2.1.10.1 General

Solutions were made up as described in (Sambrook *et al.*, 1989).

0.5 M EDTA pH 8.0

0.5 M EDTA with NaOH to pH 8.0.

1 kb+ ladder

60 ng/µl 1 kb ladder, 1X DNA loading dye in 1X TBE.

1 M Tris•HCl pH 8.0

1 M Trizma base and HCl to pH 8.0.

10 mg/ml BSA

10 mg/ml (w/v) in 18Ω milliQ water.

10% (w/v) SDS

10% (w/v) SDS in H_2O.

100 X Denhardt's solution

2% (w/v) Ficoll®400, 2% (w/v) polyvinylpyrolidone, 2% (w/v) BSA.

10X DNA loading dye

50% (w/v) Glycerol, 0.25% (w/v) xylene cyanol, 0.25% (w/v) bromophenol blue.

1X TBE

90 mM Trizma base, 90 mM orthoboric acid, 2 mM EDTA.

1X TBST

20 mM Tris•Cl pH7.6, 137 mM NaCl, 0.06% Tween-20.

1X TBSTM2.5

1 X TBST, 2.5% w/v Marvel skimmed milk powder.

1X TBSTM5

1 X TBST, 5% w/v Marvel skimmed milk powder.

2 X RNA loading Dye

5 X DNA loading dye and 0.4% SDS in DEPC-treated water.

2.5 kb molecular ruler

333 ng/µl of 2.5 kb molecular ruler, 1X TE, 1X Orange G loading dye.

20X SSC

3.0 M NaCl, 0.3 M sodium citrate.

20X SSPE

3.0 M NaCl, 0.2 M NaH_2PO_4, 20 mM EDTA, pH 8.0.

3 M Sodium acetate pH 5.2

3 M sodium acetate, pH 5.2

3% H_2O_2

3% (v/v) H_2O_2.

4% Paraformaldehyde

4% Paraformaldehyde (w/v) in PBS-$MgCl_2$.

5X Orange G loading dye

0.06 % (w/v) Orange G, 50 % (v/v) glycerol in H_2O.

70% (v/v) Ethanol

70% (v/v) absolute ethanol in H_2O.

75% (v/v) Ethanol

75% (v/v) absolute ethanol in H_2O.

80% (v/v) Ethanol

80 % (v/v) absolute ethanol in H_2O.

DAPI

10ng/µl in Vectashield mounting medium (Vector Laboratories).

Denaturing solution

0.5 M NaOH, 1.5 M NaCl in H_2O.

Depurinating solution

0.25 M HCl in H_2O.

dNTP

Equal molar amounts of dATP; dGTP; dCTP and dTTP.

Ethidium bromide

Stock solution: 25 mM in H_2O.

Guanidinium thiocyanate, 1M

0.118g/ml (w/v) in DEPC treated water.

Injection buffer

10mM Tris; 0.1mM EDTA prepared by diluting 1M Tris.Cl Ph7.4 and 500mM EDTA (Sigma) with 18Ω MilliQ water. This was sterilised with a 2 micron syringe and filter.

Lysis buffer

50 mM Tris•HCl pH 8.0, 100 mM EDTA pH 8.0, 0.5% (w/v) SDS in H_2O.

Neutralising solution

1.5 M NaCl, 0.5 M Trizma base and HCl to pH 7.5.

PBS

137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 2 mM KH_2PO_4, pH 7.4.

PBS-$MgCl_2$

PBS, 5mM $MgCl_2$, pH 7.4.

PBST

PBS; 0.1% Triton X-100

Phenol

Phenol saturated with 10 mM Tris pH 8.0, 1 mM EDTA.

Phenol:chloroform:isomayl alcohol, 25:24:1

Phenol:chloroform:isomayl alcohol, 25:24:1, saturated with 10 mM Tris pH 8.0, 1 mM EDTA.

Proteinase K

Stock solution: 20 mg/ml proteinase K in filter sterilised 50 mM Tris•HCl pH 8.0.

RNA *in situ* buffer

40% v/v formamide, 1mM DTT, 1mg/ml *E.coli* tRNA, 1mg/ml sonicated salmon sperm DNA, 10% dextran sulphate, 0.2% BSA, 2X SSC.

Southern blot hybridisation solution

7% (w/v) SDS, 1M $NaPO_4$, 2 mM EDTA.

TAE

40 mM Tris•acetate, 1 mM EDTA in H_2O.

TE

10 mM Tris•HCl pH 8.0, 1 mM EDTA pH 8.0 in H_2O.

2.1.10.2 Bacterial culture

Bacterial cultures were grown in LB supplemented with the appropriate antibiotic overnight with shaking. Colonies were grown on LB agar plates with the appropriate antibiotic, X-gal and/or IPTG as stated, inverted, overnight at 37ºC.

Ampicillin
Stock solution: 50 mg/ml (w/v) in H_2O.
Working solution 50µg/ml.

IPTG
Stock solution: 100 mg/ml (w/v) in H_2O.
Working solution 10µg/ml.

Luria-Bertani (LB) medium
3% (w/v) Bacto® tryptone, 0.5% (w/v) Bacto® yeast extract, 1% (w/v) NaCl.
LB agar contained 1.5% (w/v) agar.

SOB medium
2% (w/v) Bacto® tryptone, 0.5% (w/v) Bacto® yeast extract, 0.85 mM NaCl, 0.25 mM KCl pH 7.0 with NaOH. Sterilized 10 mM $MgSO_4$ added prior to use.

SOC medium
0.04% (w/v) glucose in SOB medium.

X-gal
Stock solution: 50mg/ml (w/v) in dimethylformamide.
Working concentration: 50 µg/ml.

2.1.10.3 Cell culture

Tissue culture media such as DMEM, GMEM and Opti-MEM, serum, antibiotics, ESGRO and solutions were obtained from Invitrogen or Sigma.

2.1.10.3.1 General

Culture medium

1 X DMEM; 1 X penicillin and streptomycin solution; 1 X non-essential amino acids; 10% foetal bovine serum.

Trypsin/EDTA solution

Stock solution: 5.0 g/l trypsin, 2.0 g/l EDTA, 8.5 g/l NaCl (stored at -20°C).

Working concentration: 0.5 g trypsin, 0.2 g EDTA•4Na/l in PBS.

100 X Penicillin and streptomycin solution

Stock solution: 10,000 U/ml penicillin, 10,000 μg/ml streptomycin. Penicillin G (sodium salt) and streptomycin sulphate in PBS.

10X Dulbecco's phosphate buffered saline (PBS)

136 mM NaCl, 26mM KCl, 15mM KH_2PO_4, 8mM $Na_2HPO_4 \bullet 7H_2O$.

Foetal bovine serum (FBS)

Origin E.C. Virus and mycoplasma tested.

Gancyclovir

2 mg/ml (w/v) in distilled water.

G418

50 mg/ml (w/v) disulphide salt solution (Sigma).

2.1.10.3.2 ES cells

Culture medium

1X GMEM, 0.25% sodium bicarbonate, 1 X non-essential amino acids, 1X sodium pyruvate, 1X glutamax, 1X penicillin and streptomycin solution, 1% β mercaptoethanol, 5 X 10^6 U leukaemia inhibitory factor (ESGRO).

G418, 50 mg/ml

Powder 1g (w/v). Filter sterilised (0.2 μM).

TVP

0.025% trypsin, 1% chicken serum, 1 mM EDTA, in PBS

Foetal bovine serum (FBS), heat inactivated

Origin E.C. Virus and mycoplasma tested, Heat inactivated. Batch tested for ES cell plating efficiency and colony morphology.

2.2 DNA methods

2.2.1 Preparation of plasmid DNA

For ES cell electroporation and some microinjections, plasmid DNA was isolated from their bacterial hosts by standard alkaline lysis followed by CsCl purification (Sambrook *et al.*, 1989). All other plasmid DNA was isolated using commercially available kits as directed by the manufacturer: Qiaquick Spin Miniprep kits (50µl distilled water used for elution) were used to obtain DNA for cloning and restriction analysis and Qiagen Endo-free Maxi prep kits were used to for the other microinjections and cell transfection.

2.2.2 Purification of DNA

After restriction digest, DNA was purified by electrophoresis through an appropriate concentration of Seakem agarose containing 15µg/ml Ethidium Bromide in 0.5 X TBE. The DNA fragment was retrieved in the following ways depending on the final use.

2.2.2.1 Transfection into cultured cells

The fragment was electrophoresed onto Na45 paper (Scleicher & Schuell), and eluted as directed by the manufacturer.

2.2.2.2 Pronuclear injection

The fragment was electrophoresed onto Na45 paper (Scleicher & Schuell), and eluted as directed by the manufacturer.

Or the fragment was excised from the gel and eluted by electrophoresis into 0.5 X TBE in dialysis tubing. The buffer containing the DNA was applied to an Elutip-d DEAE sephacell column (Scleicher & Schuell) and the DNA collected as described by the manufacturer.

Regardless of the purification method, DNA was diluted to the required concentration using injection buffer. Debris was removed by centrifugation at 14,000 rpm in a microfuge. The top 75% of the solution was transferred to a new tube and the remainder discarded.

2.2.3 DNA extraction from cultured cells

DNA was prepared using a Nucleon or Qiagen blood and tissue kit as directed by the manufacturer.

2.2.4 Preparation of mouse tail lysates

Half to one centimetre of tail tip was placed in a 1.5 ml screw top eppendorf tube with 550 µl of lysis buffer and 15 µl of 20 mg/ml of proteinase K. The tissue was incubated overnight at 60°C. The lysate was vortexed briefly and centrifuged at 21,000 g for 5 minutes to pellet the debris. The supernatant was transferred to a fresh tube and stored at 4°C. One to two µl were used in a 10µl PCR.

If PCRs during genotyping failed, the DNA was further purified. The supernatant (250 µl) was transferred into a fresh tube and an equal volume of phenol added and mixed vigorously by vortexing for 5 seconds. The mixture was centrifuged for 5 minutes at 21,000 *g* to separate the two phases. The upper phase was placed into a clean tube, and 250 µl of phenol:chloroform:isoamyl alcohol (25:24:1) was added and mixed vigorously by vortexing for 5 seconds. The mixture was centrifuged for 5 minutes at 21,000 *g*. The upper phase was removed and transferred into a clean tube. In order to remove any remaining traces of phenol, 250 µl of chloroform were added and vortexed for 5 seconds, followed by centrifugation for 5 minutes at 21,000 *g*. The upper aqueous phase was removed and transferred into a clean tube, and 25 µl of 3 M sodium acetate pH 5.2 was added and mixed briefly by inverting the tube 5 to 10 times. The DNA was finally precipitated by the addition of 500 µl of ice-cold absolute ethanol, followed by incubation at –20°C for at least 1 hour, and 30 minutes centrifugation at 21,000 *g* in a bench top centrifuge. The supernatant was decanted off, the pellet rinsed with 1 ml of ice-cold 80% (v/v) ethanol, and the precipitate collected by centrifugation at 21,000 *g* in a bench top centrifuge. The DNA pellet was either air dried at room temperature for 30 to 60 minutes, or at 4°C overnight. Once dried, the DNA pellet was dissolved in 100 to 200 µl of TE at 60°C for 30 minutes, or at 37°C overnight.

2.2.5 Determination of DNA concentration

DNA samples were diluted 1:100 in H_2O and the spectrophotometer baseline corrected with H_2O. The OD was measured and the concentration calculated as 100 x 50 x OD260 mg/ml. The purity was checked using the 260/280 ratio, which for DNA should approximate 1.8.

Alternatively, DNA concentrations were determined by electrophoresis on agarose gels, followed by densitometry estimation against low molecular weight DNA mass ladder (Invitrogen), and/or Log2 ladder (New England Biolabs).

2.2.6 Oligonucleotides and PCR

The oligonucleotides used in this project were synthesised by Genosys and supplied desalted, except for the LoxP-EGFP and LoxP-Neo forward primers, which were HPLC purified. Fluorescent primers were purified by PAGE.

2.2.6.1 Primer pairs used for construct sequences

Forward and reverse primer sets used to generate DNA fragments for cloning into the founder and targeting constructs are detailed in Table 7.

Table 7

Sequence	Forward and Reverse Primer Sequences	Template DNA
LoxP-EGFP	5'-CACGTAGTGATAACTTCGTATAGCATACATTATA CGAAGTTATGTCGCCACCATGGTGAG-3' 5'-GGATCCTCTAGAAGCTCTAGGCTCGAGTTACTT GTACAGC-3'	pIRES-EGFP
LoxP-Neo	5'-GCTAGCATAACTTCGTATAGCATACATTATACG AAGTTATAGAGCCACCATGATTGAACA-3' 5' GCGGCCGCTCAGAAGAACTCGTC-3'	pCIneo
ERDA	5'-GGATCCTGCATGCATCATATGTTGCTTTTC-3' 5'-TCTAGACCTTCCTTCCACCTTATTTTTC-3'	Human genomic DNA
DM repeat region	Round1 (DM-A) 5'-AGTGCAGTTCACAACCGCTCCGAGC-3' Round 1 (DM-BR) 5'-GTGGAGGATGGAACACGGAC-3' Round 2 (*Xho*I DM-H) 5'-CTCGAGTCTCCGCCCAGCTCCAGTCC-3' Round 2 (*EcoR*I DM-BR) 5''GAATTCGTGGAGGATGGAACACGGAC-3'	Patient sample numbers: 175 1001; 101 5053; 137 3666
BGH PolyA	5'-GAATTCAGCTCGCTGATCAGCCTC-3' 5'-GGATCCCCAGCTGGTTCTTTCC-3'	pIRES-EGFP
Thymidine Kinase	5'-GCCTTGTCGACGCCACCATGGCTTCGTACC-3' 5'- GAATTCTCAGTTAGCCTCCCCCATCTC-3'	HSV UL97 gene
PGK Promoter	Isolated on a 500bp *Taq*I restriction fragment, with filled in recessed termini.	pPNT
Neo cassette	The *Xho*I site was removed from pPNT (this was situated within the neo cassette fragment). The cassette was then isolated on an XbaI (filled-in terminus) –*Not*I fragment. This was then cloned into	pPNT

the *Not*I and *Sfo*I (filled-in terminus) sites of the TK construct.

Note –Areas underlined comprise 5' extensions which contain restriction sites and/or a *LoxP* site. They have no homology to the template DNA.

2.2.6.2 PCR Primers and conditions

Polymerase chain reactions were carried out on DNA and first strand cDNA templates in a biometra Uno II thermocycler in 1 X PCR buffer as detailed below, with 0.5U *taq* polymerase and 10 picomoles of forward and reverse primers per 10 µl reaction, unless otherwise stated. Genomic samples were denatured once at 94ºC for 2 m before cycling. Conditions are listed in Table 8.

Buffer A: 45mM Tris•Cl pH 8.8; 11mM $(NH_4)_2SO_4$; 4.5mM $MgCl_2$; 0.113mg/ml BSA; 4.4µM EDTA; 1mM each dNTP.

Buffer B: 16mM $(NH_4)_2SO_4$; 67mM Tris-HCl (pH 8.8 at 25°C); 0.01% Tween-20; 1 mM $MgCl_2$; 200nM dNTP.

Buffer C: 16mM $(NH_4)_2SO_4$; 67mM Tris-HCl (pH 8.8 at 25°C); 0.01% Tween-20; 6pMoles primer, 2.5 mM $MgCl_2$; 500mM dNTP.

Table 8

PCR: Primers and products	Cycles	
SP-PCR[A] (DMA-BR)	94°C 45s	
Forward: 5'- AGTGCAGTTCACAACCGCTCCGAGC -3'	64°C 45s	X 28, 72°C10m.
Reverse: 5'- CGTGGAGGATGGAACACGGAC -3'	72°C 3m	
Cre excision[A]	94°C 45s	
Forward: 5'-CAAGGTTACAAGACAGGTTTAAGGAG-3'	60°C 45s	X 28, 72°C10m.
Reverse: 5'-GGACACGCTGAACTTGTGG-3'	72°C 1m	
EGFP[A]	94°C 45s	
Forward: 5'-CGACCACATGAAGCAGCACGA-3'	66°C 45s	X 25, 72°C 10m.
Reverse: 5'-GTTGTCGGGCAGCAGCAC-3'	72°C 1m	
DMH-BR[A]	96°C 45s	
Forward: 5'-TCTCCGCCCAGCTCCAGTCC-3'	65°C 45s	X 25, 72°C 10m.
Reverse: 5'-CGTGGAGGATGGAACACGGAC-3'	72°C 3m	

Concatomer[B]	94°C 45s	
Forward: 5'-GG AGTGGGATTTTGGTAGGCATC-3'	62ºC 45s	X 25, 72°C 10m.
Reverse: 5'-GGGCTATGAACTAATGACCCCG-3'	72°C 45s	
Clcn1[B] (6pMoles primer)	94°C 45s	
Forward: 5'-GGAATACCTCACACTCAAGGCC-3'	63°C 45s	X 32, 72°C 10m.
Reverse: 5'-CACGGAACACAAAGGCACTGAATGT-3'	72°C 45s	
cTnnt2[C]	94°C 45s	
Forward: 5'-GCCGAGGAGGTGGTGGAGGAGTA-3'	72°C 45s	X 32, 72°C 10m.
Reverse: 5'-GTCTCAGCCTCACCCTCAGGCTCA-3'	72°C 45s	
Int-Neo[B]	94°C 45s	
Forward: 5'-CAAGGTTACAAGACAGGTTTAAGGAG-3'	57°C 45s	X 30, 72°C 10m.
Reverse: 5'-GCAGCCTCTGTTCCACATAC-3'	72°C 1m	
MuSF[A]	94°C 45s	
Forward: 5'-GCCCCTGCCTCACCGTATAG-3'	62°C 45s	X 29, 72°C 10m.
Reverse: 5'-CTGGGGTCCACCACTTCAAG-3'	72°C 3m	
DM-QR[A]	94°C 45s	
Forward: 5'-CACTGTGGAGTCCAGAGCTTTG-3'	62°C 45s	X 28, 72°C 10m.
Reverse: 5'-CACTGTGGAGTCCAGAGCTTTG-3'	72°C 45s	
GAPDH[A]	94°C 45s	
Forward: 5'-AACGACCCCTTCATTGAC-3'	60°C 45s	X 27, 72°C 10m.
Reverse: 5'-TCCACGACATACTCAGCAC-3'	72°C 45s	
CMV[A]	94°C 45s	
Forward: 5'-CATGTCCAATATGACCGCCATG-3'	58°C 45s	X 25, 72°C 10m.
Reverse: 5'-GCAGCCTCTGTTCCACATAC-3'	72°C 1m	

Superscript letters denote which PCR buffer was used

2.2.6.3 Small pool PCR

Small-pool (SP-PCR) analyses (Jeffreys *et al.*, 1994; Monckton *et al.*, 1994) were performed to assess the levels of trinucleotide repeat instability in different batches of *Eco*RI restricted pLoxEGFP250 constructs used throughout this project. Single molecule SP-PCR, where the DNA template is diluted to a low concentration, was used to amplify a rare 750 bp repeat from *Hin*dIII restricted patient DNA.

DNA was serially diluted in 1X TE, containing 0.1 µM of carrier primer DM-A, the forward primer for subsequent PCR, to give between 10 and 10,000 DNA molecules per µl. The amplification of 1µl of each DNA solution was carried out in a final volume of 7 µl, with 0.175 U of Taq DNA polymerase, 0.2 µM of each primer DM-A and DM-BR, in 1X PCR buffer A. Cycling conditions are detailed in (Table 8).

2.2.7 Gel electrophoresis

DNA molecules were separated using 0.6-1.5% agarose gel electrophoresis depending on the fragment size. NuSieve agarose gel electrophoresis was used to separate the products of the *Clcn1*and *Tnnt2* PCR products (2% (w/v) NuSieve: 1% (w/v) agarose). Gels were prepared in 0.5 X TBE and when cool ethidium bromide was added to a final concentration of 500 nM. Samples and DNA size markers were mixed with 1/10 volume of DNA loading dye and electrophoresed at 80 volts (V) at room temperature in 0.5X TBE until the fragments had separated. The gels were visualised using a UV transilluminator (wavelength 254 nm) and photographed.

2.2.8 Southern transfer

DNA was transferred to Hybond N+ membrane using the squashblot method. After electrophoresis, gels were rinsed in deionised water, incubated in depurinating solution for 10 minutes, denaturing solution for 30 minutes and neutralising solution for 30 minutes. Incubations were performed with gentle shaking.

A piece of Hybond N+ nylon membrane the same size as the gel was wet in deionised water then dipped in neutralising solution. A piece of Saran Wrap, larger in size than the gel, was placed on the bench, with a piece of gel blotting paper, the same size as the gel and wet in neutralising solution, layered on top. The gel was inverted and placed on the Saran Wrap, and the membrane layered on top of the gel. Any air bubbles trapped under the membrane were carefully removed. Two sheets of gel blotting paper were rinsed in neutralising solution and layered onto the membrane. The blot was topped with a thick layer of paper towels, a glass plate and a weight. The DNA was transferred from the gel onto the membrane by capillary action for 3 to 16 hours. The blot was dismantled and the membrane placed on a piece of dry gel blotting paper DNA side up, and baked for at least 20 minutes in an 80°C oven. The DNA was cross-linked to the membrane by exposure to 1,200 J/m2 UV light.

2.2.9 Preparation of radiolabelled DNA

Fragments (30ng linearised template DNA and 5ng 1Kb+ ladder) were radiolabelled using Ready-to-go DNA labelling beads (Amersham Biosciences) and α[32P]dCTP, as directed by the manufacturer.

2.2.10 Southern hybridisation

Filters to be hybridised were placed in a hybridisation bottle and rotated at 65°C for 30 minutes with 10 ml of Southern blot hybridisation solution. The radiolabelled probe was denatured at 100°C for 5 minutes and quenched on ice for two minutes before being added to the bottle, which contained 10 ml of fresh Southern blot hybridisation solution. Hybridisation was performed at 65°C overnight in a rotating hybridisation oven.

Following hybridisation, the filters were briefly washed inside the bottle in 0.2% (w/v) SDS, 0.2X SSPE at room temperature. The filters were then washed twice in the same washing solution for 30 minutes at 65°C. Finally the filters were transferred onto a flat tray and washed by gently shaking them at room temperature in the same solution. The filters were baked at 80°C until dry, and exposed to X-ray film at room temperature. Autoradiographs were developed after approximately 4 hours exposure.

2.2.11 Cloning techniques

2.2.11.1 Endonuclease restriction of DNA

Restriction digests were carried out to obtain DNA fragments for cloning, transfection and microinjection, and to verify inserts and whole constructs based on the fragment size or banding pattern obtained.

DNA (200ng-1µg) was restricted in a 10µl volume with 5-10 units of enzyme in the recommended restriction buffer at 37^{o}C for 1.5 to 2 hours. *Eco*RI and *Xho*I double digests were carried out by first incubating in *Xho*I buffer for 1 hr, then diluting the sample to 20µl in the recommended *Eco*RI buffer and incubating for a further hour. This avoided star activity by *Eco*RI due to low salt buffer. Where more DNA was required the reactions were scaled up proportionately.

2.2.11.2 Filling in of recessed 3' ends

Residual restriction enzymes were first heat inactivated by heating to 70ºC for 10 m One unit of Klenow (DNA polymerase 1 large fragment) (Promega) and 1µl 10mM dNTP mix, was added to 10µl of restriction endonuclease digestion mix and incubated for 30 m at 37ºC. The enzyme was then denatured by heating to 70ºC for 10 minutes.

2.2.11.3 Ligation

Linear vector DNA (50 ng) was incubated with insert DNA at 8°C overnight in the buffer provided, with 1 unit T4 ligase (Promega).

Ligations were carried out with a 3:1 insert:vector molar ratio. Where the concentration of insert was not known because the sample was too small, 3 ligations were set up with 1, 3 and 5 µl insert.

2.2.11.4 Transformation of bacterial hosts

On ice, 0.5 µl of the ligation reaction mix was added to 20 µl commercially prepared chemically competent cells and mixed by pipetting. After 30 minutes on ice, the tubes were place at 42°C for 45 s then on ice for 2 m, 180 µl SOC medium was added and the samples incubated at 37°C for 30 m to express the β-lactamase gene. One and nine tenths of the transfection mix were plated onto 50 µg/ml ampicillin LB plates, allowed to dry and incubated inverted overnight at 37°C.

2.2.11.5 Culture of bacteria

A single colony was picked using a sterile toothpick and added to either 5ml or 100ml of selective LB medium, and incubated at 37ºC overnight at 300 rpm in an orbital shaker. Cultures were generally processed for DNA or storage immediately, or stored at room temperature until required. Glycerol stocks were first streaked out to single colonies on selective L-agar plates and incubated inverted overnight, at 37ºC.

2.2.11.6 Glycerol stocks

Overnight bacterial cultures were mixed 1:1 with 40% Glycerol:peptone and stored at -80°C.

2.2.12 Transfection of DNA into cells

2.2.12.1 Transient

Most transfer of DNA into cultured cells was transient, in that the cells were used for analysis usually 24 hours later (unless otherwise stated), without selection for stable integrants. As such only a proportion of the cells resulting from transfection were positive for the construct –typically ~60% based on the EGFP positive cell count in fluorescent micrographs of 3T3 cells transfected with pLoxEGFP.

2.2.12.2 Stable

Cells with stably integrated transgenes were selected for using G418 at a concentration determined by trial and error. Cells were plated with a series of G418 concentrations in normal growth media. The lowest concentration leading to cell death after ten days treatment was chosen for selection.

Hela: 425 µg/ml

Cos7: 425 µg/ml

DmtD2976 Kidney: 300 µg/ml

CGR8.8 ES cells: 200 µg/ml

Transfected cells were grown in selective medium until colonies appeared –usually 10 days. The plates were washed twice in PBS. Colonies were dissociated using 50µl trypsin/EDTA or TVP as described previously (2.5.2), within cloning rings (Sigma) placed over the colony. The suspended cells were then transferred to 96-well plates containing normal growth medium, and incubated at 37ºC. The clones were expanded by allowing the cells to become 95% confluent, then dissociating the cells using trypsin. The suspension was then transferred to one well of a 6-well plate also containing growth medium. After the cells became 95% confluent, they were again dissociated and transferred to a 25cm^2 tissue culture flask containing growth medium. Once the cells reached 80% confluency, a proportion was frozen (2.5.4) for storage.

2.2.12.3 Liposome based

Liposome-based transfection was used for all applications except ES cell transfection. FACS samples were prepared using only Lipofectamine2000 (Invitrogen).

Either Transit LT1 (Miras) or Lipofectamine2000 (Invitrogen) transfection reagents were used as directed by the manufacturer. Briefly, for each well, the DNA and the Lipofectamine2000 reagent were diluted into separate tubes containing 250µl Opti-MEM (Invitrogen), the diluted DNA was added to the diluted reagent and gently rocked to mix. The tube was allowed to rest at room temperature for between 20 and 60 minutes whilst the transfection complexes formed. Meanwhile, the medium was removed from the cells, which were then washed with PBS. Serum-free medium was added and the complexes

added dropwise. The plates were rocked to mix and incubated for 4 hours. After this time, the medium was supplemented with foetal bovine serum to 10%. The medium was changed after 24 hours. Before preparing samples for FACS, the transfections were optimised by varying the ratio of pLoxEGFP DNA to reagent, to give the highest number of fluorescent cells. Thereafter, 4 µg of DNA and 7.5 µl Lipofectamine was used for each 35mm^2 transfection.

2.2.12.4 Electroporation

DNA was introduced into ES cells using the Genepulser II (Biorad) electoporation apparatus. First two 75 cm^2 flasks of cells at ~80% confluency were trypsinised as described in 2.5.2. The cells were washed 3 times in PBS in 15 ml falcon tubes, each time collecting the cells by centrifugation at 1000 rpm in a bench top centrifuge. The final pellet was resuspended in 0.5 ml PBS. The DNA was diluted with 150 µl PBS, then added to the cells and mixed gently. The cell-DNA mixture was then transferred to a genepulser cuvette and the current applied. The settings were 0.8KV, capacitance 3 for 0.1 second. The cells were left to recover for 5 minutes at room temperature. The transfected cells were diluted to 6mls with ES cell culture medium then divided between 6, 100mm diameter dishes containing 10mls ES cell growth medium and incubated for 24 hours at 37ºC. After this time, the medium was supplemented with G418 antiobiotic, to select for stable transfectants.

2.3 RNA methods

Care was taken to avoid contamination with RNAses. Gloves were worn at all times. All solutions were made up with water treated with 0.1% DEPC at 37ºC for two hours followed by autoclaving. Disposable plastic falcon tubes were used to store and make-up solutions, and disposable prepacked RNase-free pipette-tips were used. Benches were swabbed down with RNaseZap and rinsed with DEPC-water.

2.3.1 Isolation of RNA

Total RNA was prepared for all applications using the RNeasy RNA miniprep kit as directed by Qiagen. Purified RNA was later DNAse 1 treated if required using the RNeasy minicolumns and a modified protocol: To each 100 µl RNA, 12 µl reaction buffer and 10

µl RQ1 RNase-free DNAse1 (Promega) was added. The reaction was incubated at 37ºC for 30 minutes. Ten microlitres of stop solution (Promega) was then added and the reaction incubated at 65ºC for 10 minutes. The resulting mix was added to 350 µl RNeasy RLT buffer (provided) and mixed. Absolute ehanol (250 µl) was added and mixed, and the solution added to the RNeasy spin column. The rest of the protocol was then followed as for the isolation of RNA from animal cells.

2.3.2 Determination of RNA quality and concentration

The RNA concentration was analysed using the Nanodrop ND-1000 spectrophotometer.

The degradation level was either assessed using the Agilent 2100 Bioanalyser, or by electrophoresis on a guanidinium thiocyanate agarose gel.

To make the guanidinium thiocyanate agarose gel the gel tray and equipment were soaked in 3% H_2O_2 for 10 m, then rinsed in DEPC treated water. One gram of agarose was added to 100 mls TAE and microwaved until dissolved. Once cooled to hand hot, 2 µl 10 mg/ml ethidium bromide and 0.5 µl 1M guanidinium thiocyanate were added and the gel poured. RNA samples were prepared by adding 9 µl 2 X RNA loading dye to 1 µl RNA and incubation at 85ºC for 10 minutes. Samples were loaded and run at 80v for 45 m in TAE, then photographed under UV.

2.3.3 Complementary DNA synthesis

First strand cDNA was synthesised using Superscript II (Invitrogen) from 5 µg total RNA template using 500 nM random hexamers in a final volume of 20 µl as directed by the manufacturer. The resulting cDNA was stored at -20ºC until required.

2.3.4 Fluorescent *in situ* hybridisation

Cells attached to coverslips were washed twice in PBS; 5 mM $MgCl_2$ at 37ºC and transferred to a clean 6-well plate before fixing for 10 m at room temperature in 4% paraformaldehyde. Excess fixative was removed by washing 3 times for 10 m each in PBS-$MgCl_2$ at room temperature. Cells on each coverslip were pre-treated with 40% Formamide; 2 X SSC for 10 m, the solution removed and replaced with 200 µl *in situ* hybridisation buffer containing 10 pMol Cy3-$(CAG)_{10}$, Cy3-$(CTG)_{10}$ or Cy3-GFP. A square of parafilm was carefully placed over the cell layer and the whole plate sealed with tape. This was placed in a sealed moist chamber, at 37^{o}C overnight. Excess probe was

removed by a series of washes increasing in stringency: 2 X SSC twice for 30 m and 1 X SSC twice for 30 m. The slides were mounted with Vectastain containing DAPI at 100ng/ml. The edges of the coverslips were sealed using ClariOn mounting medium (Sigma).

2.3.5 Double labelling with ICC and FISH

Cells were fixed and washed as described previously (2.3.4). The cells were then fixed briefly in 70% ethanol for 10 m and rehydrated in PBS-$MgCl_2$ for 10 m. Non-specific sites were blocked by incubating in PBST for 30 m. The primary antibody was diluted in PBST (anti-MBNL 1:100 or anti-vimentin 1:200) and incubated with the cells for 1hr at room temperature, or overnight at 4ºC. The antibody was removed by washing three times for 5 minutes with PBST. Hybridisation of the secondary antibody (alexa 350 anti-rabbit 1:200 or FITC anti mouse 1:200) was carried out in the same way as the primary antibody. The antibody was removed by washing three times for 5 minutes with PBST. The protocol described for fluorescent *in situ* hybridisation (2.3.4) was the followed from the 40% formamide; 2 X SSC pre-treatment step.

2.3.6 Microarray analysis

RNA samples were prepared using the Qiagen RNeasy mini kit without DNase treatment as requested by the Molecular Biology Support Unit at the University of Glasgow, who carried out the target preparation, hybridisation and normalisation calculations. Whole transcript levels were assessed using the Affymetrix mouse genome 430 2.0 array, and for exon level differences, the Affymetrix mouse exon 1.0 ST array was used. Whole transcript gene lists were generated by Pawel Herzyk of the MBSU, using rank product and iterative group analysis (Breitling *et al.*, 2004; Breitling *et al.*, 2004). Partek Genomics Suite® software was used to identify mis-spliced genes within the exon array data.

2.4 Protein methods

2.4.1 Isolation of protein

Total protein was extracted from cells using the proprietary reagent CelLytic M lysis reagent from Sigma as directed, which isolates cytoplasmic and nuclear proteins.

Nuclear and cytoplasmic fractions were obtained using the Sigma CelLytic NuCLEAR extraction kit as recommended by the manufacturer.

2.4.2 Determination of protein concentration

Protein concentration was estimated using the Bradford assay (Bradford, 1976). Twenty five µl of sample or BSA standard was mixed with 750 µl Coomassie Plus reagent (Pierce) at room temperatiure and left on the bench for 10 m. Using varying dilutions of 10 mg/ml BSA to give standards of: 10 mg/ml; 2 mg/ml; 1.5 mg/ml; 1 mg/ml; 0.75; 0.5; 0.25; 0.025 and 0 mg/ml, the Nanodrop UV/Vis spectrophotometer software was used to set a standard curve. Test samples were then read using the calibrated machine.

2.4.3 Immunodetection of protein

Proteins were separated by 4-12% PAGE and transferred to Immobilon P PVDF membrane using a Novex blot module and the X-cell II apparatus as recommended by the manufacturer. Briefly, to each 12 µl of protein sample, 3 µl of 4X NuPAGE LDS sample buffer and 12 µl 10X NuPAGE reducing agent was added, and the samples heated to 70ºC for 10 m. They were then loaded onto the polyacrylamide gel and run in 1 X NuPAGE MOPS running buffer. After separation the proteins were transferred to the membrane in 1 X transfer buffer consisting of 1X NuPAGE transfer buffer; 0.1%ml NuPAGE antioxidant and 10% Methanol, at 30 V for 1 hour.

Membranes were pre-wet in methanol, then equilibrated in 1X TBST for 20 m, changing the solution after 10 m. Non-specific binding sites were blocked by incubation overnight in 1X TBSTM5, shaking slowly at 4ºC. The primary antibody was diluted 1:2000 for both anti-CUG-BP1 and anti-MBNL1 in 1X TBSTM2.5, added to the membrane and incubated overnight, again at 4ºC. The membrane was washed four times for 15 minutes each at room temperature, in 1X TBSTM2.5. The HRP-congugated secondary antibody was diluted 1:10,000 in 1X TBSTM2.5 and added to the membrane. Incubation was for 2 hrs at room temperature. The membrane was washed as described after the primary antibody incubation. A final wash of 15 m duration at room temperature was performed in 1X TBST. The chemiluminescent substrate mix was prepared as directed by the manufacturer (Supersignal, Pierce or ECL+, Amersham Biosciences) then pipetted onto the membrane and left for 5 minutes at room temperature. The membrane was blotted on Whatman paper to remove the surface liquid and covered with Saran wrap. The membrane was exposed to x-ray film for varying amounts of time, typically 30 seconds, and the films developed.

2.5 Cell culture methods

All reagent were pre-heated to 37ºC before use. ES cells were grown on gelatinised plates. Ten mls of 0.1% gelatin was placed in a 25 cm^2 flask and incubated at 4ºC overnight, or until required. Before use the gelatin was removed.

2.5.1 Feeding cultured cells

The medium was changed daily during ES cell culture and every 3 days for all other cultures unless otherwise stated.

2.5.2 Subculturing cells

Cells growing in 25 cm^2 flasks were washed twice in PBS before the addition of 0.5 ml trypsin/EDTA solution, or TVP solution for ES cells. The liquid was rocked gently over the cell-layer, then incubated at 37ºC for 5 m. The flask was tapped gently to dislodge the cells and a cell-scraper used to loosen the remaining cells if necessary. Ten mls of culture medium was added to stop the trypsin reaction. One tenth to one third (ES cells one third) of the cell suspension was transferred to a new flask, and 10 mls of culture medium added. The cells were rocked to mix and placed back into the incubator. Reagents were scaled proportionately for other dish and flask sizes.

2.5.3 Thawing frozen cells

Frozen cell samples stored in liquid nitrogen were quickly thawed by immersion of the cryovials in a water bath at 37°C. Five ml of fresh warm medium were added to the cells, and mixed gently by repeated pipetting. The cells were then collected by centrifugation at 200 *g* for 5 minutes, resuspended in fresh culture medium, and finally plated on a tissue culture flask or plate.

2.5.4 Freezing live cells for storage

A single cell suspension was obtained following trypsin digestion of cells between 80 and 95% confluency, as described previously (2.5.2). Cells were precipitated by centrifugation at 200 g for 5 minutes, resuspended in fresh culture medium at a concentration of ~1 x 10^6 cells/ml (MEF and ES cell culture medium was supplemented with 50% foetal calf serum for freezing) and transferred into 2 ml cryovials. Dimethylsulphoxide (DMSO) was added to a final concentration of 10% (v/v) (5% v/v for ES cells), and then mixed well by gently

inverting the tubes. The vials were transferred into a freezing container with isopropanol at room temperature, and then slowly cooled to -70°C over at least 4 hours. Finally the frozen samples were moved to liquid nitrogen, where they were kept until needed.

2.5.5 Determination of cell concentration

Following trypsin digestion and neutralisation with standard growth medium, cells were counted on a haemocytometer, using a phase contrast microscope. The haemocytometer was covered with the coverslip and a drop of cell suspension was dropped at the edge of the coverslip on both sides of the chamber. At least 100 cells were counted, and the number of cells/ml calculated.

2.5.6 Establishment of primary cell-lines

2.5.6.1 Transgenic tail cell-lines

*Tg*5 transgenic mice tails were tipped and kept on ice for up to 30 minutes until processed. Primary cultures were established by the explant technique. Tails were wiped using a tissue dipped in 70% ethanol, and the skin removed using sterile scissors and forceps. The skin was opened out and pressed hard onto a tissue culture dish inside down. Cuts were made 1mm apart along the length of the tail, still attached at the top and fanned out to separate. 4mls of medium was added dropwise so as not to dislodge the tissue and the dish incubated at 37ºC in a humidified 5% CO_2 atmosphere. The standard growth medium consisted of Dulbecco's modified Eagle medium with high glucose (DMEM) supplemented with 10% foetal bovine serum (FBS), 1 X non-essential amino acids, 100 U/ml of penicillin and 100 µg/ml of streptomycin. The dish was not moved for the first 3 days. The medium was changed weekly until cells became 60-80% confluent they were subsequently subcultured at a 1:3 or 1:4 ratio during the first 5 passages and at a higher ratio (varying from 1:5 to 1:10) thereafter, as described in section (2.5.2).

2.5.6.2 MEF feeder cells

2.5.6.2.1 Preparation

All solutions were maintained at 37ºC. Three pregnant FVB/N mice 14 days *post coitum* were sacrificed by cervical dislocation. The embryos were collected by dissection and sacrificed by cervical snipping. The embryos were decapitated and the livers removed. The

embryos were stored in PBS for up to 30 minutes until all were processed. The embryos were washed three times in PBS then cut into very fine pieces within the 50 ml falcon tube and again washed three times in PBS. All the supernatant was removed and 15 ml trypsin-versene added. The contents were incubated at 37ºC for 20 minutes with gentle agitation. Five ml of serum was then added to a fresh tube. Ten mls of DMEM was added to the embryo tube and the supernatant transferred to the serum tube. The volume was made up to 50 ml with DMEM, then centrifuged at 1000 rpm in a table-top centrifuge for 5 minutes. The pellet was resuspended in general culture medium and the tissue allowed to settle for 2 minutes. The supernatant was transferred to a fresh tube. The cells were counted using a haemocytometer and all but 9 X 10^7 cells frozen for storage (2.5.4). The reserved cells were inoculated into two 150 cm^2 tissue culture flasks and incubated until almost confluent. The cells were then trypsinised (2.5.2) and the total amount used to inoculate four 150 cm^2 tissue culture flasks which were incubated until confluent. These cells were also trypsinised and used to inoculate ten 150 cm^2 tissue culture flasks. Once these flasks were almost confluent the contents were arrested for growth by irradiation.

2.5.6.2.2 Growth arrest

Once the cells were confluent the flask lid was tightened the lid of the flasks and exposed to the gamma source for 3000 rads. The flask was then swabbed with 70% ethanol and returned to the incubator before trypsinisation as described previously (2.5.2) and freezing also as described previously (2.5.4).

Use of a gamma source was kindly permitted by The Beatson Institute for Cancer Research in Glasgow.

2.6 Pronuclear injection

Pronuclear injection was carried out by the Molecular Biology Support Unit, at the University of Glasgow, with DNA diluted in injection buffer. Mouse strains are detailed in Table 9.

2.6.1 Mouse strains

Table 9

Strain	Use
B6D2 -F1 Hybrid of C57B6/DBA2	Donor
ICR	Foster mother
T145	Vasectomised male

3 Design and generation of a murine model of DM1 pathogenesis

3.1 Synopsis

Myotonic dystrophy is a multisystemic disorder, not purely affecting muscle. The type 1 mutation comprises an expansion of a CTG trinucleotide repeat within the 3' UTR of the *DMPK* gene, and in type 2 the expansion of a CCTG repeat within the first intron of *CNBP*. The array-length determines the severity and age of onset of the disease. In both mutations, the repeats are transcribed but not translated. The RNA becomes trapped within the nucleus in the form of ribonuclear foci, which co-localise with the double stranded CUG binding protein MBNL1. Single stranded CUG binding protein CUG-BP1 works dynamically with MBNL1 to regulate alternative splicing during development to control mRNA isoform switching. Recruitment of MBNL to RNA aggregates results in depleted reserves within the nucleoplasm, affecting the MBNL1/CUG-BP1 balance, which leads to gene missplicing. As such, DM is an RNA mediated disease, but how this RNA brings about such pleiotropic effects is still unclear. Recent research also indicates a more direct involvement of CUG-BP, and other potential factors may be important. In order to delineate cause and effect we have attempted to use the Cre-*lox* system for the conditional expression of a toxic RNA in the mouse, to mimic the pathogenesis of myotonic dystrophy type 1 with temporal and spatial control. The nature of the transgene was essentially an *EGFP* reporter linked to an expanded CTG repeat within the 3'UTR, controlled by Cre-*lox* dependent excision of an SV40 transcriptional stop signal. Repeat tracts of 5 and 250 repeats were cloned to recreate the normal and mutant alleles. The transgene was assembled, and each component and the Cre-*lox* mechanism tested and shown to be functional. Tracts of 800 repeats were originally intended to be cloned to reflect the congenital form, but proved unsuccessful due to probable rearrangements within the bacterial host. Pronuclear injection was used to generate the mouse model: One mouse line was established with the normal transgene, but attempts to produce the mutant mouse were unsuccessful.

3.2 Introduction

Myotonic dystrophy type 1 (DM1) is one of a large group of inherited human disorders associated with the expansion of trinucleotide DNA repeats. It is an autosomal dominant

disease showing genetic anticipation and is characterised clinically by myotonia, muscle weakness, cardiac conduction defects, cataracts, mental retardation, premature frontal balding and testicular atrophy in males, and reduced fertility in females (Harper, 1998). So far two types of myotonic dystrophy have been characterised, type 1 and type 2. Both mutations are dominant, non-coding expansion disorders, consisting of expanded arrays of repeats located within the untranslated region of genes apparently unrelated to each other. The first consists of a CTG repeat expansion within the 3' UTR of *DMPK* (Fu *et al.*, 1992; Mahadevan *et al.*, 1992) -a protein kinase putatively involved in cell elongation and fusion during skeletal muscle morphogenesis (Beffy *et al.*, 2005). Type 2 consists of a CCTG repeat expansion within the first intron of exon 1 of the *CNBP* transcription factor gene (Liquori *et al.*, 2001) thought to be involved in cell proliferation and tissue patterning during anterior-posterior axis, craniofacial and limb development (Shimizu *et al.*, 2003), and gametogenesis (Liu *et al.*, 2005). Comparisons of symptoms of the two types of myotonic dystrophy show remarkable similarity, with the most notable difference being the lack of any congenital form of the disease in DM2 (Day *et al.*, 2003). How these expansions lead to the DM phenotype is still unclear. Many mouse models have been produced to address this (see chapter 1 for more detail), but none of the models to date have reproduced the complex multisystemic effects of DM pathogenesis. Knockout models that mimicked the haploinsufficiency of *DMPK* or surrounding genes perhaps affected by chromatin remodelling by the repeat, such as *SIX5*, only exhibited partial DM symptoms of mild myopathy and cardiac conduction defects (Berul *et al.*, 2000; Jansen *et al.*, 1996) and cataracts respectively (Klesert *et al.*, 2000; Sarkar *et al.*, 2000), but not myotonia, the benchmark symptom of the disease. As more mouse models were produced focussing on the *DMPK* 3' UTR, it became apparent that only those expressing the repeat showed further symptoms characteristic of the disease: Results using an expanded CUG repeat-expressing mouse showed late onset myotonia, testicular atrophy and infertility (Monckton *et. al.*, unpublished results; also in chapter 1 of Wells *et al.*, 1998). The discovery of the DM2 mutation, where patients suffer from a similar single locus mutation and multi-systemic phenotype in the apparently unrelated gene *CNBP*, concentrated hypotheses towards RNA mediated pathogenesis. Evidence to support this came from a correlation between the length of the repeat and the severity of the disease; the *DMPK* or *CNBP* message from the mutant chromosome becoming trapped in the nucleus (Davis *et al.*, 1997); sequestration of RNA CUG binding protein MBNL1 to these foci (Fardaei *et al.*, 2002) and that levels of RNA CUG binding protein CUG-BP1 are altered in DM1 patients (Timchenko *et al.*, 2001). More recent mouse models also support an RNA based

mechanism. Mankodi *et al*. recreated the myotonia and muscle wasting characteristics of adult-onset DM1 by the expression of an expanded CUG repeat alone, lacking any associated *DMPK* sequences, from the human skeletal *actin* promoter {Mankodi, 2000 #13}). They attributed the myotonia to missplicing of *chloride channel 1* gene *Clcn1* (Charlet *et al.*, 2002; Mankodi *et al.*, 2002). Subsequently, splicing defects were found to explain many of the symptoms in DM (Kuyumcu-Martinez *et al.*, 2006; for review, Mankodi *et al.*, 2002). MBNL knockout mice and mice over-expressing CUG-BP1 demonstrated the same tissue-specific missplicing seen in myotonic dystrophy patients (Kanadia *et al.*, 2003; Ho *et al.*, 2005). Although mouse models have now addressed many of the aspects of DM symptoms and pathogenesis, there are still unanswered questions. Infertility, sudden cardiac death, mental retardation, lethargy, cataracts and congenital onset are amongst some of the features still not fully explained.

It seemed plausible that if the pathogenic mechanism between DM1 and DM2 was the same, then the differences must lie in the levels, and the patterns of expression during development and between different tissue types. In ideal terms, to generate a mouse model of any human genetic disease, one would modify the orthologous murine gene to contain the human mutation. This however would only allow replication of the disease in its multisystemic entirety. In order to delineate cause and effect we wanted to generate the ability to express long and short expansions temporally and/or spatially by limiting the expression of repeats in the mouse: temporal to gain insight into the pathogenesis of the congenital form; and spatial to link tissue specificity with the symptoms directly.

3.2.1 Conditional expression

Modelling the congenital form of DM1 by expressing large numbers of repeats in mice has not yet been successful. The severity of the phenotype would be unpredictable and could result in the morbidity or reduced viability of the line. By using a conditional system, tissue specific expression patterns can be limited maintaining function in other organs, and in the inactive state, promote fecundity of the line. In order to separate the phenotypic aspects of myotonic dystrophy and to overcome fertility problems caused by the ubiquitous expression of the expanded repeat tract (seen in *Te*162 mice, Monckton *et al.*, 1998 and unpublished data), we aimed to produce a conditional CUG repeat expressing mouse model using Cre-*lox* site-specific excision. Many site-specific recombinase systems have been described from bacteria and yeast and their reactions vary in complexity. The yeast *Flp-FRT* and bacteriophage P1 Cre-*lox* mechanisms have been the most widely used in mice because of their simplicity, since they do not require additional *cis* or *trans* elements

and have high efficiency. They have been widely exploited as a tool for *in vivo* manipulation of DNA, usually for the removal of selection cassettes when targeting endogenous genes in ES cells or living animals. As a result of this there is now a growing bank of Cre-expressing transgenic mice with well-characterised limited expression patterns specific for tissues affected in myotonic dystrophy such as skeletal muscle: *myosin light chain 1f* (Bothe *et al.*, 2000), *actin* promoter(Miwa *et al.*, 2000), human *alpha skeletal actin* (Miniou *et al.*, 1999); testes (Bunting *et al.*, 1999) and eye (Utomo *et al.*, 1999). An interferon responsive Cre-expressing mouse is also available, which allows the ubiquitous expression of the activated transgene upon administration of the drug by injection (Kuhn *et al.*, 1995).

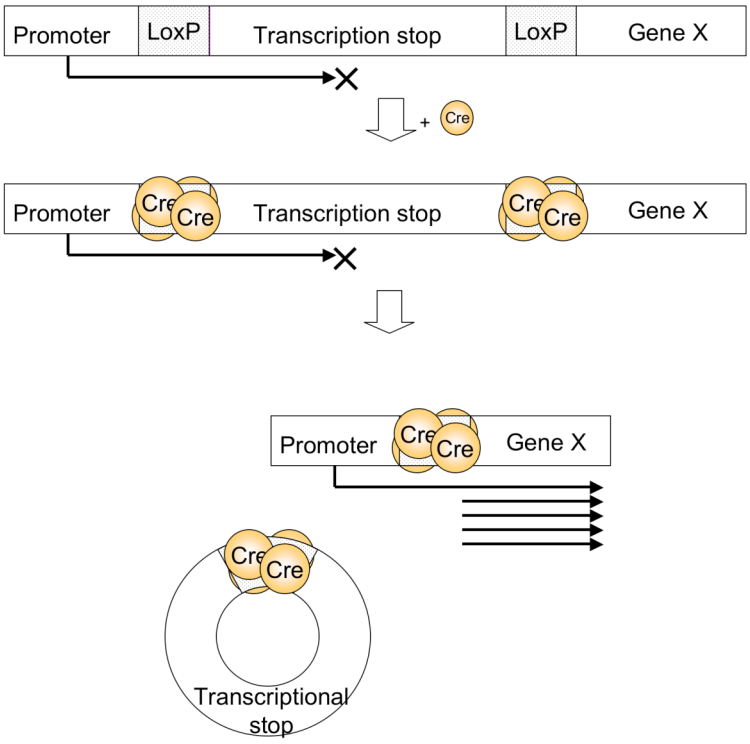


Figure 4 Cre mediated transgene activation. A simplified diagram illustrating the activation of transcription. Expression of genes 3' to the transcriptional stop signal is arrested. Cre binding results in excision of the regulatory element, allowing read-through of downstream sequences.

In nature, bacteriophage P1 exists within its host as a low copy number plasmid, and uses Cre-*lox* to maximise genomic segregation. Cre recombinase catalyses the conversion of multimeric prophage molecules into single unit copies via site-specific recombination between two *loxP* sequences (Abremski *et al.*, 1983; Hoess *et al.*, 1984). The DNA between two of these sites is circularised, resulting in single molecules. Here, we utilise this mechanism as a means of transcriptional control. By the placement of a *loxP*-flanked polyadenylation signal between the promoter and the open reading frame, transcription is inhibited (Figure 4). Expression of the transgene will occur only after Cre mediated

excision and as such is defined by its pattern of expression. This system will enable us to activate the production of CUG repeat messenger RNA in a developmental or tissue specific manner by crossing with existing Tg*Cre* mice.

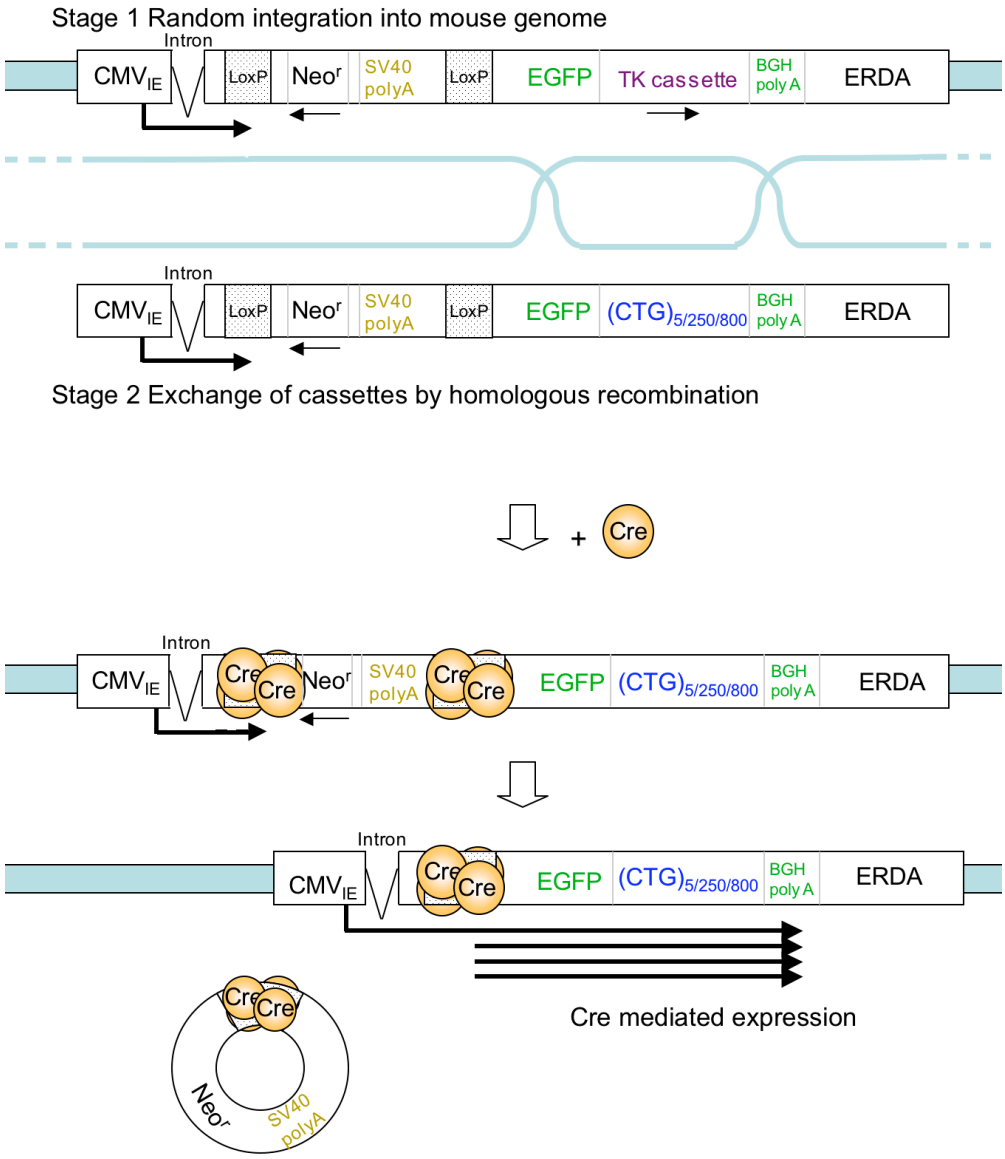


Figure 5 Primary, targeting and activated transgenes Schematic representation of the targeting strategy. Stage 1: The primary transgene pStopEGFP•TK is randomly integrated into the mouse genome using ES cells. Stage 2: Targeting constructs pStopEGFP5; pStopEGFP250 and pStopEGFP800 derived from the primary construct have repeat cassettes in place of the *TK* cassette. Using homologous recombination, these constructs direct 5, 250 or 800 repeats to the 3'UTR of the *EGFP* gene, at identical positions within the genome. *EGFP* and repeat expression is defined by the Cre recombinase expression pattern. Constructs and transgenes post Cre mediated excision contain one *lox*P site and are referred to as pLoxEGFP (0 repeats); pLoxEGFP5 (5 repeats) and pLoxEGFP250 (250 repeats).

In order to obtain transgenic mice that are free from positional effects differing only in the number of expanded repeats, we decided to use a two-stage process in ES cells. In the first stage, the primary *TK* construct would be randomly integrated into the genome and several clones selected based on Cre activated *EGFP* expression. These cell-lines would then be

used as a target for the integration of 5, 250 and 800 CTG repeats by homologous recombination (Figure 5). Once selected, cells from the clones would then be injected into blastocysts and transferred to the uterus of the foster mother. Chimeric offspring would then be set up for breeding and the F1 generation typed for germline transmission of the transgene.

3.2.2 Transgene structure

Selection of the promoter to drive *EGFP* and repeat expression was given much consideration. Although ideally we would have used the *DMPK* promoter, in this instance it was not possible since the full extent of the region was not known. Temporally, the chosen promoter needed to be active at least as early in development as the *DMPK* promoter, and in the tissues involved in this study. *DMPK* transcripts have been shown to be present in mice as early as day 10.5 *p.c.* (Jansen *et al.*, 1996), and can first be detected in whole embryos at mid-gestation (Unigene expression analysis for Mm6529, based on EST data). This is equivalent to 9-11 days p.c. since total gestation time is 19 days in the mouse. Baskar and co-workers analysed the developmental expression pattern of the CMV immediate early promoter using a lacZ transgene reporter, and first detected expression as a dorsal stripe in the neural folds of mouse embryos at day 8.5 *p.c.* A broader pattern was exhibited at day 9.5 *p.c.* and included the somites, from which muscle is derived (Baskar *et al.*, 1996). The CMV promoter is generally considered to be active over a broad range of cell-types, capable of producing high levels of transcript *in vitro* and *in vivo*, and has been used successfully in other transgenic mouse models (Hallauer *et al.*, 2000, skeletal muscle; van den Pol *et al.*, 1998, neurons). The expression profile and well-characterised aspect of the promoter made it a solid choice, although the level of expression is higher than that of the *DMPK* promoter, which could potentially pose problems since high levels of expression can stress the cell. This higher level may however result in earlier onset of disease, which could be beneficial when considering the lifespan of the mouse. When constructing living models, expression levels are difficult to control because of the strong influence of genomic context. The integration site can effect the levels of gene expression and even silence them in some cases (Mankodi *et al.*, 2000). If the transgene integrates into an essential gene, morbidity or lethality can also occur. In this strategy, these positional effects can be selected for after integration since clones can be selected for expression levels in tissue culture by pCre activation before blastocyst injection, and for viability by generation of the mouse itself.

The pCIneo expression vector (Promega) was used as the backbone since it contained the chosen promoter and other elements required for expression in mammals, namely an intron based on that of chicken β-globin and the BGH polyadenylation signal.

The *loxP* sequence is 34 bp long and arranged as an inverted repeat of 13 bp separated by an 8bp spacer region. This region contains two ATG codons, which could be utilised as initiation codons (Figure 6) resulting in missense translation of the *EGFP* reporter. To avoid this, both *loxP* sites were incorporated in reverse orientation. If two *loxP* sequences are present in *trans*, inverted with respect to each other, the intervening sequence is flipped to the opposite direction by Cre, so in order to give directionality during cloning, the sequences themselves were incorporated into the *neomycin* and *EGFP* forward PCR primers.

To free the model of variations due to position rather than repeat length, the secondary repeat constructs needed to be targeted to the randomly integrated primary *TK* construct (Figure 5). Homologous recombination is a rare event, the frequency lying around 1 in 10^{-5} to 1 in 10^{-9} integrants. Higher efficiencies are generally associated with longer flanking homologous sequences (Pinkert, 1994), with approximately 5000 bp of flanking homologous sequence optimal (Thomas *et al.*, 1987). In our transgene there is 4465 bp 5' homologous flanking region from the CMV promoter to the end of the *EGFP* ORF, but only 411 bp 3' provided by the BGH PolyA, so in order to make the transgene sufficiently large for second stage targeting we included a 2600 bp stretch of the human *ERDA1* region 3' to the non-homologous repeat section. The human *ERDA1* region (17q21.3) is associated with CTG repeats polymorphic within the population which account for most of long CTG repeats detected by the repeat expansion detection (RED) method in the genome (Sidransky *et al.*, 1998). Even though extensively researched (Schraen-Maschke *et al.*, 1999; Bowen *et al.*, 2000; Meira-Lima *et al.*, 2001; Mendlewicz *et al.*, 2004), *ERDA1* has not been linked to any disorder and is considered genetically inert. It has no homology to the mouse genome. In these constructs, part of the non-repeat *ERDA1* DNA is used as random sequence to extend the *TK*/repeat flanking region of the construct, to facilitate second round targeting.

Figure 6 The sequence of *loxP*. The inverted repeats are highlighted with arrows and possible initiation codons are in bold.

Both the *neomycin* and *thymidine kinase* genes need to be expressed in ES cells. *Neomycin* to enable positive selection for stable first-round integrants using the primary construct, and *thymidine kinase* (*TK)* to enable negative selection for second-round homologous recombinants, and so are placed under the control of the *PGK* promoter, which is widely used for ES cell targeting (Tybulewicz *et al.*, 1991: Yu, 2000 #612; Rijli *et al.*, 1994; Seidl *et al.*, 1998).

3.3 Transgene assembly

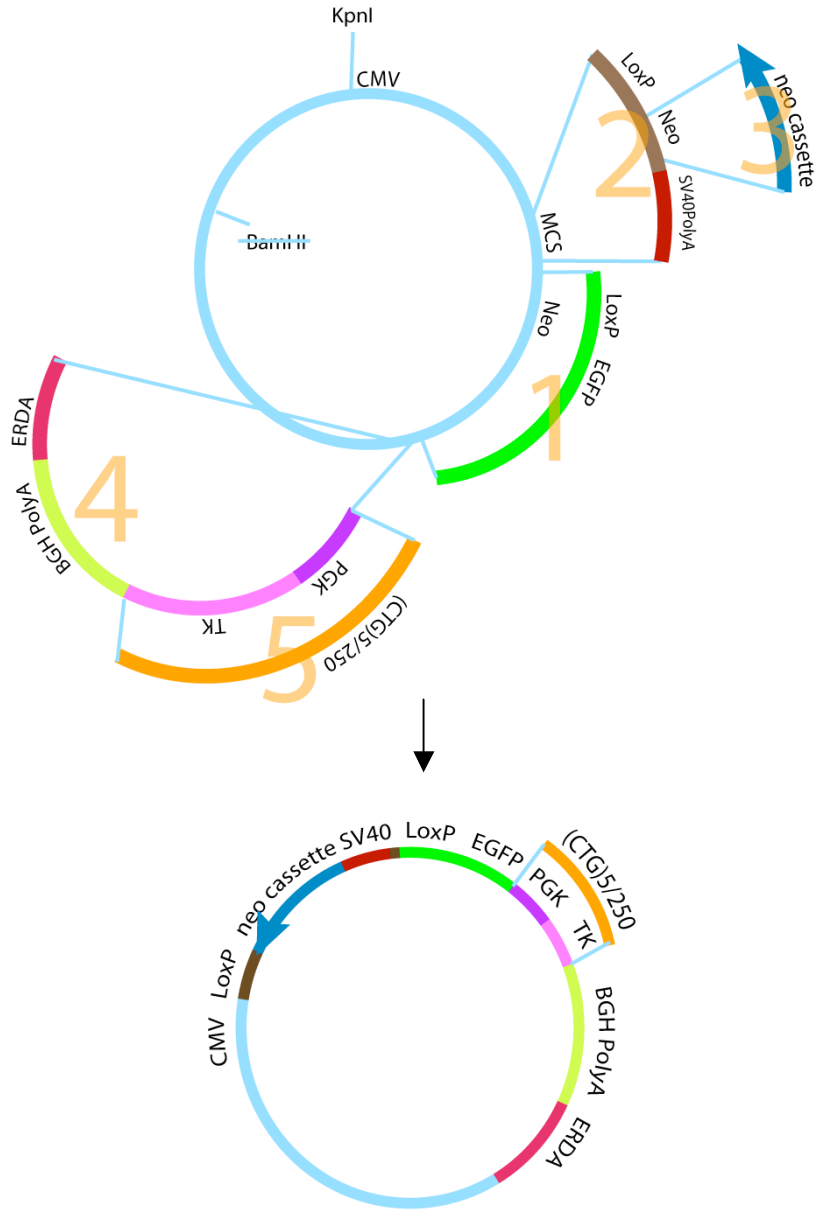


Figure 7 Cloning strategy. Diagram depicting the assembly of the component parts. Numbers refer to the order of ligation into pCIneo. Fragments 2 and 4 were first assembled in pBluescript SKII+. Partial replacement of fragment 4 with fragment 5 converts the founder to the targeting construct. Primers, restriction fragments and templates are detailed in materials and methods. Not to scale.

The constructs were pieced together from components obtained by restriction digest, or PCR amplification, with each primer designed to incorporate a unique restriction site. A *loxP* site was incorporated into the forward amplimers 5' to the Kozak sequence and the initiation codon of both the *EGFP* and *neo* ORF. PCR products were cloned into pTEasy and verified by sequencing. A cloning vector was used to assemble fragments before cloning into a modified pCIneo vector. These modifications comprised replacement of the vector (pCIneo) *neomycin* resistance with *loxP-EGFP* ORF, removal of the *Bam*HI restriction site and the addition of a unique *Kpn*I site upstream of the CMV promoter. At first, *neomycin* resistance (required to select for recombinant ES clones) was designed to be transcribed from the CMV promoter and terminate by the SV40 PolyA, but concerns arose about the function of the CMV immediate early promoter in undifferentiated cells after reports that the CMV promoter was inactive in ES cells (Chung *et al.*, 2002). To counteract this, the PGK-*neo*-PolyA cassette from the targeting vector pPNT was cloned into the *neo* ORF in the reverse direction. Primer pairs and restriction fragments used for generation are detailed in chapter 2 "Materials and Methods". Finished constructs were verified by restriction digest.

3.4 Component function

3.4.1 CMV promoter, EGFP and *loxP* interference

Once gene expression has been activated by Cre excision, a single *loxP* site remains within the transcribed region. This sequence is 34 base pairs long: A 13bp inverted repeat separated by an 8bp spacer region. It was possible that this sequence could form a hairpin structure within the transcribed RNA, inhibiting translation, resulting in low levels of protein production (Figure 8A). The Kozak consensus sequence for ribosome binding lies 2bp 3' of the possible hairpin and so may not interfere with binding, allowing translation. To test the sufficiency of protein expression from the CMV promoter with the *loxP* site present, the *loxP-EGFP* ORF fragment was cloned into the MCS of the pCIneo vector (to generate pLoxEGFP), and transfected into mouse 3T3 cells. A high proportion of these cells glowed green when visualised by fluorescence microscopy (Figure 8B). No fluorescence was seen in cells transfected with reagent alone. It is clear then that the *EGFP* fragment used for the founder and targeting constructs is functional and that expression from the CMV promoter is strong enough to be visualised even when the *loxP* sequence is present within the transcript. It is impossible to know in a transient transfection if the

EGFP is transcribed from a single copy gene as would be the situation created in the mouse, or if multiple copies are present within the cell and are in fact required in order to accumulate sufficient levels of the fluorescent protein. Therefore, we transfected pCDC2-EGFP, also a strong promoter, which was available in our laboratory and pLoxEGFP in parallel to compare EGFP levels. Because the fluorescence was at least equivalent when judged by eye (figure 5C), and the *loxP* site is unavoidable in this strategy, we did not see it necessary to generate a further control construct with the CMV promoter and *EGFP*, without the *loxP* site. Since the CMV promoter is strong and frequently chosen for expression vectors carrying EGFP as a reporter (*e.g.* pIRES-EGFP, Clontech; pcDNA 6.2 DEST series, Invitrogen) high levels of transcript should be produced.

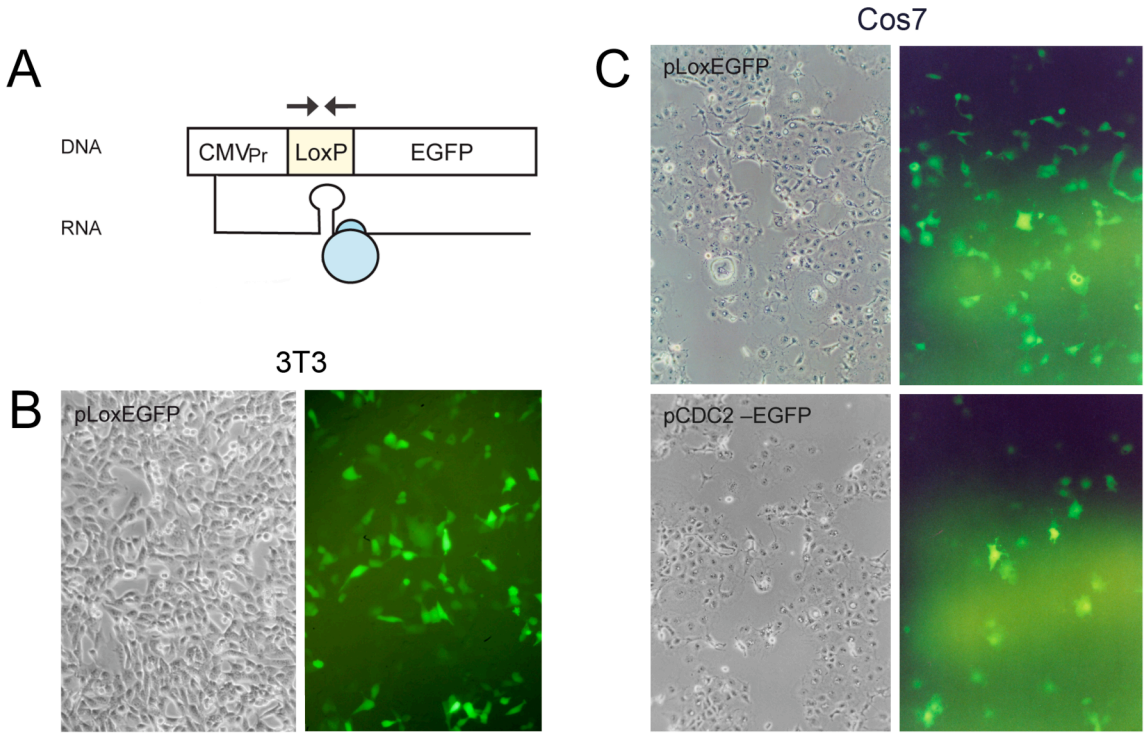


Figure 8 EGFP expression in the presence of the *loxP* site. A Schematic diagram of pCI lox EGFP Expression showing possible hairpin formation. The ribosome is shown in blue at the Kozak sequence. **B** Fluorescent micrograph showing EGFP expression in mouse 3T3 cells. Photographed under UV excitation and brightfield. Magnification X10. C Fluorescent micrograph showing a comparison of EGFP expression from pLoxEGFP and pCDC2-EGFP plasmids in Cos7 cells (African green monkey).

3.4.2 Neomycin

The *neomycin* gene from Tn5 encodes an aminoglycoside 3'-phosphotransferase, which confers resistance to G418, an aminoglycoside antibiotic. G418 blocks polypeptide synthesis by inhibiting the elongation step in both prokaryotic and eukaryotic cells, and so cells expressing the *neomycin* gene survive whilst those without it die. This would allow us to select ES cell clones transfected with the construct by adding G418 to the medium.

Stable cell-lines were generated to test *neomycin* function. The lowest concentration of G418 resulting in cell death after ten days treatment was determined for the cell-types *Dmt*D2976 Kidney pg 21; Cos7 and HeLa cells. The targeting transgenes pStopEGFP containing 5 and 250 repeats were then linearised and isolated on a *Kpn*I-*Xba*I restriction fragment for *Dmt*D2976 Kidney and Cos7 cells and an *Acl*I fragment for the second set of Cos7 and HeLa cells (Figure 20), transfected into the cells, and selected for until colonies appeared. Colonies from each were picked and propagated. Untransfected cells from the same lines died in the same time period; therefore the *neomycin* gene is functional conferring G418 resistance. When analysed by PCR all clones with amplifiable DNA present were positive for the construct (Figure 9).

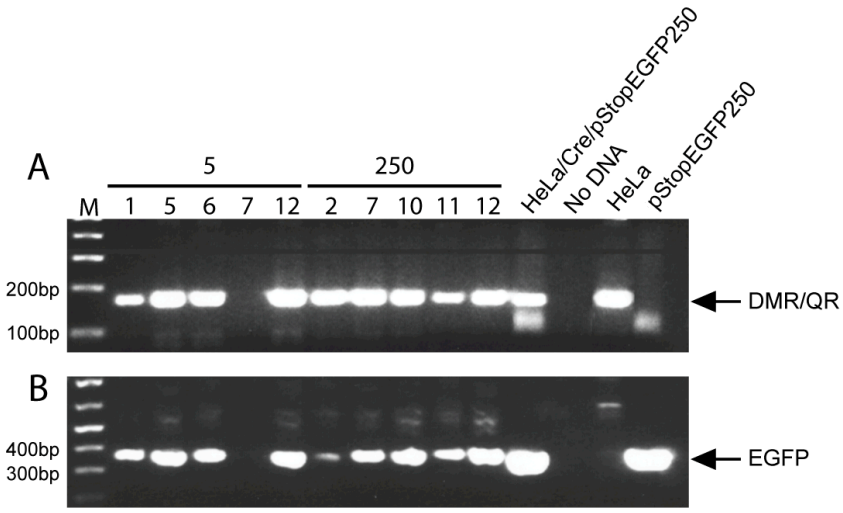


Figure 9 A representative DNA analysis of stable cell lines Representative HeLa clones for 5 repeats (1, 5, 6, 7, 12) and 250 repeats (2, 7, 10, 11, 12); co-transfected HeLa control (pStopEGFP250 and pCre); no DNA control; untransfected HeLa control and pStopEGFP250 control. **A** PCR using DMR and QR amplimers specific for the human *DMPK* 3' UTR and not present in the transgene. This confirms the presence of DNA in the sample. **B** PCR using primers within the *EGFP* ORF confirming the presence of the construct.

3.4.3 Thymidine kinase

Viral *thymidine kinase* (*TK*), the product of the CMV *UL97* gene, specifically phosphorylates gancyclovir, a nucleoside analogue, precipitating chain termination and causing cell death once incorporated into the DNA during replication. It would be used as negative selection for cells that have lost the gene through homologous recombination, required for targeting the repeat constructs to the primary gene construct.

The mouse cell lines described in 3.4.2 harbouring the founder construct, and control cells without were treated with 560ng/ml Gancyclovir for 5 days. The test cell lines died whilst control cells grew to confluence indicating that the *TK* gene is functional.

3.4.4 Transcriptional stop, *loxP* and Cre excision

It is important to prevent transcription of the expanded repeats in the uninduced state. To facilitate this the SV40 PolyA signal, a widely used strong transcriptional terminator, has been positioned between the *loxP* sites. To further hinder any possible read-through transcription prior to Cre activation, the *neomycin* selection cassette has been positioned 3'-5' relative to the direction of transcription from the CMV promoter. To test the mechanism of activation and repression, murine 3T3 cells were transfected with the pStopEGFP5 targeting construct with and without Cre-expressing plasmid or pCre alone (Figure 10 top three rows).

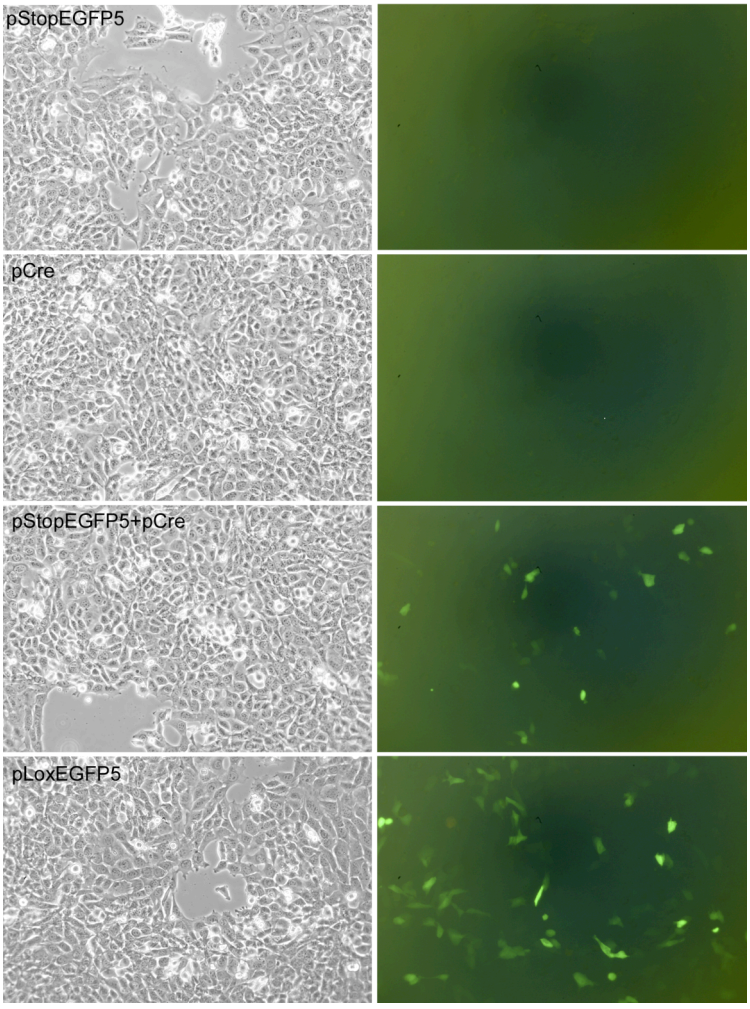


Figure 10 Cre activation of EGFP fluorescence. Brightfield and fluorescent micrographs showing EGFP fluorescence activated by Cre excision in 3T3 cells. EGFP fluorescence was only detected when cells were transfected with both pStopEGFP5 and pCre, or with the constitutively-expressing pLoxEGFP5. There were fewer EGFP fluorescent cells with pStopEGFP5+pCre compared to pLoxEGFP since the amount of transcriptional target (pStopEGFP5) was reduced by half in the co-transfection, the other half made up by pCre.

Fluorescence could only be detected in those cells transfected with both pStopEGFP5 and pCre plasmids indicating that Cre recombinase was required for activation of fluorescence,

and that fluorescence was repressed by the transcriptional control elements. The mechanism was also tested by using recombinant Cre *in vitro*. PStopEGFP5 and pStopEGFP250 plasmids were treated with recombinant Cre, and the resulting mixture transformed into JM109 cells. Cre-excised plasmids were isolated (Figure 11) and the pLoxEGFP5 derivative used to transfect 3T3 cells (Figure 10 bottom). The proportion of fluorescent cells was much higher in the pLoxEGFP5 sample compared to the co-transfected samples because the total amount of DNA in each transfection was constant, and two plasmids were required for fluorescence in the co-transfection. Secondly, the starting amount of pStopEGFP5 plasmid was reduced by half, the other half made up by pCre, so fluorescence was further reduced. The same result has been obtained using *Dmt*D Kidney; Cos7 and HeLa cells (data not shown).

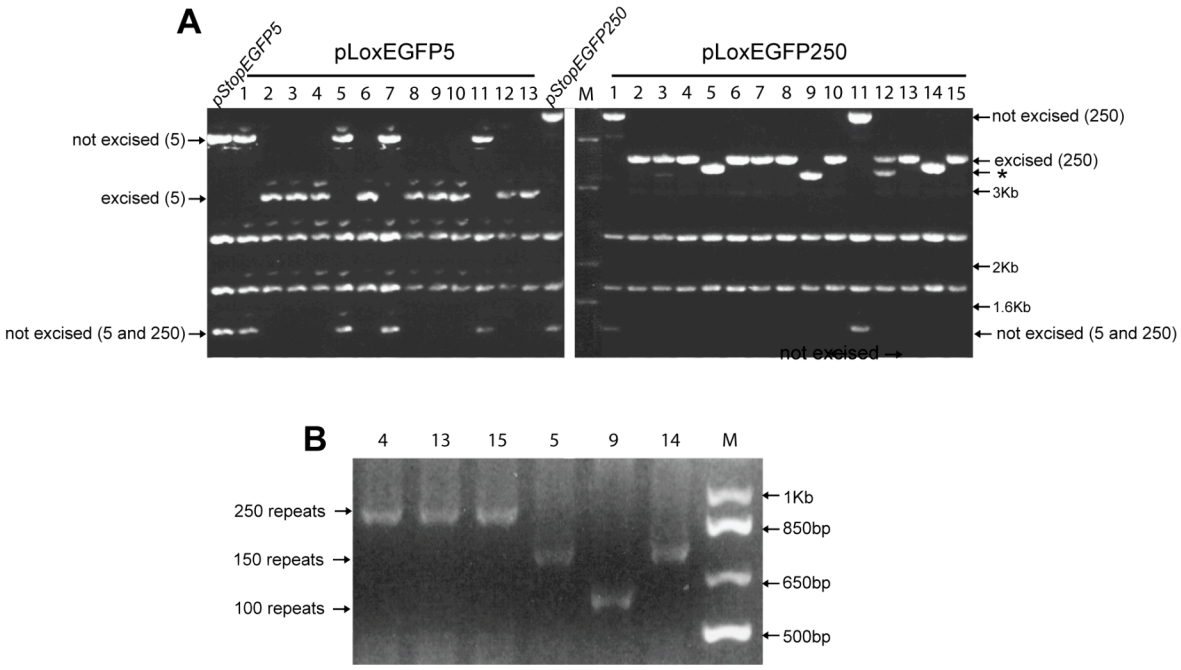


Figure 11 Analysis of lox constructs after *in vitro* treatment with recombinant Cre. A *Bgl*II restriction analysis of transformants. pStopEGFP5 and pStopEGFP250 are untreated parental controls. The positions of fragments specific to excised or untreated constructs are indicated, unlabeled fragments are common to all. The asterix denotes plasmids identified with putative spontaneous deletions in the number of repeats. **B** DMH-BR PCR over the repeat region. Sample numbers 5, 9, and 14 are spontaneous deletions containing approximately 150, 100 and 150 repeats respectively. Numbers correspond to 250 repeat clones in A.

3.4.5 The repeats

The CTG repeat has been shown to form a single hairpin *in vitro* (Michalowski *et al.*, 1999) which was thought may interfere with *EGFP* transcription if placed directly 3' to the termination codon. To avert this, a portion of the *DMPK* 3' UTR (113bp 5' and 112bp 3'of

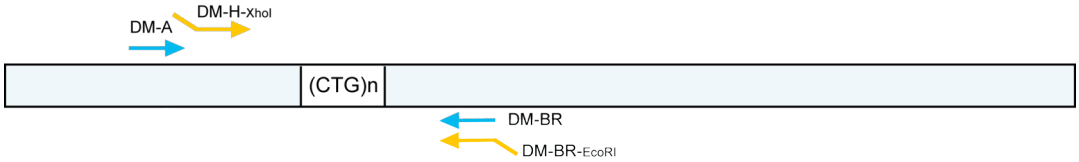


Figure 12 Human *DMPK* 3' UTR showing the position of nested primers used to amplify repeat sequences. A schematic diagram depicting the relative primer positions for semi-nested PCR. First round primers are blue. Second round primers are orange and include restriction site 5' extensions for cloning.

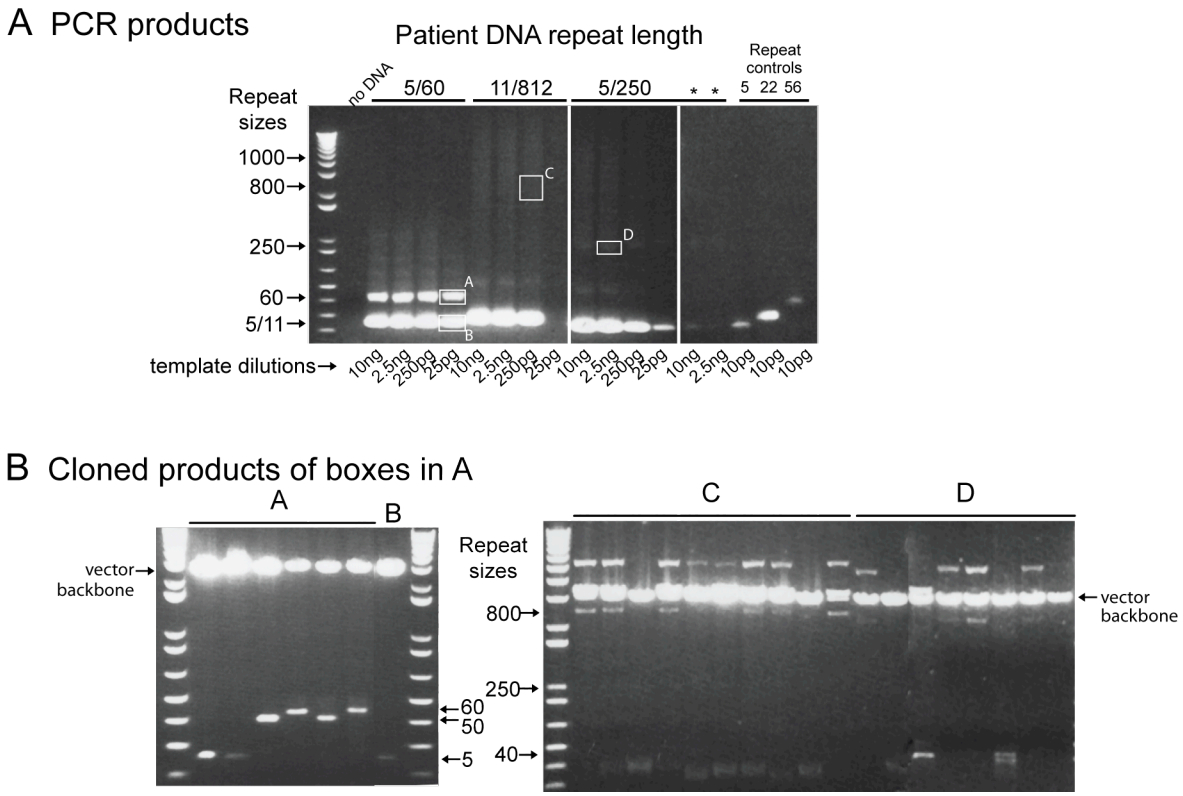


Figure 13 Cloning the repeats. A Products of nested PCR used for cloning. Bands were excised as indicated by white boxes. Templates consisted of *Hin*dIII digested patient genomic DNA. PCR was carried out for 28 cycles or 7 cycles (*) and with varying dilutions of DNA. Products obtained using the highest dilution and lowest number of cycles was used for cloning, in order to minimise aberrant nucleotide incorporation during PCR. **B** Repeats after cloning. Note that there is slight variation in the length of cloned "60" repeats in sample A. The smaller 5 repeat band may be due to contaminating 5 repeat PCR product within the agarose. Unexpected band sizes were obtained for the 800 and 250 repeat samples C and D. 800 repeats would be 2620bp and 250 repeats 970bp. The vector backbone is 3Kb which indicates that many of the clones in sample C and D are probably undigested DNA. The largest insert contains no more than 40 repeats.

the repeats: Accession # L19268), was included to act as a spacer between the end of the *EGFP* reading frame and the polyadenylation sequence. This region used includes the subsequently discovered splicing factor binding sites for hnRNP C; PTB; U2AF and PSF (Tiscornia *et al.*, 2000). *DMPK* 3' UTR regions containing CTG repeats were isolated from patient DNA samples previously characterised for the length of repeat (Hogg *et. al.*

unpublished results) by semi-nested PCR (Figure 12). The mutant alleles in these patients contained distributions around 60; 250 and 812 repeats, with normal alleles of 5; 11 and 5 respectively. These samples would act as a template to amplify 5; 250 and 800 repeat arrays to model unaffected, and the adult onset and congenital forms of myotonic dystrophy type 1. The 60 repeat was also cloned for possible use at a later date. Once amplified the fragments were gel purified based on size as bands, or regions in the case of the 800 repeat since smears were obtained due to the mosaicism of repeat length in somatic tissues (Figure 13A). The DNA was isolated from the agarose and ligated into pTeasy, a PCR-product cloning vector. Using blue-white screening in JM109, transformants for the 5 repeat and 60 repeat ligations were obtained, but not for 250 or 800 repeat ligations. Insert sizes were confirmed by restriction digestion and found to be correct for the 5 repeat, and varied between 55 and 60 repeats for the 60 (Figure 13B samples A and B). UTRs were verified by sequencing (MBSU, Glasgow). The transformations for 250 and 800 repeat lengths were repeated increasing the amount of ligation reaction. Colonies were analysed but did not digest completely with *Eco*RI, some insert bands were present, but not of the expected size ranging from mostly 5 repeats from the 800, to no more than 40 repeats for the 250 repeat insert (Figure 13B samples C and D).

PCR amplification and cloning was repeated with more product obtained using 10ng *Hin*dIII DNA, *Xho*I-DM-H/*EcoR*I-DM-BR PCR, and a larger PCR volume. Bands were excised, ligated and cloned and as before. Unexpected results were obtained once again. From the results it looked as if some of the samples were not fully restricted since there were very large bands present. *Eco*RI digestion should release inserts of 1kb for 250 repeats and 2.6Kb for the 800 repeat leaving a 3Kb backbone, yet no restriction patterns showing 3Kb and 1Kb or 3KB and 2.6Kb were observed. Most samples (Figure 13B samples C and D, and Fifure 14A top) contained multiple fragments of varying intensities often seen with undigested plasmid samples. For clarification, the samples were Southern blotted and hybridised with a DM56 repeat probe to check the position of the repeat fragments on the gel (Figure 14A bottom). This probe was in general use in the laboratory for the detection of repeat alleles by Southern blot and consists of a CTG repeat PCR product amplified with oligonucleotide primers DM-C and DM-ER from the DM1 3'UTR (Figure 20). The repeat positive clones were digested with a fresh aliquot of *Eco*RI enzyme, and electrophoresed alongside undigested and *Xho*I restricted samples. *Xho*I digestion was included for the samples for clarification of plasmid size since only one *Xho*I site should be present in the clone, from the 3' end of the insert. If multiple inserts had been cloned, fragments would be released by *Xho*I. After restriction, only one band was

present, so this was not the case. Insert fragments released by *Eco*RI were smeary, indicating a range of repeat lengths. These were judged to be smaller than 250 repeats in all but two samples (Figure 14B arrows). From the Southern and the second digest, samples 5 and 8 (originally from the 800 repeat pool) contained the largest inserts at around the correct size for a 250 repeat (Figure 14A top and B arrows). These were streaked out to single colonies from the same culture used to isolate plasmid DNA for the restriction analysis. Plasmid DNA was isolated and digested with *Eco*RI and the products electrophoresed beside undigested sample or *Xho*I digested sample (Figure 14C). There was a large amount of insert variation present, but the average size of the bands was equivalent to the most predominant bands in the parent clones, 5 and 8. The asterisked, clone 8 derivative was selected for cloning into the final targeting construct. Verification of the insert by sequencing was attempted, but it proved impossible to sequence across the full stretch of repeats so the expansion may be imperfect, however the first 50 repeats at the 5' end (where sequence information was good) are pure.

3.4.5.1 Further attempts to clone 800 repeats

To try a different approach to clone an 800 repeat array, small-pool PCR was used to amplify single molecules from the patient DNA. One of the amplified aliquots contained a 750 repeat and also a 5 repeat allele (Figure 15A). The 750 repeat allele fell into the range for congenital DM1 classification, and so attempts were made to clone it in place of 800 repeats. A second round of PCR was performed using the aliquot with semi-nested primers as previously described (Figure 12; 25 cycles). The product was gel isolated using low melting temperature agarose, ligated directly and transformed into Top10f' with blue-white selection. Transformants were restricted with *EcoR*I and found to have massively variable insert sizes (Figure 15B). Clone 10 was selected for further purification, as the top insert band was the correct size for 800 repeats, and present at a high level compared to the smaller insert fragments. It was hoped this plasmid could be made homogeneous for a large insert. The clone was streaked out to single colonies as previously described. *EcoR*I restriction analysis revealed that the insert ranges had again reduced in size (Figure 15C). This clone was abandoned and clones 3, 12 and 24 were analysed by restriction digest to check the restriction sites (Figure 15D). *Eco*RI should release the insert because of flanking sites within the MCS. *Pvu*I should not cut within the insert so a duplex digest with *Eco*RI and *Pvu*I should leave a fragment of the same size as that when *Eco*RI is used alone, yet both larger fragments have reduced in size and only smaller band sizes remain constant. *Xho*I does not cut within the vector, and so should result in a single band of 5665bp for an insert of 800 repeats and the vector backbone. Yet digestion with *Xho*I gave

a doublet band. If two inserts had been cloned then one would see a doublet band with *Xho*I, and also if two plasmids were present with different array lengths -which could happen if repeats result in deletions during replication or mismatch repair.

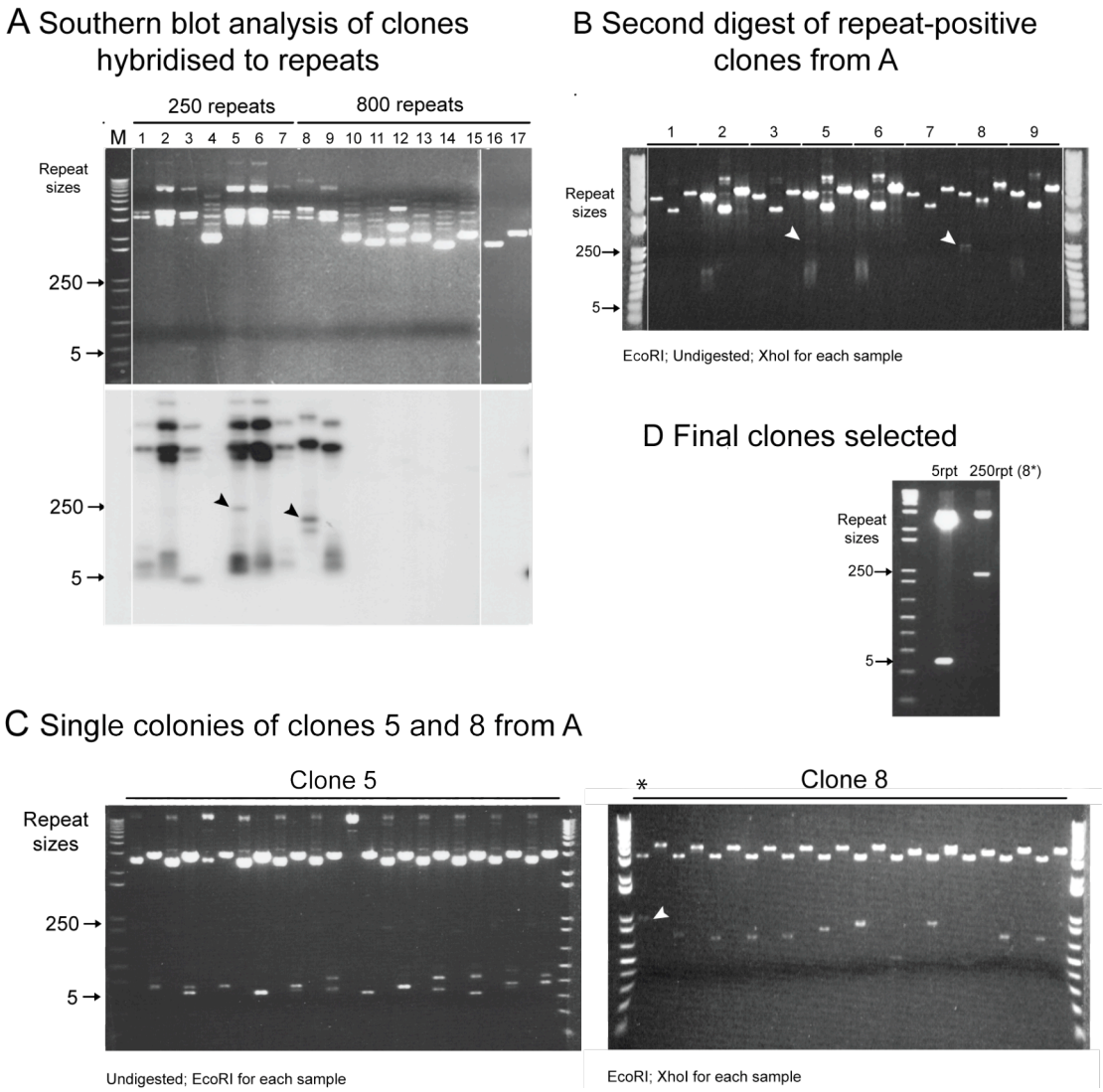


Figure 14 Unexpected restriction pattern of repeat clones A Top *Eco*RI restriction did not release inserts. No bands corresponding to repeat arrays were observed. **A Bottom** Southern blot of the same gel confirms presence of repeats. Arrows indicate inserts of the correct size for 200-250 repeats. **B** Clones found to be positive for the repeats by Southern blot in A were re-digested with fresh *Eco*RI and electrophoresed beside undigested samples to confirm enzyme activity; *Xho*I linearisation was included to confirm plasmid size and rule out rearrangements. Arrows indicate inserts of the correct size for 250 repeats. Note that insert bands appeared smeary possibly indicating insert heterogeneity within the clones. **C** Clones 5 and 8 (identified as 250 repeat clones in B) were streaked out to single colonies, DNA was isolated and re-analysed for insert size. Note the variation in insert size between the clones. The range of variation is similar to the predominant fragment size in the parental clone. The asterix denotes the clone selected for construct generation. The arrow indicates insert size corresponding to 250 repeats. **D** Clones selected for the final construct. *Eco*RI-digested. 8* corresponds to the 250-repeat clone in C. The 5 repeat clone was selected previously (Figure 13B left hand clone of A).

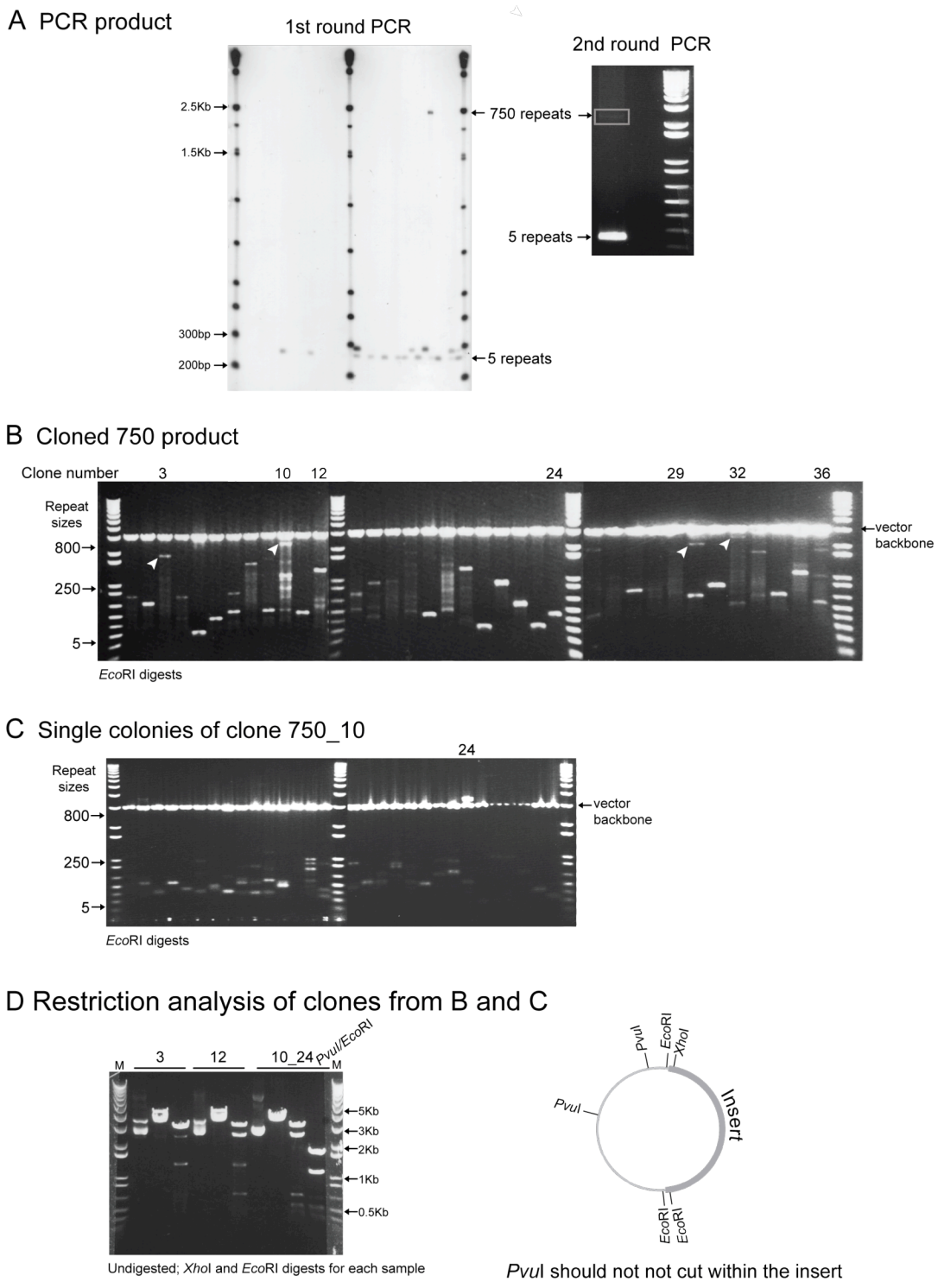


Figure 15 Cloning 750 repeats. A A 750 repeat band was identified from SP-PCR of patient DNA and used as a template for the second round PCR (Figure 12) unfortunately this sample also contained the normal 5 repeat allele which preferentially amplified. **B** Massive variability between repeat-containing inserts in the cloned product. Some clones contain large repeat inserts (arrows). Most have a mixture of large and small fragments. Clone 10 was selected for further analysis because the large band was proportionally high in the mix compared to other clones. **C** The overnight culture of clone 10 used for DNA analysis in B, was streaked out to single colonies. DNA was isolated and re-analysed with *Eco*RI. The large band disappeared, with no insert bands above 250 repeats observed. **D** Restriction analysis of selected clones with larger inserts. The sizes for a 750 repeat band are: *Eco*RI: 2653 bp + 3000 bp; *Xho*I: 5665 bp; *Eco*RI/*Pvu*I: 2653 bp + 1710 bp +1096 bp + 191 bp + 16 bp. The *Xho*I doublet probably originates from different sized inserts.

Attempts so far involved the analysis of many clones since it looked as if the repeats were reducing and expanding during the cloning and bacterial transformation. Streaking out existing transformants and probing with repeats should reveal those clones with the largest inserts since more probe would bind. Clones 29; 32 and 36 (Figure 15B) were chosen based on insert size, streaked out to single colonies, picked and transferred to nylon membrane and hybridised to a DM56 repeat probe. There was an obvious variation in the strength of the hybridisation signal between clones (Figure 16A), those corresponding to the strongest were propagated and analysed by *EcoR*I restriction. Again there is a massive variation in the length of the repeat fragments, with large deletions apparent between sibling isolates (Figure 16B), and compared to the parental clones (from Figure 15B). We concluded that the 800 repeat was too unstable to clone.

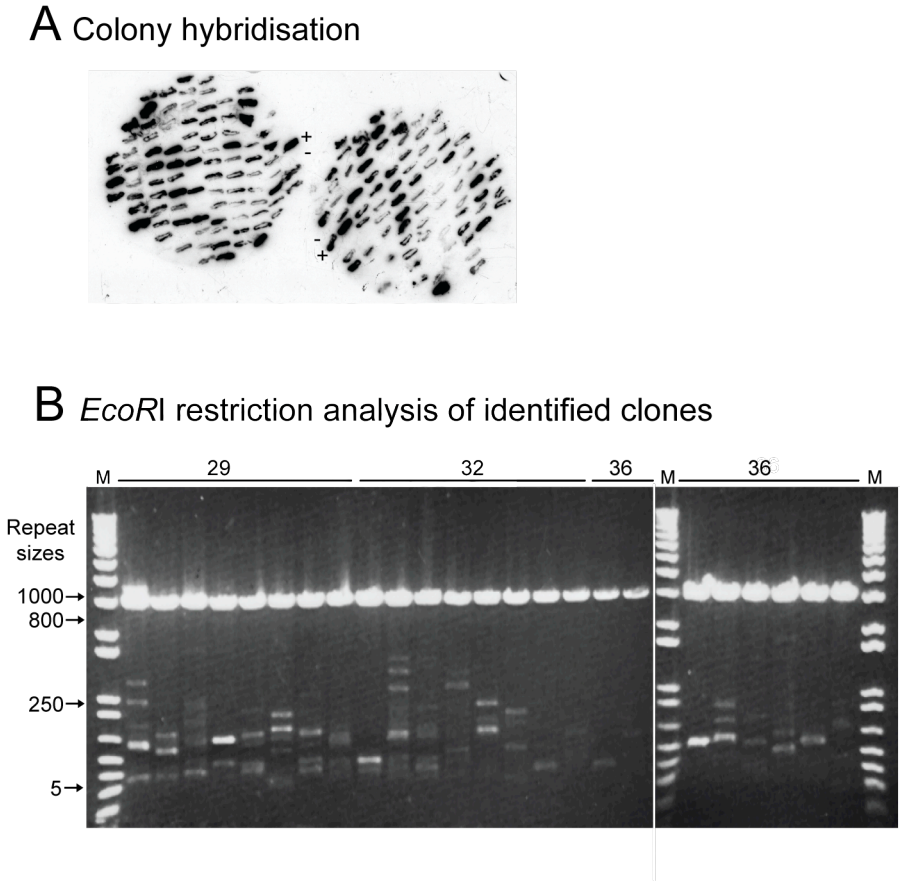


Figure 16 Identification and DNA analysis of repeat clones. A Single colonies from clones 29; 32 and 36 (Figure 15B) probed with DM 56 repeat probe. **+** (positive control) = pLoxEGFP250 in JM109; **-** (negative control) = pLoxEGFP in JM109. **B** *Eco*RI restriction analysis of DNA from the high signal colonies reveals high variability of insert sizes both within and between clones.

By this time Mankodi *et al.* had published evidence that 250 repeats expressed within the 3' UTR of an mRNA unrelated to DM1 was sufficient to induce DM-like symptoms in mice. Because of this information and time constraints the mouse model was limited to 5 and 250 repeats.

3.4.5.2 Propagation of repeat constructs

The hallmark of myotonic dystrophy pathogenesis is the formation of ribonuclear foci; aggregates of the mutant RNA trapped within the nucleus. Until the floxed 250 repeat construct was tested for foci formation in mammalian cells (which is addressed in chapter four), we did not realise that there was any significant instability with expansions of 250 repeats whilst amplifying the DNA in bacteria. During testing, not all EGFP positive cells harboured CUG ribonuclear foci, which made us wonder if there were smaller length repeat inserts within the DNA preparation used. To test this, DNA was linearised and serially diluted to single molecules, then subjected to PCR across the repeat tract. Surprisingly, the preferentially amplified size within this particular preparation corresponded to 27 repeats, not 250 (Figure 17A left). Restriction analysis of the insert fragment revealed a ladder of repeat lengths ranging from 250 to 5 repeats (Figure 17 A right). The original overnight culture was then streaked out to single colonies and analysed for insert size. The clone with the most homogeneous population of repeat tract was selected (Figure 17 B), streaked out and reanalysed. Clone 250a* was selected as the master stock for future floxed plasmid preparations (Figure 17 C).

Throughout the time of this thesis three large-scale preparations each of the original and the floxed 250 repeat targeting constructs have been made. Each time it was necessary to reselect the correct repeat length. For each batch, the construct was re-transformed into JM109 without blue-white selection and a single colony used for inoculation of each of ten 100ml overnight cultures. A 1.5ml sample was taken from the culture and analysed for insert size whilst the remaining culture was pelleted by centrifugation and stored at -20^{o}C. Clones were selected based on the size and homogeneity of repeat length (Figure 18), and the corresponding pellets prepped. Different host cells have been used to culture the plasmid, DH5α, Top10F' and JM109. The latter two contain *lacI*^{q}, which affords constitutive expression of the repressor giving tighter control of gene expression. Whilst Top10F' and JM109 give much higher yields compared to DH5α, there seems to be no significant difference in homogeneity of repeat length (data not shown).

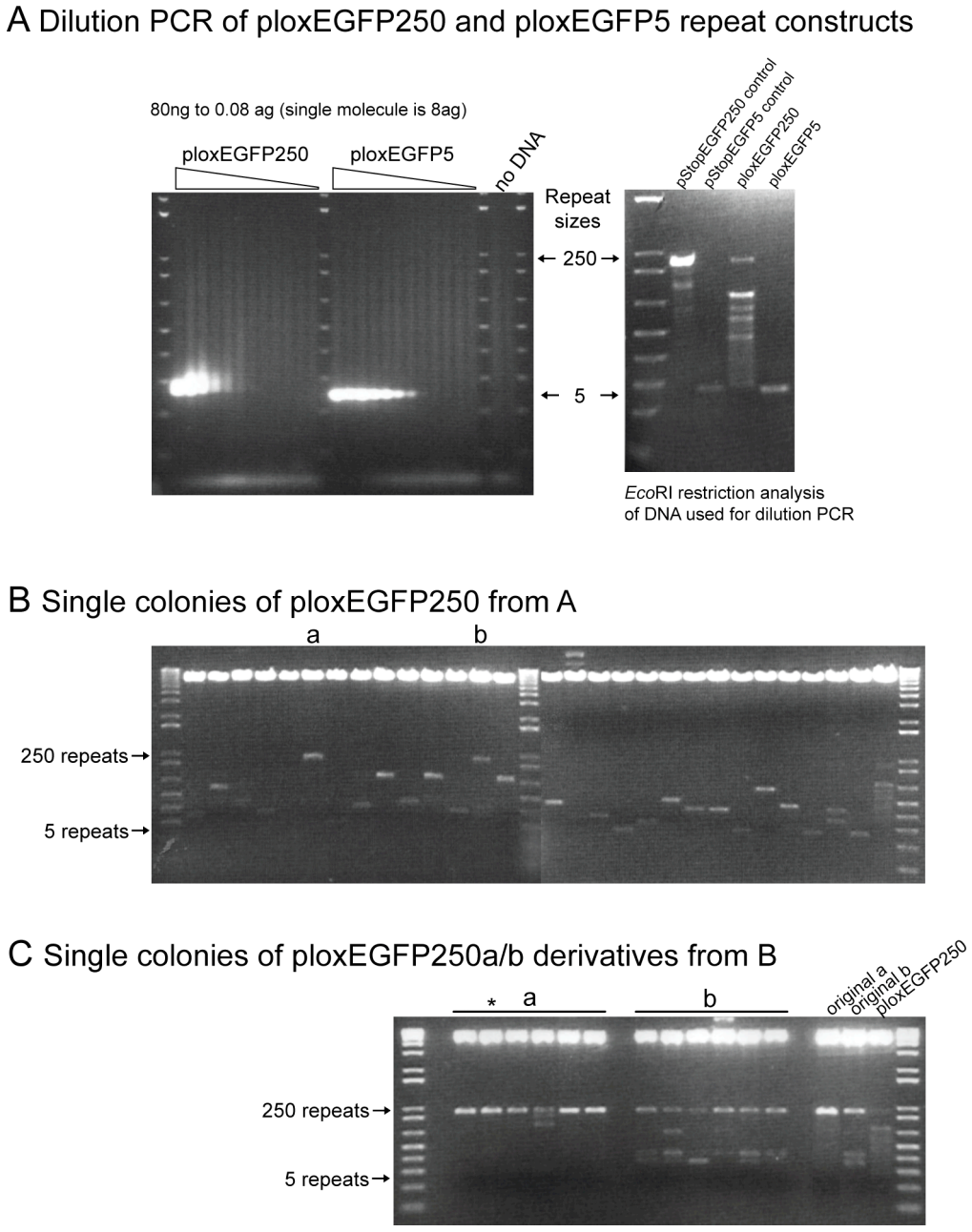


Figure 17 Repeat array contraction during plasmid propagation. A Dilution PCR and restriction analysis showing variation in repeat length after propagating pLoxEGFP250 from a glycerol stock. **B** *Eco*RI restriction analysis of subcultures of the original pLoxEGFP250 glycerol stock in A. The number of bands in each derivative has reduced, and the overall range of the band-sizes is similar to the original. **C** A second set of subcultures from **a** and **b** in B. The number of bands in each derivative has reduced again, the overall range of the band-sizes similar to the original.

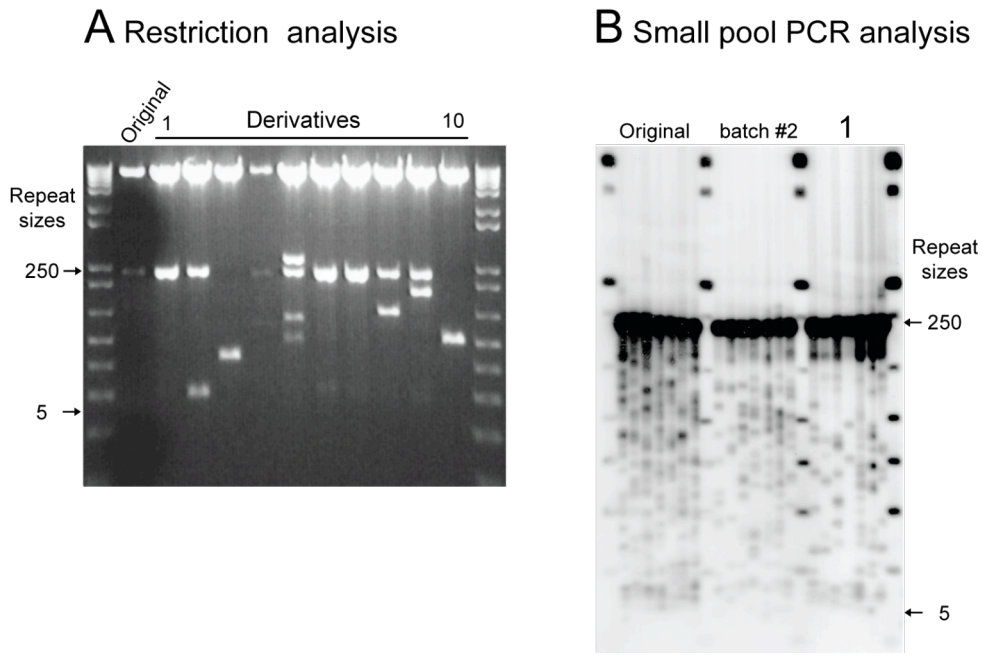


Figure 18 Heterogeneity of repeat length in floxed 250 preparations. A Original denotes master stock plasmid pLoxEGFP250a*(Figure 17C) used for transformation. 1-10 denotes subcultured single colony transformants, analysed by *Eco*RI digestion to release the cloned repeat array. Sample 1 was selected and is shown in SP-PCR analysis in B. **B** SP-PCR analysis of pLoxEGFP250 constructs used throughout this research. 500 molecule dilutions; 7 replicates. Samples: Original floxed 250; batch #2 and transformant 1 from A. Note the presence of deletion products ranging from 250 to approximately 5 repeats.

3.5 The mouse model

3.5.1 Targeted integration

Precise alteration of endogenous genes is a method usually used for null or knockout models. Here we attempted to use ES cells for random integration of the master construct followed by targeted integration to that construct, to produce a gain-of-function model where control and mutant mice differ only in the number of repeats, free of positional effects.

ES cells derived from preimplantation embryos retain totipotency which reduces as the number of passages increase (Nagy *et al.*, 1993). This would allow genetic manipulation by targeting in the early passages. We proposed that if the passage number was kept low, two rounds of manipulation are possible. The first, a random integration of the master construct followed by clone selection. The second, targeted integration of the 5 and 250 repeats. Once manipulated, ES cells from selected clones would be injected into

blastocysts, which would then be implanted into the uterus of the pseudopregnant foster mother to produce chimeras, hopefully with colonisation of the germ-line.

The ES cell-line CGR8.8 was a kind gift from Professor R.W. Davies (I.B.L.S., University of Glasgow), originally derived from mouse strain 129j/Olac (Nichols *et al.*, 1990). These cells had already been shown to give germ-line transmission in the Davies laboratory. Before proceeding with the integration, a kill-curve was done to select the optimum dose of G418 to use for selection of transgene-positive cells. It was important to use the lowest effective dose to ensure single copy integrants would be isolated, and that selection was not biased towards those cells expressing large amounts of *neomycin* due to the presence of multiple copies of the transgene. One hundred CGR8.8 embryonic stem cells were seeded into each well of a gelatinised 12 well plate. After 48 hours, 0; 100; 200; 300 and 400µg/ml G418 was added to the medium in duplicate wells. The G418 supplemented medium was replaced each day and the state of the cells noted. Cell death began in the higher concentrations after 48hrs, and cells were completely dead by day 6. Few cells remained at 200µg/ml G418. At day ten, no colonies could be seen at 200µg/ml G418, and so this concentration was chosen for the selection of transfectants.

To test promoter function in ES cells, pLoxEGFP was linearised and transfected into CGR8.8 embryonic stem cells using electroporation, and the media supplemented with the cytokine leukaemia inhibitory factor (Lif) to maintain the undifferentiated state of the cells (Nichols *et al.*, 1990). Unfortunately it was not possible to include a cell-type positive control such as 3T3 cells for electroporation due to contamination prevention guidelines of the ES cell facility. G418 was added after 24 hours to select for stable transfectants. Colonies appeared after 7 days and were viewed using fluorescent microscopy. Few cells exhibited EGFP fluorescence. This could have been due to a low rate of transfection, often experienced in primary cell-lines such as ES cells. ES cells require close cell-cell contact in order to maintain the undifferentiated state. The cells that did fluoresce were situated on the perimeter of the ES cell colony where the cells are most likely to be differentiating (Figure 19). Since each colony is derived from a single cell, EGFP fluorescence should be apparent throughout the colony. This was not the case so it is reasonable to conclude that the promoter is not functioning in the undifferentiated state, as has been reported (Chung *et al.*, 2002).

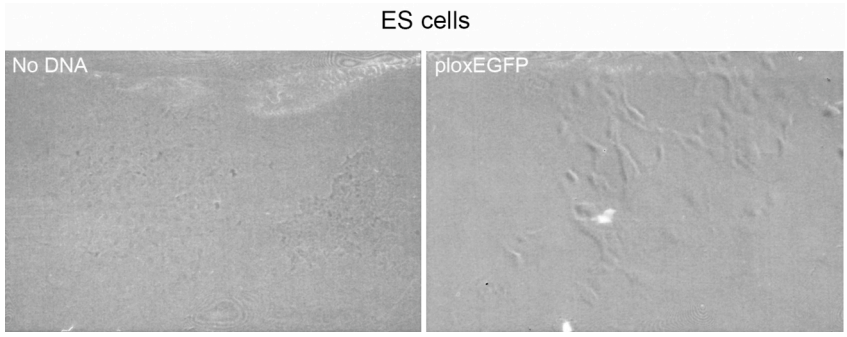


Figure 19 EGFP fluorescence in CGR8.8 129/Ola ES cells. Fluorescent micrograph after transfection showing EGFP fluorescence at the perimeter of a colony.

Although it was disappointing to conclude that the CMV promoter was not available for clone selection in the undifferentiated cells, it was not relevant for the mouse model, and may be beneficial. As explained in 3.4.1, the CMV promoter satisfies the criteria required in that it is at least as powerful and as active temporally as the *DMPK* promoter, which drives expression of the DM1 mutation. It did mean that in order to select clones based on EGFP expression levels, the clones would need to be differentiated after transfection with Cre.

The master transgene *Tg*StopEGFP•TK was excised from the pStopEGFP•TK construct on a *Kpn*I-*Xba*I fragment (Figure 20) removing the unnecessary bacterial growth genes, and introduced into ES cells by electroporation. Cells were plated onto a MEF feeder layer and after 24hrs, transformants were selected for using G418 supplemented medium. Massive cell death occurred on day six revealing 20 bright circular colonies. Once they had reached 1-2mm in diameter, 18 undifferentiated colonies were picked. The cells were disassociated and the clones expanded by growth first in 96 well plates, then 24 well plates, again on a MEF feeder layer. The differentiation state of each clone was determined by visual microscopic inspection and 70% of the sample frozen down. The remaining 30% was expanded by 3 days further culture. DNA was extracted and subjected to three different PCR analyses: '*EGFP*' to confirm the presence of the construct by amplification of a portion of the *EGFP* ORF 'Concatemer' to ascertain the presence of arrays. In this assay primers are positioned to face away from the body of the construct such that a product is only made when the construct has integrated as a tandem array; and '*mUSF*' to verify the presence of PCR grade DNA in the sample. Primers mUSF-A and mUSF-BR were used to amplify a 1019 bp fragment within the *mouse upstream stimulatory factor* (*mUSF*) housekeeping gene. The generation of a PCR product in this reaction confirmed the presence of mouse DNA in the sample.

The *EGFP* PCR was also subjected to Southern hybridisation to an *EGFP* probe for sequence confirmation. Twelve of the clones were positive for *EGFP* and so contained the transgene (Figure 21A). Of these clones, two were determined by concatemer PCR to contain more than one insert (Figure 21B). DNA usually integrates into the genome at a single point either as a single molecule, or a head to tail array (Brinster *et al.*, 1981). The expected band size if an array was present was 437 bp if the array is in the usual head to tail arrangement, which is what we see in clone 12. Clone 15 shows a smaller band of ~220 bp. If the clones were head to head, or tail to tail, the band would be 410 bp or 450 bp respectively. This sample was not positive for *EGFP* (Figure 21), so possibly this clone resulted from a DNA rearrangement.

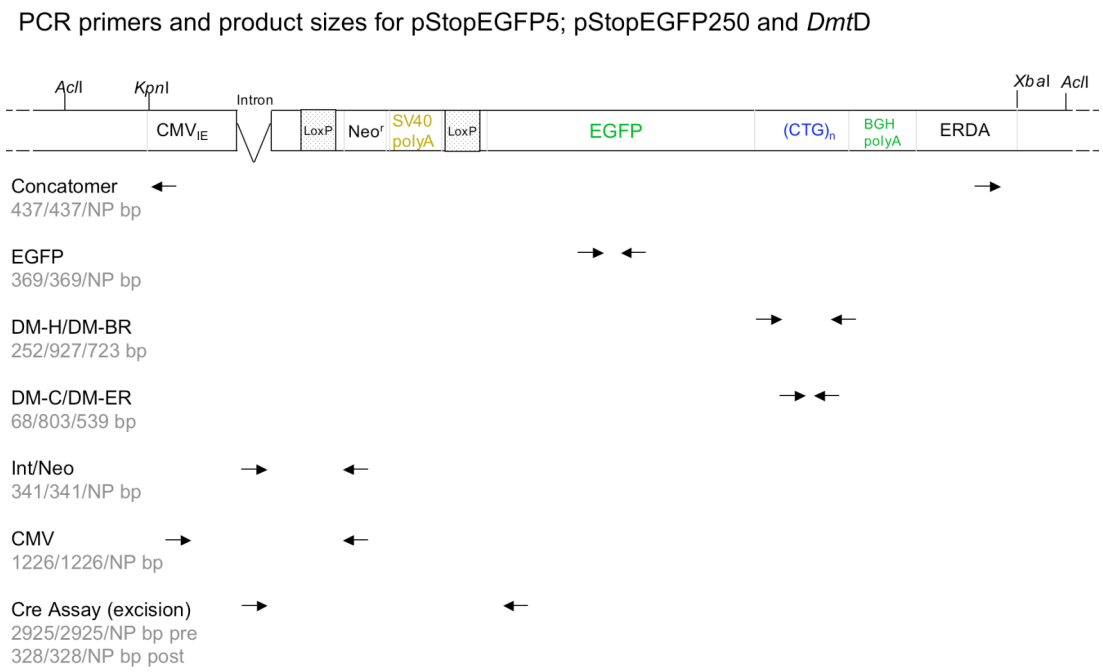


Figure 20 The position of PCR amplimers relative to the transgene. Schematic diagram of linear pStopEGFP5/250 (not to scale) illustrating the position of amplimers used in PCR analyses. Product sizes are given for pStopEGFP5; pStopEGFP250 and *Dmt*D162 (genotyping control) templates. NP = template sequence not present. Restriction sites used for transgene isolation are also shown.

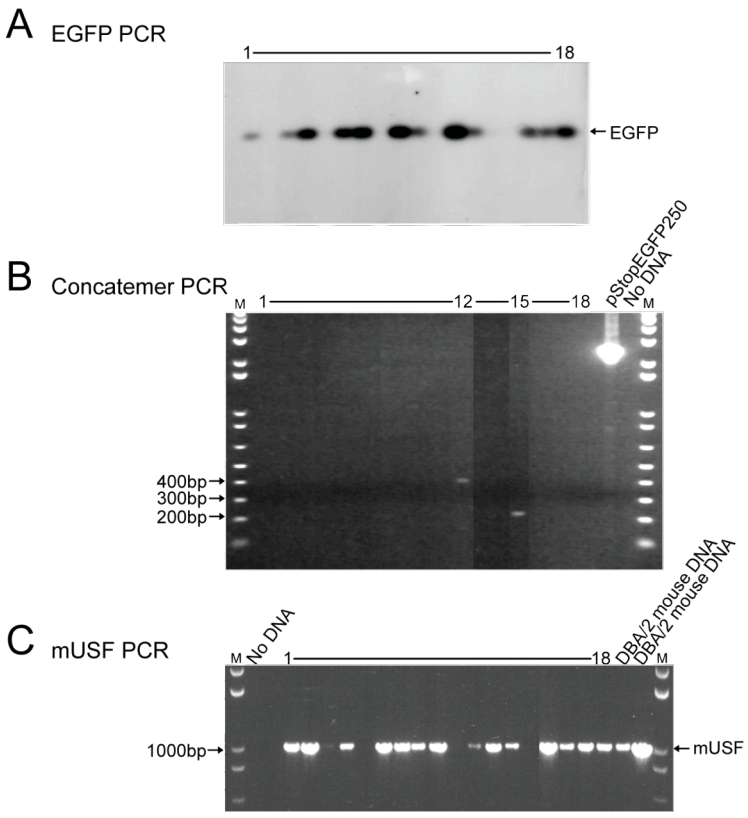


Figure 21 PCR analyses of pStopEGFP•TK ES cell clones A Southern blot of *EGFP* PCR hybridises to an *EGFP* probe. Twelve clones were positive for the transgene. **B** Concatomer PCR: Clones 12 and 15 were positive for arrays. **C** Most preparations were positive for PCR grade DNA.

The cells were at passage 11 as they were plated into 96 well plates and a high proportion of colonies looked undifferentiated. That is to say that the cells upon visual microscopic inspection were growing as densely-packed bright circular colonies with few patches of flattened differentiating cells (Nichols *et al.*, 1990). As the clones were expanded, the number of undifferentiated colonies was much reduced and only in clones 5 and 14 did the colonies look dense, bright and round. Unfortunately both were negative for *EGFP* (Figure 21A).

The design of our model was such that the only difference between lines was the length of the repeat tract. To do this we were attempting to randomly integrate the master construct and then using negative selection to target differing repeat lengths to this construct by homologous recombination. To free the model of positional effects was advantageous because it would remove the further consideration of the integration site in an already complicated disease. Transgenic models such as ours are usually generated by random integration however, with several lines selected from each transgene to control for position effects. For example during the generation of $HSA^{SR/LR}$ myotonic mice, in some lines the transgene was not expressed, or ‘silenced’, possibly because of the integration site

(Mankodi *et al.*, 2000). At this point in our experiments, doubts had been raised by another researcher over the integrity of the CGR8.8 cells we had both been using, since they had failed repeatedly to obtain homologous recombinants (C. Winchester, personal communication). This coupled with the differentiated look of our 'construct positive' clones led us to suspend work on the ES cell strategy, and attempt random integration of the second round constructs (pStopEGFP5 and pStopEGFP250) directly to generate the model by the usual random integration route. This would also establish the function of the Cre-*lox* strategy *in vivo* before returning to, or perhaps repeating the lengthier ES cell targeting strategy at a later date.

3.5.2 Random integration

Pronuclear injections into B6D2 hybrid zygotes were carried out by the Central Research Facility at the University of Glasgow by the procedure documented in "Manipulating the Mouse Embryo" (Hogan, 1994). The transgene was excised on an *Acl*I fragment (Figure 20) which contained 1146bp 5' and 583bp 3' flanking DNA to insulate the ends of the transgene from exonuclease digestion. The DNA was purified using DEAE sephacel, and for the *Tg*StopEGFP5 and the first round of *Tg*StopEGFP250 injections was diluted to 1.8ng/µl in injection buffer, corresponding to approximately 100 molecules per injection. The initial injection concentration was low in order to bias the procedure towards single copy integrants. Since Cre recombinase excises DNA between two *loxP* sites, arrays present within the genome would complicate the excision process, and therefore unpredictably affect activation of *EGFP* expression.

The number of live births totalled 16 after pronuclear injection with *Tg*StopEGFP5 and166 after pronuclear injection with *Tg*StopEGFP250. The number of live births was smaller for *Tg*StopEGFP5 simply because fewer eggs were injected after the detection of a successful integrant. For the *Tg*StopEGFP250 injections, three batches of DNA were used, each prepared from separate cultures and verified by restriction digestion. The first batch was at a concentration of 1.8ng/µl, as for *Tg*StopEGFP5. The second was diluted to three different concentrations to increase the chances of integration: 1ng/µl; 5ng/µl and 10ng/µl, all single-use aliquots to avoid freeze-thaw nicks in the DNA. The third batch was supplied at 2ng/µl, also in single use aliquots.

3.5.3 Genotyping

Once implant mice were weaned, usually between four and six weeks of age, 1cm tail-tips were removed under anaesthetic and the DNA extracted. The presence of the transgene was assayed using MTT multiplex PCR, routinely used in our laboratory for the genotyping of *Dmt*D162 mice. This PCR amplifies three separate bands in the control *Dmt*D162 mice (*mUSF*; DM-C/DM-ER and DM-R/DM-QR) and two in *Tg*StopEGFP5 and *Tg*StopEGFP250 mice (*mUSF* and DM-C/DM-ER). The *mUSF* set of primers are an internal control for the presence of PCR grade DNA, so any samples which did not show up positive for this 1 Kb product were re-analysed with increased amounts of DNA sample, or repurified DNA. If the result was still negative, a second tail tip was used. The extra 180bp PCR product in the *Dmt*D162 control originates from the presence of further *DMPK* 3'UTR sequence 3' to the repeat region, not present in the *Tg*StopEGFP5 and *Tg*StopEGFP250, amplified by DM-R and DM-QR. The third set of primers, DM-C and DM-ER amplify the repeat region in both *Dmt*D162 and the *EGFP* constructs (Figure 20).

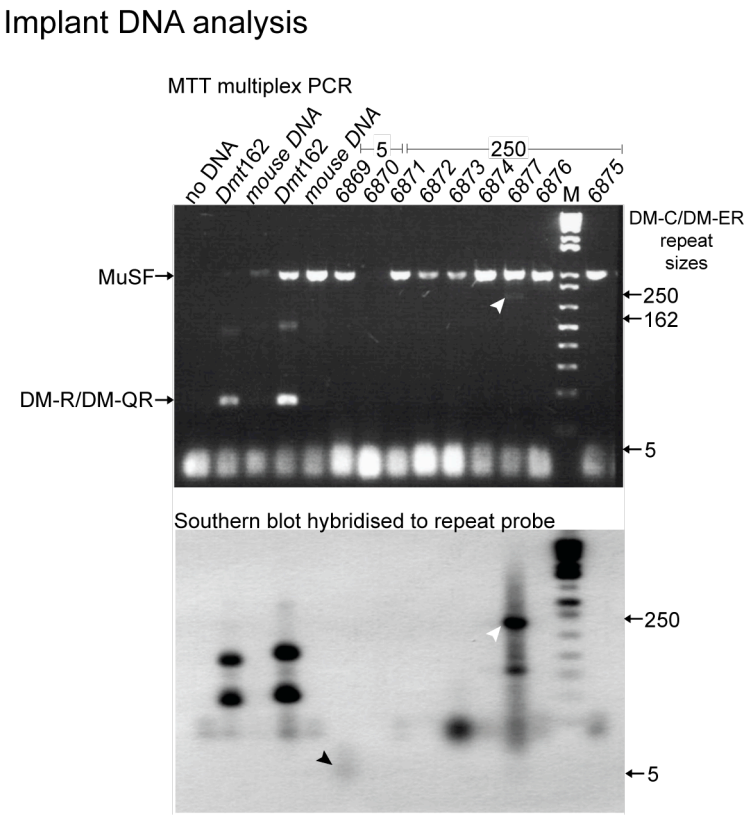


Figure 22 Implant genotyping by PCR Top: MTT multiplex PCR DNA analysis on tail-tip lysates. Bottom Southern blot of the above agarose gel and hybridisation to the DM56 repeat probe. Arrows indicate putative positive mice, black *Tg*StopEGFP5; white *Tg*StopEGFP250.

Two mice positive for the repeat region of the transgene were identified, number 6869 *Tg*StopEGFP5 and 6877 *Tg*StopEGFP250 (Figure 22). The same batch of tail tips was

then subjected to three further PCR assays for verification: '*EGFP*' and 'Intron' PCR to confirm the presence of the construct and DM-H/DM-BR PCR to confirm the presence and length of the repeat region. At this stage mouse number 6869 *Tg*StopEGFP5 was confirmed as positive for the transgene, but 6877 *Tg*StopEGFP250 was negative (Figure 23). The analyses were repeated using second tail tips from the same batch of mice and the results were still found to be positive for mouse number 6869 *Tg*StopEGFP5 and negative for 6877 *Tg*StopEGFP250 (data not shown). Tail-tip lysates from mouse 6869 *Tg*StopEGFP5 were used as a positive control for all further PCR genotyping of implants.

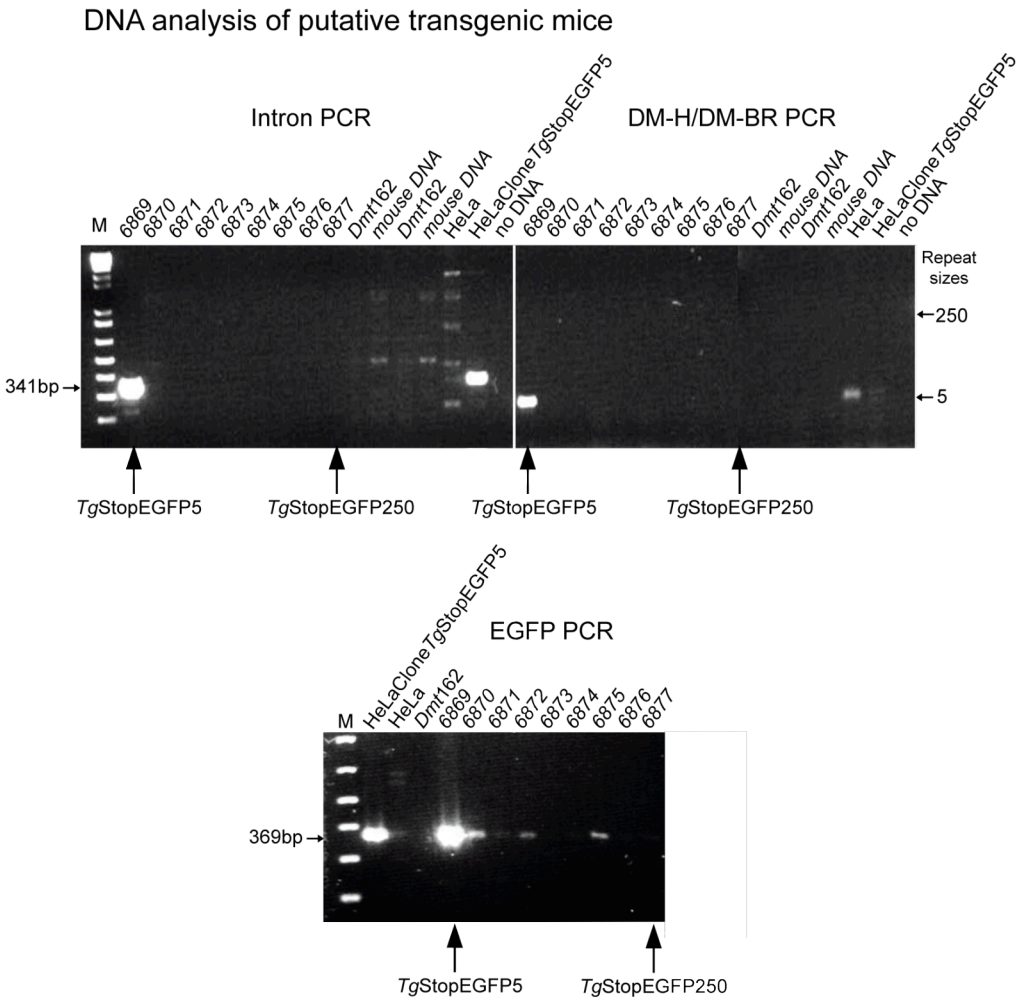


Figure 23 Further analysis of putative *Tg*StopEGFP5/250 mice. Three DNA analyses: Mouse number 6869 *Tg*StopEGFP5 is positive, 6877 *Tg*StopEGFP250 is negative. Intron PCR produces a product of 341bp if the construct is present regardless of repeat length. DM-H/DM-BR PCR spans the repeats producing band sizes of 252bp for 5 repeats and 927bp for 250. *EGFP* PCR produces a 369bp product within the ORF.

The donor egg genotype used for pronuclear injection was derived from a female C57Bl6 - male DBA/2 cross, so to obtain a pure background positive mouse 6869 *Tg*StopEGFP5 was set up with strain C57Bl6 for breeding. Offspring were genotyped using MTT

multiplex PCR and Southern blot hybridisation to the DM56 repeat probe; DM-H/DM-BR PCR was used to verify the repeat length and *EGFP* PCR to show the transgene was transmitted through the germline and to identify mice for breeding (Figure 24). Note that segregation of the transgene was not 50:50. The founder mouse 6869 *Tg*(StopEGFP5) fathered 74 offspring in total, 11 of which were positive for the transgene, and so could have been mosaic, the transgene having integrated after the single cell stage. This apparent effect on segregation could also be the result of embryonic lethality, caused by the unfortunate integration of the construct into an essential gene. Since the founder mouse was male, any male offspring produced will be DBA/2 for the Y chromosome, so for the F2 generation, female transgenics were selected to establish the line.

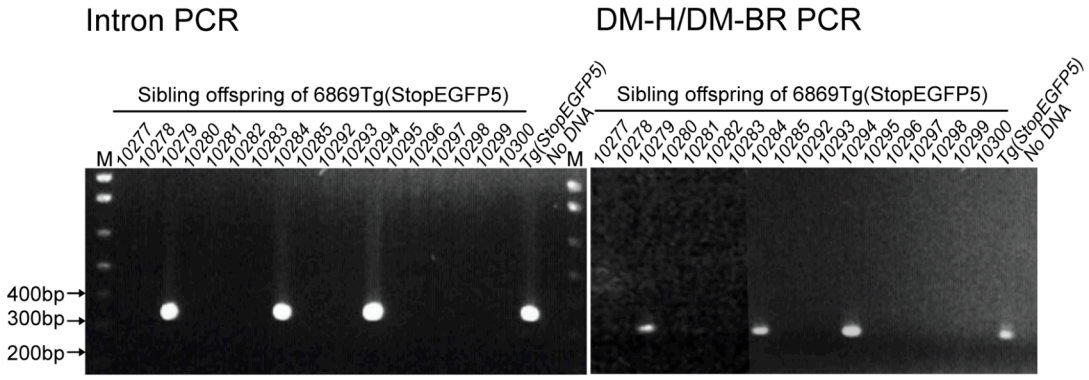


Figure 24 Genotyping of the *Tg*(StopEGFP5) F1 generation. A sample of 6869 *Tg*(StopEGFP5) offspring PCR genotyping detecting transmission of the transgene. F1 mice 10279; 10284 and 10294 are positive. *MUSF* was also amplified to confirm the presence of PCR grade DNA (data not shown).

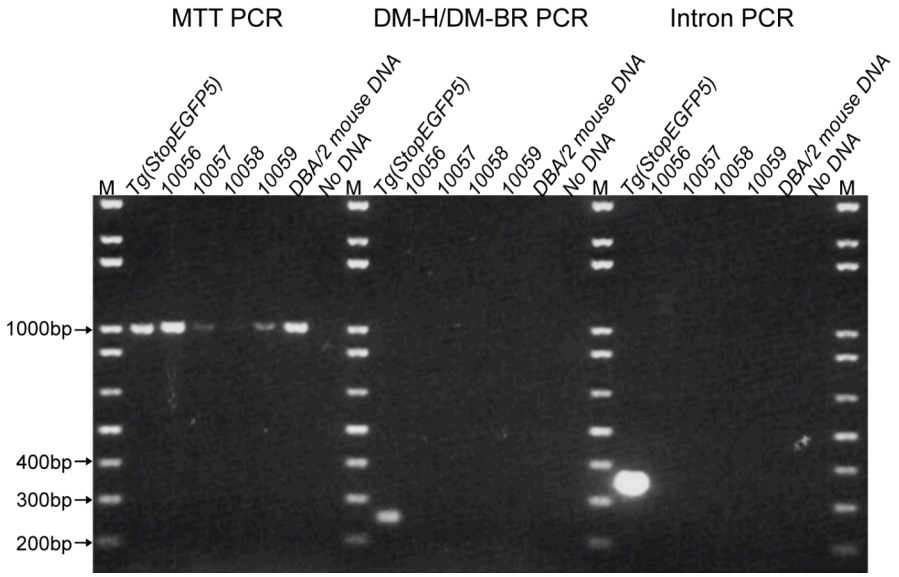


Figure 25 A representative proportion of implant genotype analyses. The presence of amplifiable DNA is determined using MTT PCR. Repeat (DM-H/DM-BR) and Intron PCR verify the presence of the construct and repeat (DM-H/DM-BR) PCR also verifies the CTG array length.

In total 182 live births were analysed after pronuclear injection, 16 with *Tg*StopEGFP5 and 166 with *Tg*StopEGFP250. Finally, to ensure no positives were missed, before sacrifice all implant mice were assayed using *EGFP* PCR; H-BR PCR; and either *mUSF* or MTT PCR (Figure 25), followed by Southern blot analysis using the *EGFP* or DM56 repeat probe as appropriate (data not shown). No positive mice were obtained for *Tg*StopEGFP250.

3.6 The Cre-*lox* mechanism *ex vivo*

To test the function of the Cre-Lox mechanism *in vivo*, cell lines were established from the tails of the F3 and F4 C57Bl6 backcross of mouse *Tg*(StopEGFP5) (Figure 26).

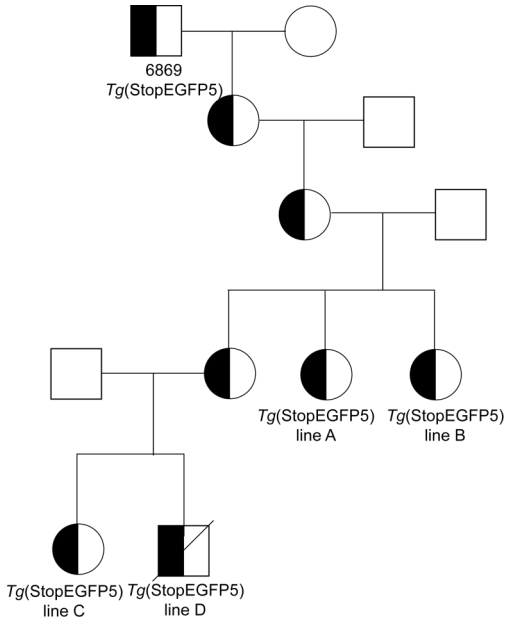


Figure 26 Pedigree of tail cell-lines Derivation of tail cell-lines from founder mouse 6869 *Tg*(StopEGFP5) during C57Bl6 backcross. Line D did not become immortal within the time scale.

These lines strictly-speaking were *ex vivo* and as such should be a good reflection of the situation in the living animal. Tail tips were removed and wiped with 70% ethanol before removing the bones using a scalpel. The tail tissue was flattened out and parallel 1-2mm cuts made almost all the way along the length but leaving a small piece of tissue uncut to keep the strands together. The tissue was then placed onto a 60mm culture dish skin side up and the strands splayed out. The tissue was pressed very firmly in order to adhere to the dish surface, and prewarmed culture medium added gently dropwise. The dishes were incubated at 37ºC until cells could be seen growing out from the tissue, which was then removed. Cells continued to be incubated and passaged until growth slowed. The cells

were then incubated without passage until they began to grow once again. At this stage the cells had overcome senescence, whereby the cells lost the ability to divide, and were considered immortal. Once the lines were continuously dividing, cells were transfected with the Cre-expressing plasmid pCre. The experiment was controlled for using mouse fibroblast 3T3 cells: The constitutively expressed pLoxEGFP served as a positive transfection control, whilst co-transfection with pStopEGFP5 and pCRE acted as a positive Cre-*lox* mechanism control. The pStopEGFP5 and pCRE plasmids were also transfected alone to show no fluorescence prior to Cre activation. The *Tg*(StopEGFP5) cell-lines were also transfected with pLoxEGFP to show that the cells were able to be transfected. All cells on the coverslip were analysed 24 hours later using fluorescent microscopy. None of the *Tg*(StopEGFP5) cell-lines tested showed Cre activated EGFP fluorescence after transfection with pCre (Figure 27). This experiment was carried out three times in total. The same result was obtained; no EGFP fluorescence could be detected after pCre transfection of the *Tg*(StopEGFP5) tail cell-lines.

Why transfection with pCre did not activate EGFP fluorescence was not clear in this experiment. Possibly the construct had suffered rearrangement within the mouse genome causing the mechanism to malfunction. To test whether activation was working at the DNA level, cells were transfected again as described in the previous experiment. DNA was extracted and subjected to PCR with primers that lay outside the region of Cre excision (Figure 20). This detects a shift between product sizes from 2925 bp before excision to 328 bp after excision. The mechanism was found to be functioning correctly in all three lines since both bands could be detected after pCre transfection (Figure 28).

It is possible that the amount of EGFP produced was too low to be visualised directly by fluorescence microscopy. To enhance the EGFP signal α-GFP primary antibody was used with an AMCA-conjugated secondary, which according to the manufacturer can give an eightfold increase in signal. The cell-lines and control 3T3 cells were grown on coverslips and transfected with pCre, or pLoxEGFP as a positive control for both transfection and the antibody hybridisation. Control 3T3 cells were also transfected with pCre or pLoxEGFP, also pStopEGFP5 or pStopEGFP5 + pCre as a positive control for pCre function, and no DNA. After antibody hybridisation, using fluorescence microscopy, all samples transfected with pLoxEGFP showed blue and green fluorescence (Figure 29) data not shown for 3T3/pLoxEGFP). All the cells on each coverslip were analysed. In 3T3 cells, no fluorescence was detected with pCre (Figure 29) or pStopEGFP5 alone (data not shown), or no DNA (data not shown). PCre plasmid was shown to be functional since 3T3 cells co-

transfected with pStopEGFP5+pCre were EGFP positive (Figure 29). It was noted that the signal was at a similar level for cells exhibiting very strong EGFP fluorescence but was generally enhanced for those cells with lower levels of EGFP fluorescence (Figure 29 arrows). However, there was still no signal for the lines tested for Cre activation (Figure 29

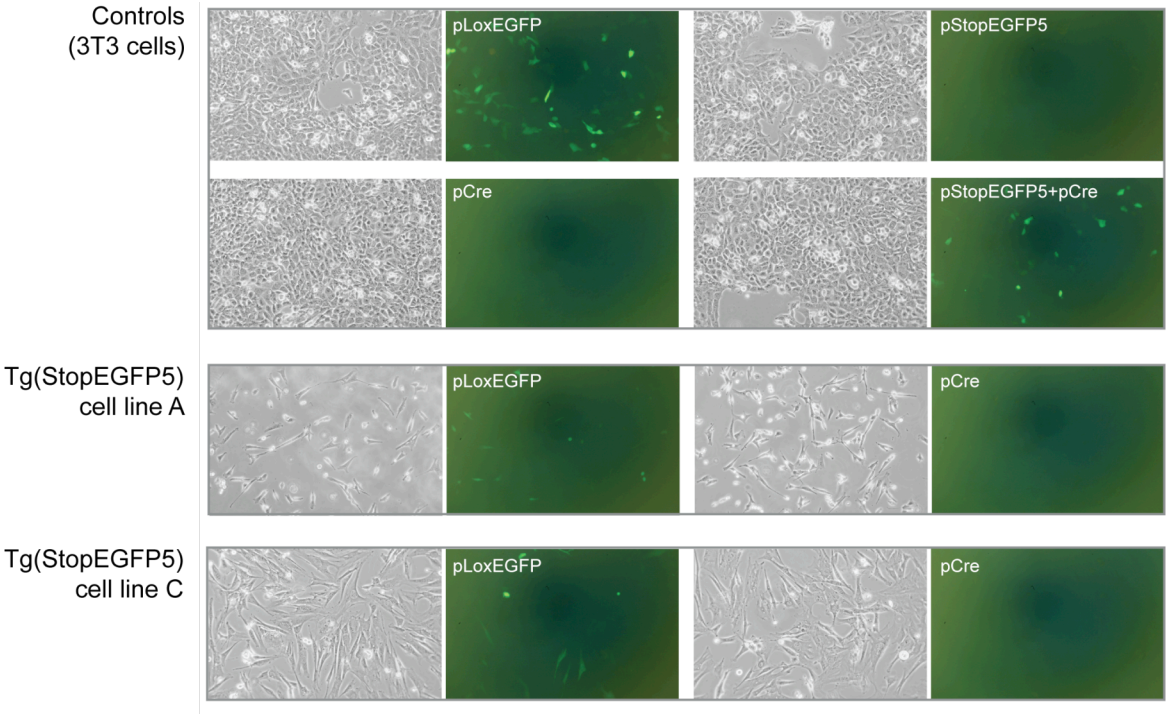


Figure 27 Cre activation of *Tg*(StopEGFP5) tail lines. Brightfield and fluorescent micrographs. PCre transfection of the *Tg*(StopEGFP5) tail cell-lines should have activated EGFP fluorescence by removal of the *loxP*-flanked polyA signal but no signal was detectable using fluorescent microscopy. The controls show that technically the experiment is valid, since transfection of pStopEGFP5 or pCre alone does not result in EGFP fluorescence, whereas co-transfection of pStopEGFP5 and pCre activates fluorescence showing that the excision mechanism is working. PLoxEGFP transfections were used to confirm the ability of the cells to take up and express plasmid DNA.

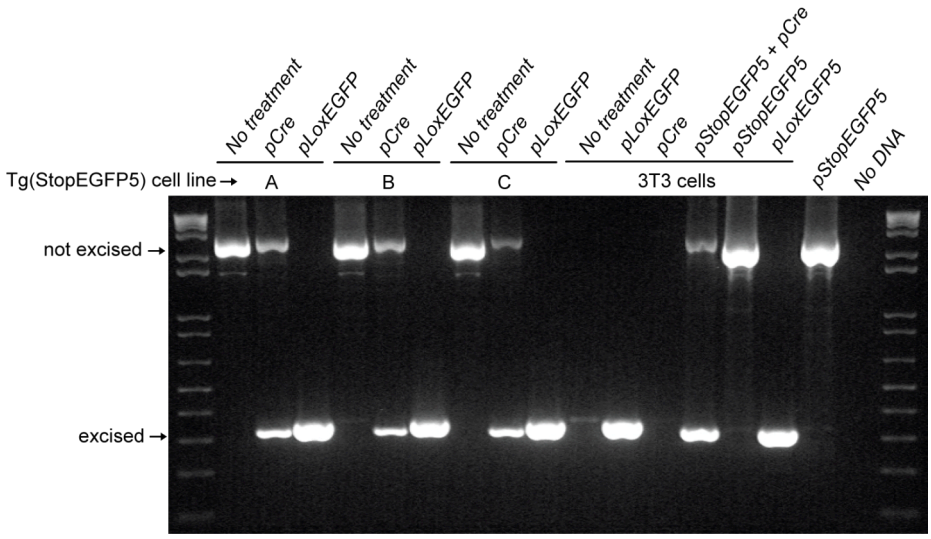


Figure 28 Cre-*lox* excision at the DNA level. Cre assay PCR showing a reduction in product size from *Tg*(StopEGFP5) cell-line DNA after Cre excision. Constitutively expressing pLoxEGFP functions as a positive control for transfection, and a size marker for Cre excision.

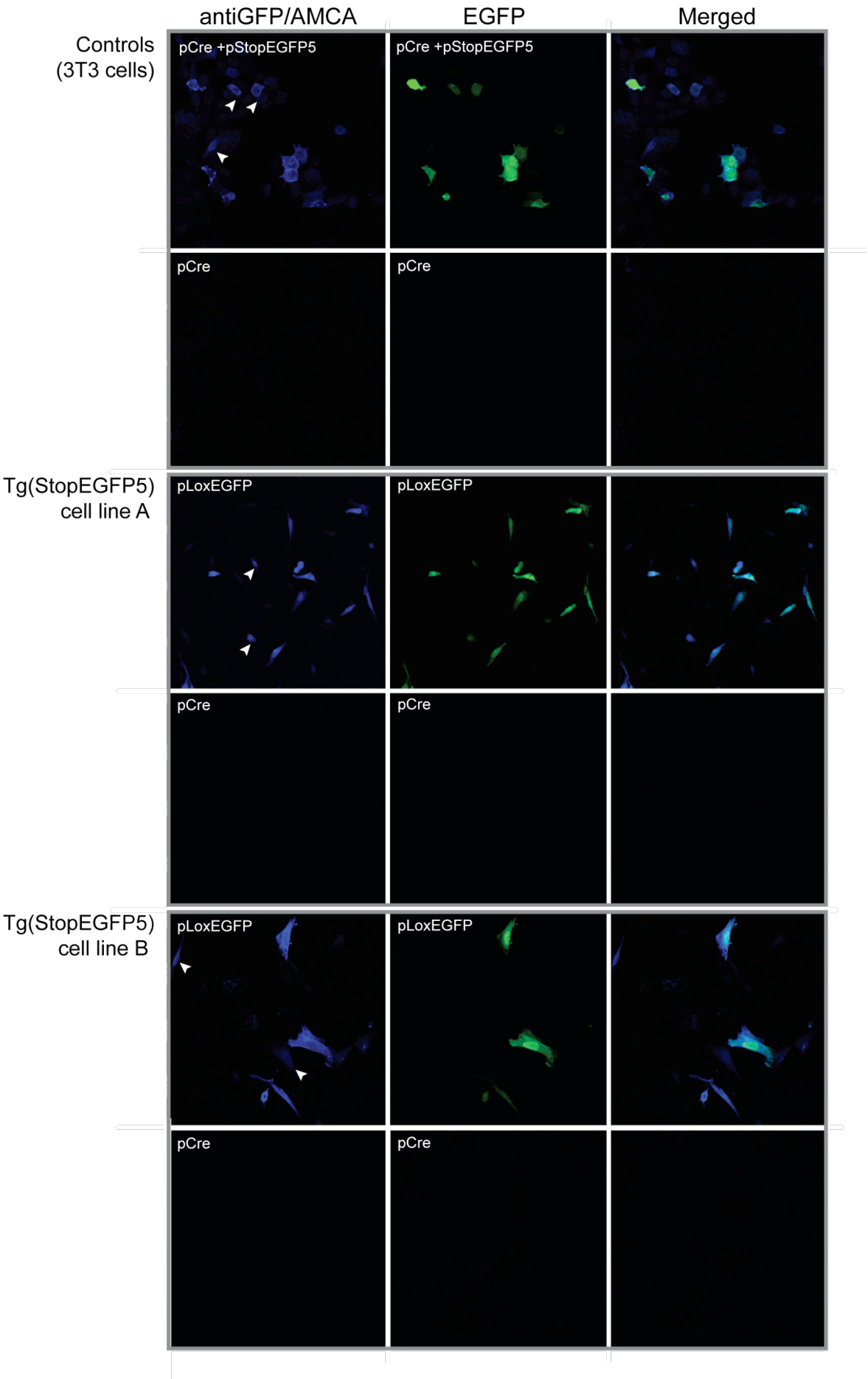


Figure 29 Enhancement of EGFP detection using αGFP. Fluorescent micrograph. *Tg*(StopEGFP5) cell lines were transfected with pCre, which should activate *EGFP* expression. Anti-GFP primary antibody and AMCA (blue) conjugated secondary antibodies were used against

EGFP to try to amplify the fluorescent signal. There is a slight increase in the detection of EGFP expressing cells using these antibodies (white arrows) when compared to EGFP alone. However, even after enhancing detection with αGFP/AMCA, EGFP still could not be detected in pCre-transfected *Tg*(StopEGFP5) cell lines.

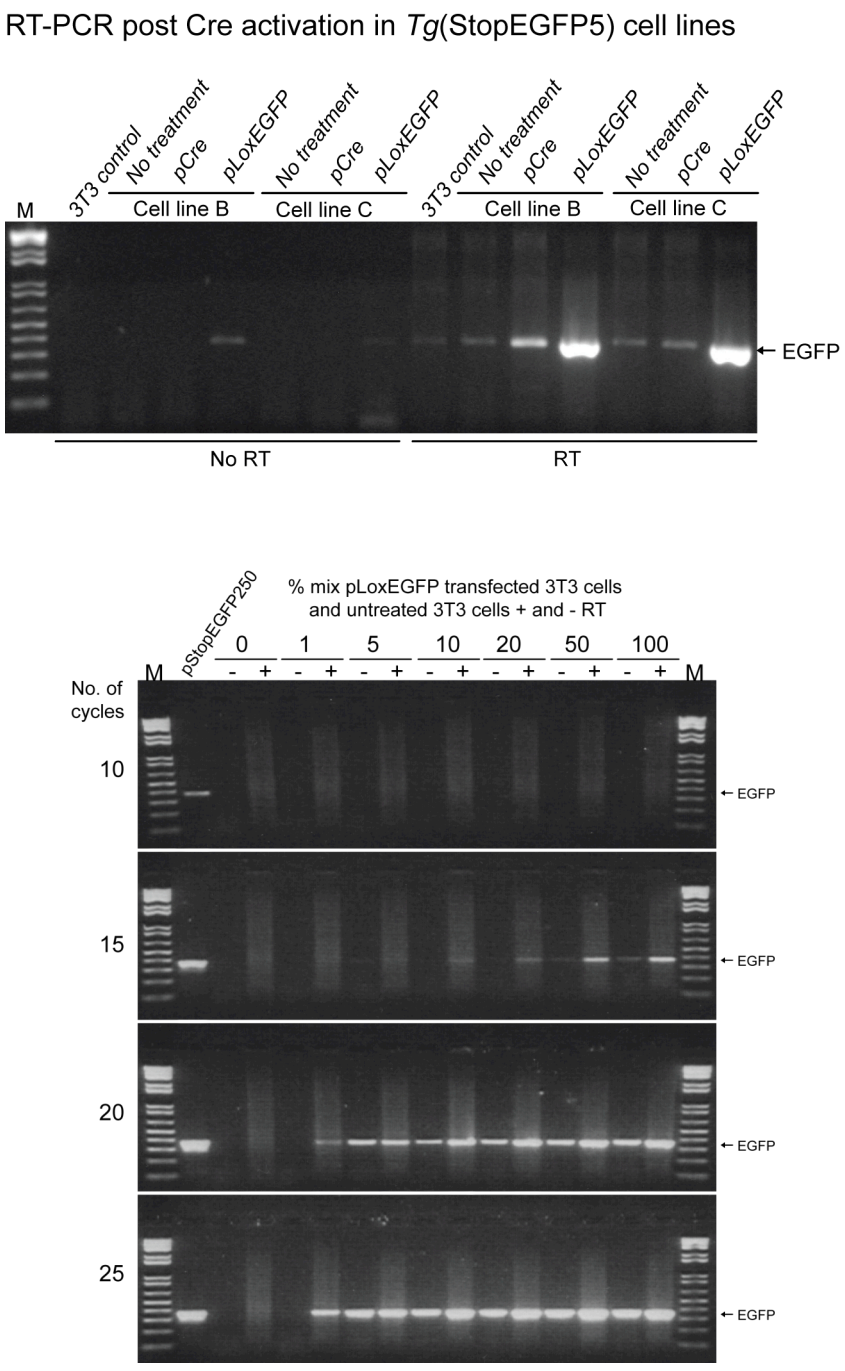


Figure 30 *EGFP* RNA analysis pre and post Cre-*lox* excision. Top *EGFP* RT-PCR. The increase in transcript after pCre activation of expression is small. 20 cycles were used with the conditions described below. **Bottom** Optimisation to ensure amplification is within the linear range. Cells were transfected with pLoxEGFP as previously described, and harvested after 24 hrs. This preparation was '100%' since it was comparable to the test sample transfection rate. The sample was then diluted using untransfected 3T3 cells. Total RNA was prepared from the mixes, and 500ng of RNA was reverse-transcribed in a 10µl volume, 1µl of which was used in each PCR reaction. Reverse transcriptase was omitted from the 'no RT' controls. Amplification was limited to 10; 15; 20 and 25 cycles. 15 or 20 cycles were deemed to be in the linear range depending on the amount of starting sample. NB There is contamination in the 'no RT' controls probably due to plasmid DNA binding to the RNA purification column.

To check whether the *EGFP* transcript was being expressed pStopEGFP5 cells were transfected with pCre; reagent alone (no DNA) or pLoxEGFP as a positive control, and RNA isolated after 24hours. Samples were subjected to optimised *EGFP* RT-PCR (Figure 30 Bottom) to ensure that product levels seen after amplification were a reflection of the amount of starting template. Unfortunately the experiment shows contamination in one of the 'no RT' controls and so should ideally have been repeated (Figure 30 Top). If the increase in *EGFP* transcript after Cre activation is real, and not due to contamination, then the increase is not large. If time allowed, this experiment would have been repeated, using Cre assay primers which span the intron to distinguish between pre and post Cre-excised transcripts. For the mouse model, it is important to ascertain whether there is read-through transcription of the SV40 PolyA signal in the absence of Cre. In this case RNA transcribed would not result in EGFP fluorescence, but would still contain the repeat region -which is the basis of DM pathogenesis, although low levels of expanded triplet repeats have been shown to be without pathogenic effect in mice (Mahadevan *et al.*, 2006).

3.7 Discussion

Our aim was to create a conditional mouse model of DM1, based on the limited expression of expanded CUG repeats linked to an *EGFP* reporter, using the Cre-*lox* system. Expression of the repeat tract would be defined by the pattern of Cre recombinase expression, obtained by crossing our *Lox*-repeat transgenic line with available Cre expressing lines. We planned to generate the transgenic lines using two rounds of integration in ES cells; the first, a random integration of the construct backbone, to obtain an integration site target to which the second round constructs containing differing lengths of repeat tract could be targeted by homologous recombination. Five constructs were created as described in Figure 5: One founder construct pStopEGFP•TK and a 5 and 250 repeat version each of the pStopEGFP targeting construct and the constitutively expressed pLoxEGFP construct. The 800 repeat form proved too elusive for the time available to carry out this work. Three further pLoxEGFP constructs arising from spontaneous deletions during *in vitro* Cre excision were isolated and the 3'UTR length verified by restriction digest. The component parts of all these constructs were tested in at least one stage of their construction and shown to be functional.

One transgenic mouse was generated using the normal repeat allele *Tg*StopEGFP5 by random integration. Cell-lines were established from tail-tip tissue taken from generation

F3 and F4 and used to test the Cre-*lox* mechanism *in vitro*. EGFP fluorescence could not be detected after pCre transfection by fluorescent microscopy directly, or by using α-GFP to increase the signal. PCR analysis of the transfectants revealed the excision mechanism to be functional at the DNA level.

3.7.1 Repeats

We successfully generated the pStopEGFP•TK construct required for first round random integration and pStopEGFP5 and pStopEGFP250 for second round targeting of repeats to integration by homologous recombination. It was our intention to use 5; 250 and 800 expanded repeat tracts for our mouse model to reflect the normal; adult-onset and congenital forms of the disease. Whilst the cloning of 5 repeats was straightforward, 250 repeats was difficult and 800 repeats unsuccessful due to deletions generated during bacterial propagation of the DNA. It was not clear at the outset that the repeat tract itself was changing. We were attempting to clone from a size distribution generated from patient DNA, and often a large band when gel purified will contain a small amount of contaminating short sequences. During cloning of the 250 repeat it was assumed that these fragments were also cloned, and selected for because they might be expected to grow more rapidly since the insert would be smaller. For this reason small colonies were also picked, which would be expected to contain larger or 'difficult to propagate' inserts, but these did not show any increase in size or insert homogeneity. It became clear as larger repeat lengths were attempted it was probable that clones containing long repeats were undergoing a rearrangement either during transformation into the host, or during replication within the bacterium, creating a population of plasmids whereby repeats had been lost, and also perhaps part of the 3'UTR, losing the *Eco*RI or *Xho*I site, or duplicating it leading to the appearance of partially digested samples and large doublets respectively. Intraplasmid rearrangements have been shown to be mediated by CTG repeats in *E. coli* (Hashem *et al.*, 2002; Wojciechowska *et al.*, 2005). Another factor influencing rearrangements may have been introduced as the repeats were cloned. Here, blue-white screening was used whereby the insert is cloned into the *lacZ* gene. This disrupts translation of β-galactosidase and the subsequent hydrolysis of x-gal substrate to an insoluble blue dye, resulting in white colonies for cloned inserts. Using this mechanism means that the cloned insert was also expressed. It has been shown that in *E.coli*, induction of transcription of repeats results in an increase of deletions within the array (Bowater *et al.*, 1997). Expression must not have been the only factor since heterogeneity within the repeat was still apparent upon re-transformation and propagation of further batches of 250

repeat constructs without blue-white screening, where the *lacZ* gene was repressed by LacI (*lacI*^{q}) (as in Figure 18A). It would be interesting to see how recently developed bacterial hosts such as SURE or stbl2 cells would fare. SURE cells have tight lacZ control and are recombination (*recABJ*), UV and SOS repair deficient, designed to grow up plasmids harbouring repeated sequences and those with abnormal secondary structure such as Z-form DNA.

An alternative method designed to create large repeat tracts is to clone arrays of interrupted repeats as done by de Haro *et al.* to create a *Drosophila* model of DM1 (de Haro *et al.*, 2006). Here 24 blocks of $(CUG)_{20}CUCGA$ were used to create an interrupted repeat of 480, $(iCUG)_{480}$. In this instance expression of $(iCUG)_{480}$ led to DM-like symptoms of muscle wasting and degeneration. Furthermore, $(iCUG)_{480}$-induced symptoms could be modified by altering levels of MBNL1 and CUG-BP1, RNA binding proteins implicated as pivotal in DM pathogenesis. So clearly a pure repeat tract is not necessary to induce DM-like symptoms. Whether interruption of the repeats modifies the effect of the mutated array in any other way has not yet been established, but recent research indicates that interruptions reduce the rate of repeat expansion (Braida, unpublished) and may modify disease progression as proposed by Matsuura *et al.* in SCA10 (Matsuura *et al.*, 2006).

3.7.2 Random integration

We set out to use ES cells to generate our transgenic model, but because of perceived doubts about the integrity of the cell line, we switched to random integration, the usual method for generation of transgenic models such as ours. We used the new transgenic services of Glasgow University's Central Research Facility to carry out the injections. They successfully generated a *Tg*(StopEGFP5) transgenic mouse using pronuclear injection. However, multiple attempts to generate the mutant *Tg*(StopEGFP250) mouse failed. According to the literature (Hogan, 1994; Pinkert, 1994) and our in-house training course ("Transgenic mice: Applications and Methodology" run by Dr. J.B.Wilson) one would expect to obtain four positive transgenic animals from each 100 eggs injected. In generating transgenic mice by pronuclear microinjection, typically 50-80% of eggs survive the injection, and 10-30% of implanted microinjected eggs can be expected to reach term. Of those mice born usually 20-40% are transgenic (Hogan, 1994). We obtained 1 transgenic mouse from 182 live births. We have been unable to obtain complete records as to the number of eggs injected to create all 182 live mice, since it transpires that records were not kept by the facility in the early days. Of those records that were made, of 2071

eggs that were injected with *Tg*StopEGFP250, 716 survived injection (34%), 710 were implanted and of these 82 (12%) were born live, and 1 pup was found dead. None of these animals were positive for the transgene (dead pup not tested). The survival rate in these experiments was half the expected rate for injections. This could have been due to the injection process, handling, culture conditions or the DNA preparation. Efficiency in terms of DNA integration and development of eggs to term depends on the concentration of the DNA, with increasing concentration resulting in higher frequencies of integration, but a reduction in survival (Hogan, 1994). Brinster *et al.* found 1-2µg/ml to be optimal (Brinster *et al.*, 1985). Here 4 batches of DNA were prepared by 3 different recommended methods of purification: CsCl/agarose gel/NA45 paper; endofree maxiprep/agarose gel/dialysis, and endofree maxiprep/agarose gel/DEAE sephacell. Although a low concentration of DNA was used to bias towards incorporation of single molecules, it was within the optimal range.

The number of mice born was low, 12% of implants, but just within the expected range for live births after implantation (10-30%). Reasons causing this could be technical such as overculture of embryos followed by oviduct rather than uterine transfer, or construct related such as embryonic lethality caused by 'leaky' transcription of repeats through the SV40 terminator. During genotyping of the *Tg*StopEGFP5 F1 progeny it was noted that the transgene segregation was not 50:50 as would be expected from integration during the single cell stage. Analysis of the F1 progeny using a chi square test showed that transgene integration most likely happened during the 4 cell stage. Results calculated for transgene integration at each of the early stages of embryonic cell division were: 1 cell, 36.54; 2 cell, 4.05; 4 cell, 0.38 and 8 cell, 9.37. The control *Tg*StopEGFP5 was the only positive transgenic mouse to come from the facility, the number of total injections unknown, but at least 2 other research groups were using the service. Although aliquots had been given it transpired that the DNA had been stored at 4ºC and repeatedly used. This would result in reduced concentration due to molecular adhesion to the tube wall over time, and perhaps degradation. Subsequent batches were given as single use tubes to overcome this, storage specified as -20ºC. The 5 and10ng/µl concentrations in this second batch would also have compensated for reduced concentration, although (assuming injection of equivalent numbers) the numbers of mice being born seemed to reduce with increasing concentration; 20; 14 and 11 from 1; 5 and10ng/µl respectively, indicating that this perhaps was not an issue. Technically, it could be that integration is more difficult for larger repeats, but DM mouse models constitutively expressing up to 300 repeats have been successfully

generated (300 repeats: Seznec *et al.*, 2000 ; 250 repeats: Mankodi *et al.*, 2000; 162 repeats: Monckton *et al.*, 1997), so this strategy should also have been possible.

It has been recently published that over expression of 5 CTG repeats in a mouse model results in a major phenotype a few days after induction or after nine months without induction due to a leaky promoter (Mahadevan *et al.*, 2006). Our *Tg*StopEGFP5 mouse was produced in 2003 and since it was deemed a normal control line and appeared healthy, no phenotypic analyses were carried out. It was later sacrificed due to unspecified ill-health aged 22 months.

3.7.3 Cre-*lox* mechanism and activated fluorescence

The mechanism of Cre-*Lox* activation of EGFP expression is functional *in vitro* at the DNA level and in cell culture, by pre-treatment of pStopEGFP5 or pStopEGFP 250 with recombinant Cre prior to transfection, and *in vivo* by co-transfection of pStopEGFP5 or pStopEGFP 250 and pCre into *Dmt*D162 kidney; Cos7; Hela and 3T3 cells. It is not clear why no fluorescence was seen after pCre transfection of *Tg*StopEGFP5 tail lines even after α-GFP enhancement. Is it that the *EGFP* expression levels are too low? Ideally the EGFP RT-PCR analysis on pCre transfected *Tg*StopEGFP5 tail lines should be repeated, since contamination was evident within the controls, albeit low level (Figure 30). So, whilst it is not possible to determine whether low levels of RNA are expressed before Cre activation from this analysis, it is apparent that the levels of RNA present after Cre activation are not high. It is assumed during the transfection process that equal amounts of DNA enter the cell when comparing tail cells transfected with pCre to the pLoxEGFP control, but the actual *EGFP* transcriptional target gene may vary hugely in copy number between the two, since the target for pCre is endogenous. Cre recombinase also catalyses the integration of the circularised excision product, so at any one time it can be assumed that only half of the target gene will be activated for fluorescence. It could still be that the *EGFP* levels are too low to be detected, which could have been investigated by comparing expression levels in another GFP fluorescent mouse. Another explanation in this *ex-vivo* experiment, is that Cre recombinase itself is interfering with transcription. If high levels of EGFP are produced in the control EGFP transfection, then it can be assumed that high levels of Cre are produced in the pCre transfection since the promoters are of a similar strength. Large amounts of Cre may lead to continuously occupied *lox* binding sites, which would be expected to interfere with transcription. Allowing plasmid loss post-transfection by extended incubation of the cells could test this theory. As transfected cells divide, the plasmid is lost naturally in the absence of positive selection due to uneven segregation of the molecules. After the plasmid

disappears, the amount of Cre within the cell would diminish, freeing *lox* binding sites and allowing production of EGFP. These cell-lines are representative of a single integration event not fully characterised. Positional effects can contribute to the level of expression within a gene. There may be more rearrangements, tandem integration or arrays, each of which could adversely affect Cre excision, so there are a multitude of reasons as to why Cre does not activate EGFP fluorescence in these single-founder cell-lines. Whilst the *ex vivo* approach should give a good reflection of what's happening in the mouse model, the amounts of pCre DNA entering the cell in a transient transfection are extremely variable and unpredictable. In the true *in vivo* situation where Cre and *lox* mouse lines are crossed in order to induce expression, only one transgene would be present in each cell so this approach may still have been successful.

4 Design and characterisation of a cell culture model of DM1 pathogenesis

4.1 Synopsis

In both myotonic dystrophies type 1 and type 2, the mutant transcript becomes trapped within the nucleus forming foci, which co-localise to three members of the muscleblind family of double-stranded RNA binding proteins MBNL1; MBNL2 and MBNL3. MBNL proteins act in opposition to CUG-BP1, a single-stranded binding protein, to regulate the alternative splicing of target genes during development. Recruitment of MBNL1 to nuclear foci of expanded repeat arrays transcribed from the mutant gene, is thought to alter the dynamic balance between these regulators, resulting in the missplicing of genes, some of which can be directly related to a specific symptom of the disease.

In chapter three we generated founder and targeting constructs to create a conditional mouse model of DM1, based on the limited expression of CTG repeats positioned within the 3'UTR of the *EGFP* gene using the Cre-*lox* system. Here we have used the targeting constructs pStopEGFP5 and pStopEGFP250, which allow transcription of *EGFP* and the repeat tract only after Cre excision of the upstream SV40 polyA signal, and their constitutively-expressing floxed derivatives pLoxEGFP5 and pLoxEGFP250, to create an inducible cell-culture model of DM1 to study early pathogenic changes. Our model replicated the key markers of pathogenesis observed in DM patient cells. In HeLa cells, we have shown that expression of the expanded repeat transcripts from constitutively expressed construct pLoxEGFP250 (chapter 3) formed foci within the nucleus, which co-localised with MBNL1 protein. Foci were also formed in Cos7 and 3T3 cells. Since MBNL1 and CUG-BP1 proteins are central to the aberrant splicing seen in DM we also confirmed their presence within the cell. During co-localisation experiments using an MBNL1/GFP fusion protein it was noted that expression of MBNL1/GFP, therefore increasing the amount of available MBNL1 within the cell, increased the size of foci formed, supporting the MBNL sequestration hypothesis, but also suggesting a protective role of MBNL1 in DM1.

4.2 Introduction

Recent evidence suggests that the genes affected in myotonic dystrophy employ alternative splicing to produce different isoforms, regulated during development through a dynamic balance between two splicing regulators MBNL1 and CUG-BP1 (Ho *et al.*, 2004). These proteins are antagonistic, favouring either the inclusion or exclusion of exons by adjacent site-specific binding during splicing of the messenger RNA, producing the embryonic or adult form of the gene when appropriate. Sequestration of MBNL1 protein within nuclear foci, and/or the increase in nuclear CUG-BP1 activity seen in DM1 patient cells, is thought to alter the balance between the two regulators and causes regulation of splicing to shift in favour of the production of embryonic or dysfunctional forms (by using cryptic splice sites) of the proteins concerned (Ladd *et al.*, 2005; Lin *et al.*, 2006). Although it is clear that missplicing is caused by this shift in the balance of regulation, little is known about the mechanism of regulation itself, and what regulates the regulators. CUG-BP1 is a single stranded binding protein whereas MBNL1 binds double-stranded RNA, so it may not be as straightforward as binding stoichiometry during splicing. In fact recent research indicates the mechanism may involve complex formation between either CUG-BP1 or MBNL1 and another splicing factor, heterogeneous nuclear ribonucleoprotein H (hnRNP H) (Paul *et al.*, 2006). In DM1 patient cells, and in cells expressing expanded CTG arrays, the activity and concentration of CUG-BP1 increases in the nucleus (Timchenko *et al.*, 2001), but the overall total of nuclear and cytoplasmic CUG-BP1 is constant. What causes this and the reasons for it are unknown, so significant gaps remain in our understanding of DM pathogenesis.

Myotonic dystrophy is a multisystemic disease, which varies in severity between individuals and progresses throughout the lifetime of the individual. The mutation expands somatically at varying rates between tissues and continues as the patient ages, the symptoms increasing in severity caused by the increase in array length. Sample donations from patients are invasive involving tissue biopsies and so are infrequent, nowadays rarely done since the advent of molecular CTG repeat analysis used for diagnosis, rather than skeletal muscle histology (Harper, 2001). The poor availability of patient samples and the variation between them confounds analyses done using them. Mouse models, and cell-lines generated from patient tissue have been used and are useful for pathogenic study, but the disease process is fixed and already underway. When we look at samples from patients it is not clear whether primary or secondary changes are being studied. For instance, researchers have reported decreased levels of DMPK in patient muscle (Fu *et al.*, 1993;

Koga *et al.*, 1994). Salvatori *et al.* reported that decreased expression of DMPK correlated with the proportion of type 1 fibres compared to type 2, and that the decrease therefore was probably due to secondary effects resulting from type 2 fibre atrophy rather than a direct effect of the repeat on DMPK levels (Salvatori *et al.*, 2005). It is therefore difficult to use patient samples to study the primary events arising from the earliest transcription of the mutant gene, as may be the case with the congenital form of DM1. An inducible tissue culture model would help to fill these gaps. Cell-culture based systems are easy to use, readily available and consistent. Depending on the cell-type used, the model would help to unite data already obtained from both human cells and samples and mouse models to study the effects of expanded CTG arrays on the processing of downstream targets. Since DM is a multisystemic disorder not purely affecting muscle, we chose to use non-muscle derived cell-lines to encompass rather than exclude any possible global effects.

4.3 Validation of an inducible model

Here we utilise transient transfection as a means to activate expanded array expression from the targeting constructs generated in chapter 3, allowing the study of pathogenesis from the first hours of expanded repeat transcription.

4.3.1 Cre, Lox and foci

The model is based on the expression of normal length or expanded CUG repeats placed within the 3'UTR of an *EGFP* reporter transcript. The induction of EGFP fluorescence and therefore expanded array expression by the Cre-*lox* excision mechanism has already been shown in 3T3 cells (Chapter 3.44, Figure 10). Here, pStopEGFP5 and pCre were transfected into 3T3 cells, EGFP fluorescence could only be detected when the two plasmids were transfected together. In the same experiment, *in vitro* Cre-excised derivative pLoxEGFP was also transfected into 3T3 cells and produced EGFP fluorescence. Activation of EGFP by Cre excision has also been successfully achieved in Cos7; HeLa and *Dmt*D162 kidney cell-lines using co-transfections of pStopEGFP5 or pStopEGFP250 and pCre plasmids (data not shown).

In both myotonic dystrophy type 1 and type 2, the mutant transcripts become trapped within the nucleus and are retained in the form of foci (Davis *et al.*, 1997; Liquori *et al.*, 2001). To determine whether our expressed transcripts form foci within the nucleus, Cos7 cells were co-transfected with pCre and pStopEGFP5 or pStopEGFP250, and fluorescent

in situ hybridisation with Cy3 (red) conjugated $(CAG)_{10}$ or $(CTG)_{10}$ oligonucleotides performed. The CAG repeat oligonucleotide is expected to bind to the CUG expanded array, whilst the CTG repeat oligonucleotide should not, acting as a negative control. Nuclei were counter-stained with DAPI, a fluorescent blue dye.

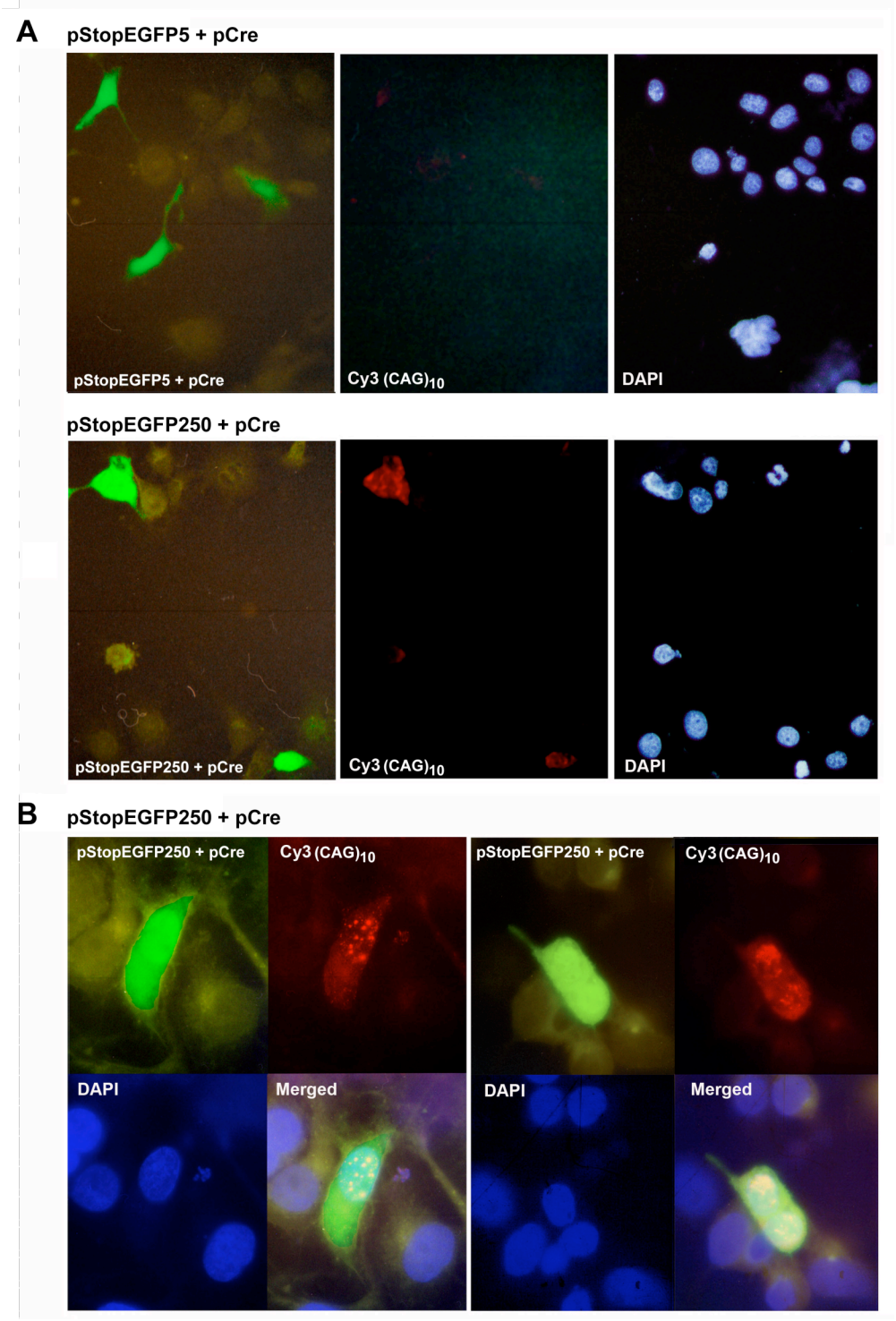


Figure 31 Nuclear foci in Cos7 cells expressing 250 repeats. A Fluorescent micrographs Low power magnification of Cos7 cells co-transfected with pStopEGFP250 or pStopEGFP5 and pCre. Foci can bee seen in the pStopEGFP250 repeat/pCre co-transfectants but not in the pStopEGFP5 repeat/pCre co-transfectants. **B** High power magnification (X63) detailing foci within the nucleus in Cos7 cells co-transfected with pStopEGFP250 and pCre plasmids.

Foci were detected in the nuclei of cells when co-transfected with pCre and pStopEGFP250, but not with pCre and pStopEGFP5 (Figure 31). No foci were detected using the Cy3-$(CTG)_{10}$ (data not shown). The Cy3-$(CAG)_{10}$ fluorescent *in situ* hybridisations were repeated using HeLa and 3T3 cell-lines. Foci were detected in nuclei when co-transfected with pCre and pStopEGFP250 but not with pCre and pStopEGFP5 (Figure 32). Therefore, in our model, the transcribed RNA containing expanded CUG repeats forms foci within the nucleus, and in cell-types other than muscle; monkey kidney (Cos7), human epithelial (HeLa) and mouse fibroblasts (3T3). It was observed that in general the foci were larger in Cos7 cells, and smaller and more well-defined in HeLa and 3T3 cells (Figure 31 and 32), which, if MBNL1 is limiting in foci formation, could be due to differences in endogenous MBNL1 levels between the cell-lines.

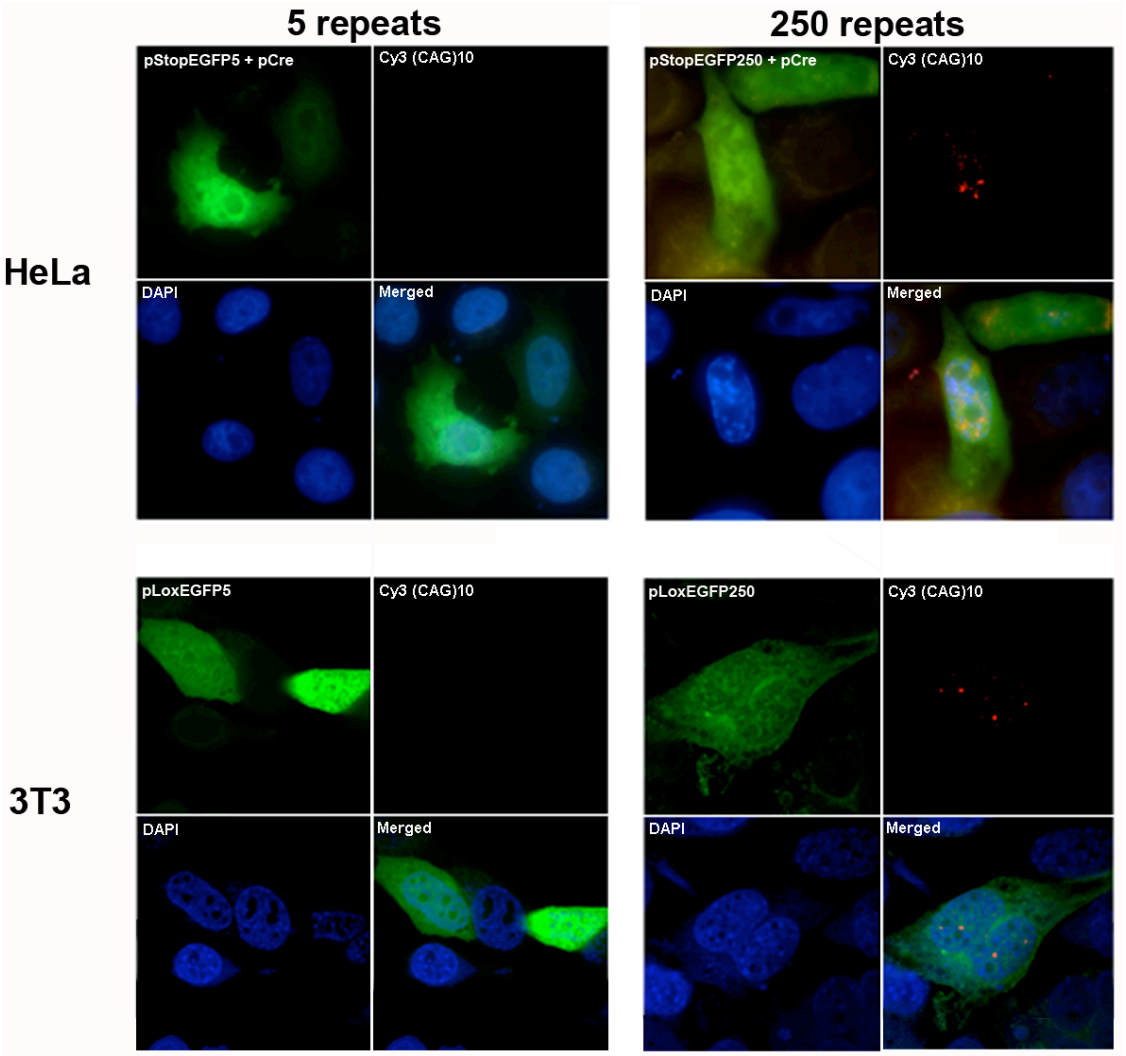


Figure 32 Fluorescent micrographs showing nuclear foci of expanded repeat transcripts in HeLa and 3T3 cells Fluorescent micrographs. No foci are detected by Cy3-$(CAG)_{10}$ *in situ* hybridisation using constructs expressing only 5 repeats. Hela cells were transfected with pStopEGFP5 and pCre. 3T3 cells were transfected with constitutively-expressing pLoxEGFP5. NB All cells on the coverslips were assessed for foci formation, representative samples have been shown as detailed close-ups in order to visualise foci.

Here we used co-transfection to test the Cre-*lox* mechanism and foci formation, which results in a low number of cells expressing both plasmids and consequently a low number of cells expressing the expanded array. Single versus double transfection has been shown in chapter 3 to give significantly increased EGFP fluorescence when using equivalent amounts of DNA (Chapter 3.44 Figure 10, pStopEGFP5 + pCre *vs.* pLoxEGFP5). Single transfections could be used to activate expression in our model using the cell-lines made in Chapter 3 (3.4.2) for testing neomycin function: HeLa and Cos7 cells stably transfected with pStopEGFP5 or pStopEGFP250. Expression of the reporter and expanded array could be activated by the simple transfection of pCre.

4.3.2 Stable cell-lines

To confirm Cre-*lox* function in these lines, Cos7 and HeLa pStopEGFP5 and pStopEGFP250 stable cell-lines (*Kpn*I-*Xba*I fragment, chapter 3, Figure 20) were transfected with pCre. Cos7 and HeLa cells were transfected with pLoxEGFP as a positive control for transfection and fluorescence, or pCre alone as a negative control. EGFP fluorescence was detected microscopically in the pLoxEGFP control transfection, but not in the clones, HeLa or Cos7 cells using pCre (data not shown). The experiment was repeated, and produced the same negative result (data not shown). PCR analyses had been performed previously to confirm the presence of the transgene: EGFP PCR on the Cos7 clones and for the Hela clones, EGFP (chapter 3, Figure 9) and DMH-BR PCR over the CTG repeat region (Chapter 3, Figure 20 for primer positions). To confirm the presence of the promoter, CMV PCR was carried out on a subset of the Cos7 clones, selected to reflect high, medium and low levels of EGFP product, since this could relate to the transgene expression level which may have been useful in future analyses. The construct was present in most clones as determined by EGFP and H-BR PCR analysis, but in the PCR obtained for the CMV promoter, product was absent in most isolates (Figure 33). Ideally the reaction should have been repeated since there is contamination present in the Cos7 untransfected control, however this has no bearing on the interpretation of the negative result. Perhaps the close proximity of the 5' end to the promoter region had resulted in promoter sequence loss due to endonuclease activity during integration. To check this, the *Kpn*I -*Xba*I transgene fragment (Chapter 3, Figure 20) was tested by direct transfection into Cos7 cells to make sure it was functional. No fluorescence was evident (data not shown). A larger *Acl*I fragment, incorporating a further ~300 bp of flanking region (Chapter 3, Figure 20) was isolated and co-transfected into Cos7 cells with pCre, and successfully expressed EGFP (data not shown). The *Acl*I fragment was then stably integrated into Cos7

cells and ten clones from each of pStopEGFP5 and pStopEGFP250 isolated. Unfortunately, once again, transfection with pCre did not activate EGFP fluorescence (data not shown).

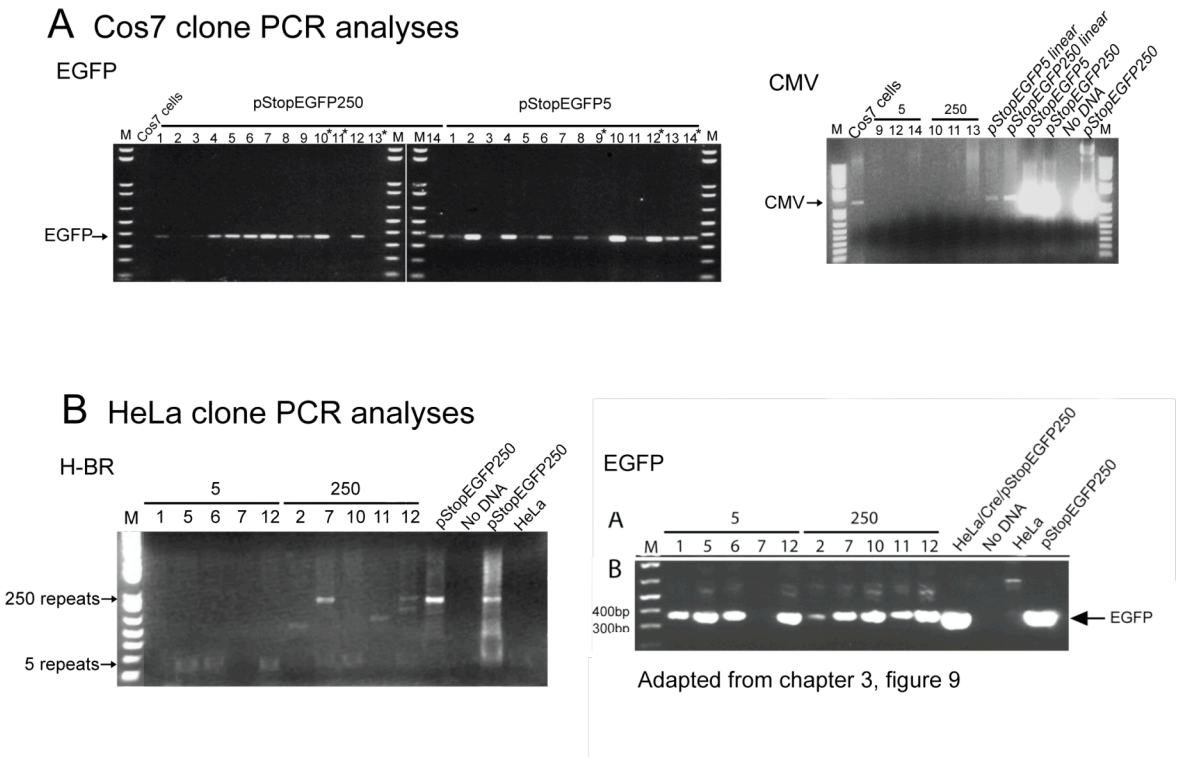


Figure 33 The transgene is present within the stable clones. PCR analyses carried out to verify the transgene integrity. **A** Cos7 clones. CMV PCR was carried out on a subset of samples asterisked in the EGFP analysis. **B** HeLa clones. Most clones positive for DNA amplified both *EGFP* and H-BR PCR products. Note the different repeat lengths in "250" clones.

It is not clear why this approach was unsuccessful. Possibly, the transfection of pCre used to activate the stable lines, resulted in many copies of the recombinase in comparison to the low copy number of the integrated transgene. This could have generated high levels of recombinase relative to the target transgene, interfering with transcription of the reporter gene by physically blocking RNA polymerase with incessant *loxP* binding.

4.3.3 The constitutively expressed transgene

The advantage of using stably pStopEGFP5/250-transfected cells was to increase the number of cells actively expressing EGFP after transfection. Since cells already harbour pStopEGFP5/250 integrated within the genome, cells need only a single plasmid transfection of pCre for EGFP gene expression, rather than the pCre + pStopEGFP5/250 co-transfections previously used for transgene expression. So, rather than induce expression of EGFP via transient Cre activation of stable pStopEGFP5/250 cell-lines, this single transfection approach can also be done using pre-excised pLoxEGFP5/250 (Figure

34). In fact, transient transfection of the constitutively expressing floxed transgene confers an advantage over stably integrated transgenes due to the lack of position effects, often noted after genome integration.

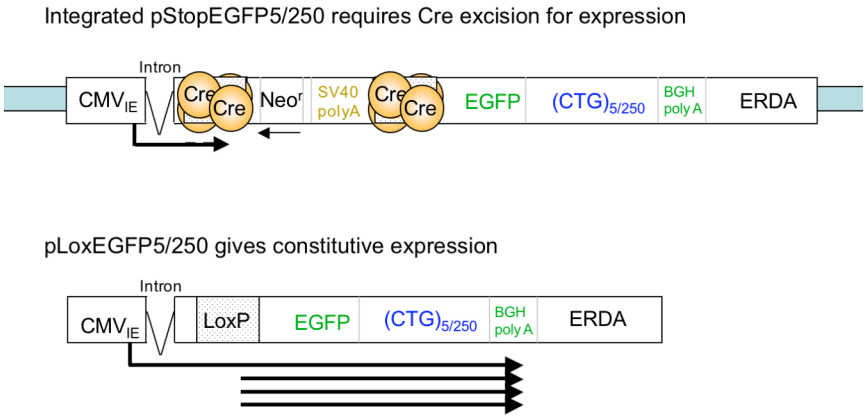


Figure 34 Two different approaches for single plasmid activation. Schematic diagram depicting the differences between two approaches to activation of transcription using a single plasmid transfection. **Top** pStopEGFP5/250 is first integrated into the genome of the chosen cell-line and clones selected by neomycin resistance. *EGFP* transcription is then activated by excision after transfection with pCre. **Bottom** Constitutively expressing pLoxEGFP5/250 is transfected directly into the chosen cell-line.

Floxed constructs were generated in chapter 3 (chapter 3, Figure 11B) by *in vitro* excision of pStopEGFP5 and pStopEGFP250 using recombinant Cre recombinase. During isolation, some constructs were identified with variations in the length of cloned repeats. PCR analysis revealed the arrays to be reduced to approximately 150, 100 and 110 triplet repeats. To confirm that the floxed pLoxEGFP250 products formed nuclear foci, HeLa cells were transfected with floxed products pLoxEGFP5 and pLoxEGFP250. The 100 and 150-repeat floxed constructs were also included in the experiment to discover to what extent shortened arrays would form foci. Transfected cells were subjected to *in situ* hybridisation using Cy3-labelled $(CAG)_{10}$ oligonucleotides. PLoxEGFP5 was included as a negative control for foci formation and no foci were detected. In pLoxEGFP constructs containing 100, 150 and 250 repeats however, foci were detected in the nucleus.

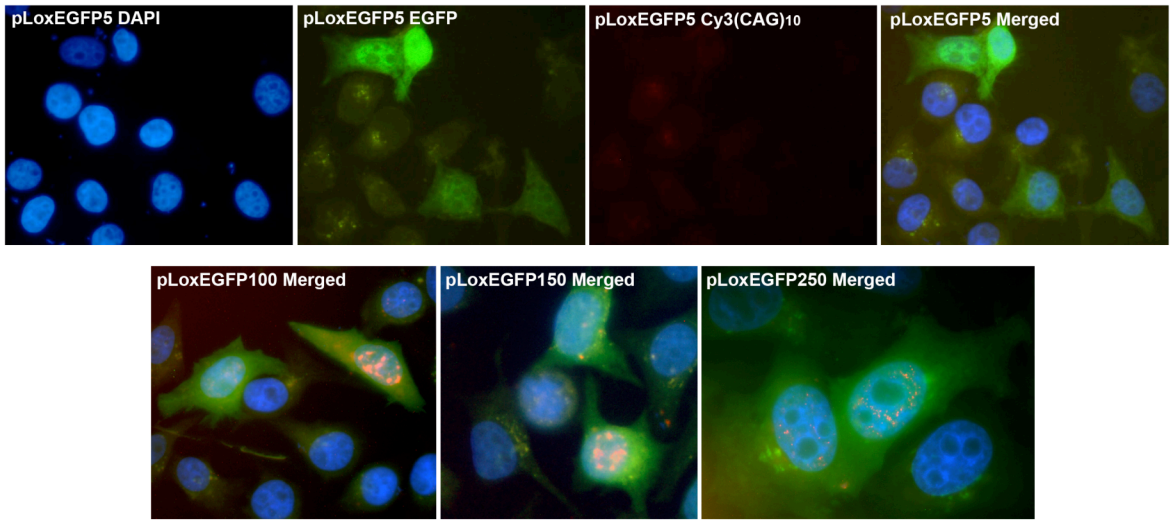


Figure 35 Foci formation with 100, 150 and 250 repeats in HeLa cells The top row of pictures show the three channels DAPI; EGFP and Cy3 separately and merged for the pLoxEGFP5 transfectants. No foci are detected with Cy3-$(CAG)_{10}$. The bottom row shows the three channels DAPI, EGFP and Cy3-$(CAG)_{10}$ merged for pLoxEGFP constructs with 100, 150 and 250 repeats. Foci are apparent within the nucleus.

4.3.4 EGFP and RNA foci

From the experiments conducted so far, it is clear that although the EGFP transcript is retained in the nucleus in the form of foci, enough transcript escapes into the cytoplasm to produce detectable levels of EGFP. However, it is not always the case that EGFP and foci are detected within the same cell. Looking at transfections so far in Cos7; HeLa and 3T3 cells, on occasion EGFP positive cells were devoid of foci, and in contrast some foci positive cells showed no green fluorescence (Figure 36). The occurrence of EGFP-positive foci-negative cells can be explained by the heterogeneous nature of repeat length in any one plasmid preparation (discussed in chapter 3, figure 15); cells transfected with small numbers of repeats would not be expected to form foci. Those cells with foci but no EGFP fluorescence are harder to explain, but foci formation must be a dynamic process: Repeats must be present before MBNL can bind. During a transient transfection, we are using a population of cells in different stages of the cell cycle with transfection complexes present over a period of 24 hours; not every complex will enter the cell and the contents transcribed at exactly the same moment. If a cell is transfected later rather than sooner within this timeframe, and if MBNL binds as the transcript is generated, in these particular cells the RNA binding protein would not be limiting, so all transcripts could be retained within the nucleus resulting in no EGFP translation. A time-course correlation between EGFP fluorescence and the presence of foci is investigated in chapter 5, section 3.1.1, and

is found to be most consistent at 48hrs, whereas these samples are processed at 24 hours post transfection, which could further affect the correlation.

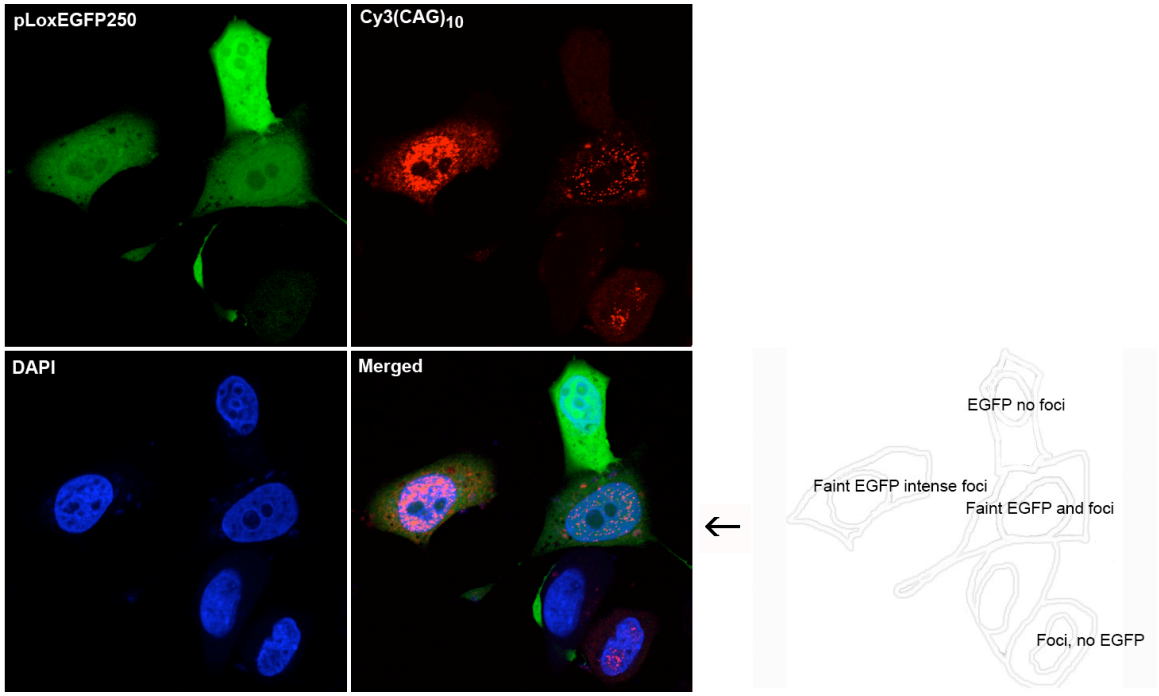


Figure 36 EGFP positive cells do not always exhibit foci Fluorescent micrograph illustrating the lack of correlation between EGFP fluorescence and the presence of foci in HeLa cells transfected with pLoxEGFP250

In our model, expansions of 100 repeats and above form nuclear foci. It is possible that in cells harbouring the 5 repeat construct, the RNA becomes aggregated within the nucleus but is not detected using the Cy3-$(CAG)_{10}$ oligonucleotide because there are 50 fold fewer target repeats: 5 vs. 250. To address this, a Cy3-GFP oligonucleotide was synthesised. The body of the two constructs is identical and so an oligonucleotide directed to the EGFP portion should have equivalent amounts of target RNA. Any differences in the pattern of staining should be due to the location and density of the target rather than absolute number of repeats. HeLa cells were transiently transfected with pLoxEGFP5 or pLoxEGFP250 construct and after 24 hours *in situ* hybridisation was performed with Cy3-GFP, or Cy3-$(CAG)_{10}$ oligonucleotides as a positive control for the presence of foci. Unfortunately the Confocal 488nm laser was out of alignment and awaiting maintenance, rendering the EGFP signal more faint and grainy than usual, but still useful to identify EGFP positive cells, the Cy3 and DAPI channels were unaffected. The Cy3-$(CAG)_{10}$ control samples showed the expected pattern of staining with nuclear foci apparent in the pLoxEGFP250 transfectants, and no staining in the pLoxEGFP5 transfectants (Figure 37). Of the Cy3-

GFP samples, no foci or general staining could be detected in the pLoxEGFP 5 transfectants. In the EGFP positive portion of the pLoxEGFP 250 transfectants some cells, but not all exhibited foci-like staining in the nucleus, and many cells showed bright staining in the cytoplasm (Figure 37). Throughout this project *in situ* hybridisation to pLoxEGFP 250 transfectants with Cy3-$(CAG)_{10}$ does highlight repeat RNA within the cytoplasm which peaks at 16 hours post transfection in 3T3 cells (see Chapter 5.3.1.1 'the dynamics of foci formation') but not as seen here with Cy3-GFP. In this experiment few cells showed foci, most staining was found in the cytoplasm, and a significant number of cells exhibited EGFP fluorescence with no staining, something not seen as often with Cy3-$(CAG)_{10}$ probes. If the cytoplasmic staining is seen in EGFP positive cells in the pLoxEGFP250 transfectants, it is strange then that no cytoplasmic staining is seen in cells transfected with pLoxEGFP5. To conclude from this experiment, either the hybridisation itself failed, which is unlikely since the control Cy3$(CAG)_{10}$ samples showed the expected results, or the Cy3-GFP is in fact not sensitive enough to detect a single copy target without amplification as afforded by a 50 fold increase in repeat number (250 *vs.* 5 repeats). If foci were formed with arrays of 5 repeats, they would not be detected by Cy3-$(CAG)_{10}$ fluorescent *in situ* hybridisation. Taneja *et al.* looked extensively at the distribution of the *DMPK* transcript using a mixture of thirteen different 40-45bp probes placed throughout the transcript, in normal and DM1 patient cells. They observed focus formation only in DM1 patient cells and not in normal cells (Taneja *et al.*, 1995). This conclusion has since been confirmed using riboprobes (Houseley *et al.*, 2005), which confer more sensitive detection of target nucleic acid, since they are single-stranded and have a higher specific activity. Foci detected with $(CAG)_{10}$ oligonucleotides therefore are a feature of expanded repeat expression and are specific to these cells.

4.3.5 RNA binding proteins

The most recently published literature suggests that the missplicing seen in myotonic dystrophy is caused by a shift in equilibrium between two opposing splicing regulators with *trans* dominant effects on specific pre-mRNA targets; primarily CUG-BP1 and MBNL1 RNA-binding proteins (de Haro *et al.*, 2006; Kanadia *et al.*, 2006; Lin *et al.*, 2006; Paul *et al.*, 2006).

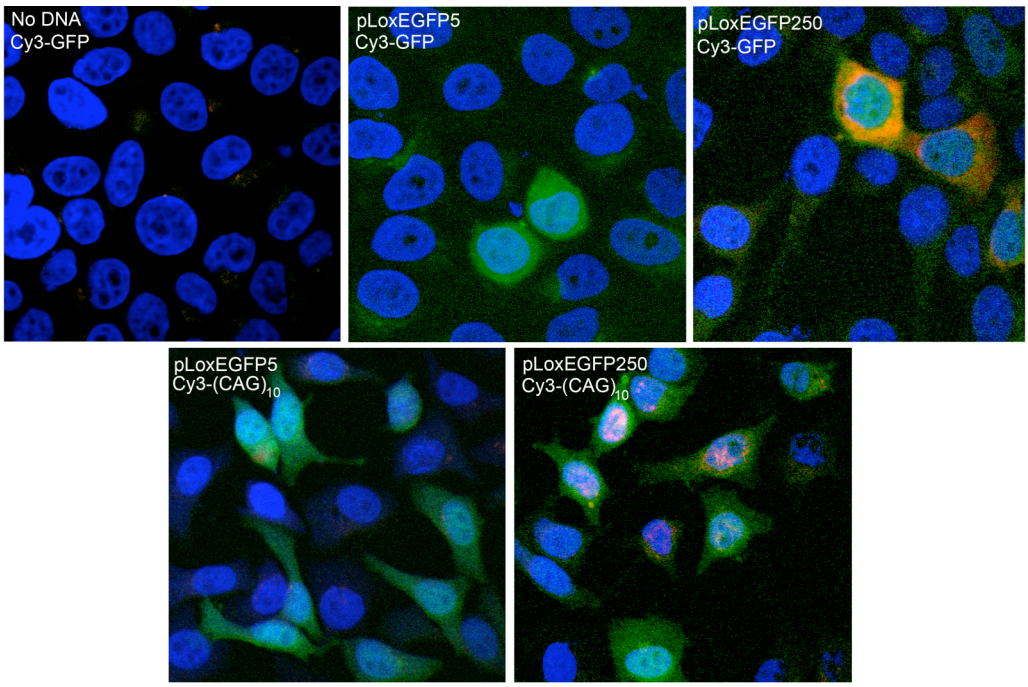


Figure 37 Foci detection using Cy3-GFP *in situ* hybridisation. Merged 3 channel fluorescent micrographs (EGFP; DAPI and Cy3) showing the failure of Cy3-GFP oligonucleotides to detect foci in control pLoxEGFP250 transfected HeLa cells, as well as the test pLoxEGFP5 transfected cells. Foci can clearly be seen in the control pLoxEGFP250 transfected HeLa cells using Cy3$(CAG)_{10}$. Note that the green colouration originates from construct EGFP and not the fluorescent probe GFP, which is red (Cy3).

4.3.5.1 CUG-BP1

CUG-BP1 is a member of the CELF (CUG-BP and ETR-3 like factor) family of binding proteins. These proteins regulate the splicing of target mRNAs during development. In myotonic dystrophy patient cells, it had been noted that the concentration and activity of CUG-BP1 increased within the nucleus (Roberts *et al.*, 1997; Timchenko *et al.*, 2001). This led Philips *et al.* to look for CUG binding sites. They noticed CUG binding sites within muscle specific enhancers of the chicken (*cardiac troponin T*) *cTnT* gene, and went on to confirm it to be alternatively regulated during development by CUG-BP1, and more importantly to be mis-spliced in adult DM1 patients (Philips *et al.*, 1998). Since this first data was published a plethora of CUG-BP1 target genes, some which can be directly related to symptoms have been discovered to be mis-spliced in DM1, such as *chloride channel 1* causing the hallmark myotonia (Cooper *et al.*, 2001; Mankodi *et al.*, 2002) and the *insulin receptor*, insulin resistance (Savkur *et al.*, 2001).

Therefore in order to model this dynamic antagonism in the chosen cell-lines, it is important that CUG-BP1 is present. HeLa and 3T3 cells were transfected with pLoxEGFP5; pLoxEGFP250; pLoxEGFP (3T3 cells only) or transfection reagent alone and harvested after 24 hours. HeLa transfectants were lysed and washed before the nuclear

proteins were extracted. Total protein was extracted from the 3T3 transfectants. Each protein sample (50µg) was separated by SDS-PAGE (4-12%) before transfer and immobilisation onto a membrane. Protein bands were detected using α-CUG-BP1 hybridisation followed by HRP-conjugate-chemiluminescent detection. CUG binding protein was detected in 3T3 extracts and the nuclei of HeLa cells (Figure 38). The levels of CUG-BP1 between the three HeLa nuclear samples appeared equivalent, although because no protein loading control was used it is impossible to confirm this.

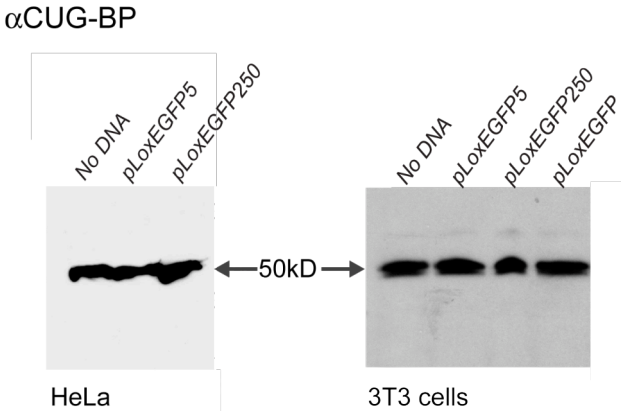


Figure 38 CUG-BP1 is expressed in HeLa and 3T3 cells 50µg nuclear extracts (HeLa) whole cell (3T3). The doublet seen in 3T3 cells is likely to consist of hyper and hypo-phosphorylated forms of CUG-BP1 (Roberts *et al.*, 1997).

Anti-CUG-BP1 analysis was repeated in HeLa cells to study the levels of the binding protein in separated cytoplasmic and nuclear fractions. Although in the transient transfection not every cell expresses the repeat (routinely in our hands, ~65% express the EGFP reporter as visualised by fluorescent microscopy), it may still be possible to detect an increase in the levels of nuclear CUG-BP1 caused by the expression of 250 CUG repeats.

HeLa cells were transfected with pLoxEGFP5; pLoxEGFP250 or transfection reagent alone and harvested after 24 hours. Nuclear and cytoplasmic fractions were prepared, and 50µg of each protein sample separated by SDS-PAGE (4-12%), before transfer and immobilisation onto a membrane. Protein bands were detected using α-CUG-BP1 hybridisation followed by HRP-conjugate-chemiluminescent detection (Figure 39). CUG-BP1 was detected in the nucleus of HeLa cells at much higher levels than in the cytoplasm. But there was no difference in CUG-BP levels between samples.

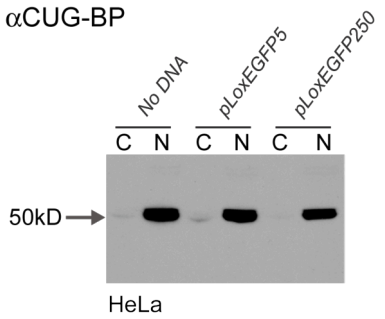


Figure 39 CUG-BP1 in HeLa nuclear and cytoplasmic fractions. CUG-BP1 is present at much higher levels in the nucleus compared to the cytoplasm of HeLa cells.

4.3.5.2 MBNL

MBNL is recruited to foci of mutant RNA within the nucleus (Miller *et al.*, 2000; Fardaei *et al.*, 2001; Fardaei *et al.*, 2002), and works in opposition to the splicing regulator CUG-BP1. Using mouse models that over-express CUG-BP1 (Ho *et al.*, 2005), or are nullizygous for MBNL1 protein (Kanadia *et al.*, 2003), it has been shown that alteration of the balance between these two proteins leads to missplicing equivalent to that seen in myotonic dystrophy. In addition, chloride channel defects resulting from the over-expression of CUG-BP1 in the mouse, have been reversed by increasing expression of MBNL1 (Kanadia *et al.*, 2006). The current model of pathogenesis suggests that in myotonic dystrophy, MBNL depletion by foci formation may lead to the shift in MBNL1-CUG-BP1 equilibrium, resulting in the splicing of developmentally inappropriate isoforms.

The presence of MBNL in cultured cells was investigated by immunocytochemistry using αMBNL1 polyclonal antibody (αEXP42, a gift from Maurice Swanson, Miller *et al.*, 2000). Cells were fixed during exponential growth and the antibody hybridised overnight. Control cells were incubated without αMBNL1. Antibody binding was detected using an anti-rabbit AMCA conjugated secondary antibody followed by fluorescence microscopy. Control cells with no αMBNL1 showed slight background secondary antibody staining throughout the cytoplasm and the nucleus, but staining in the αMBNL1 sample was stronger, and primarily in the nucleus (Figure 40).

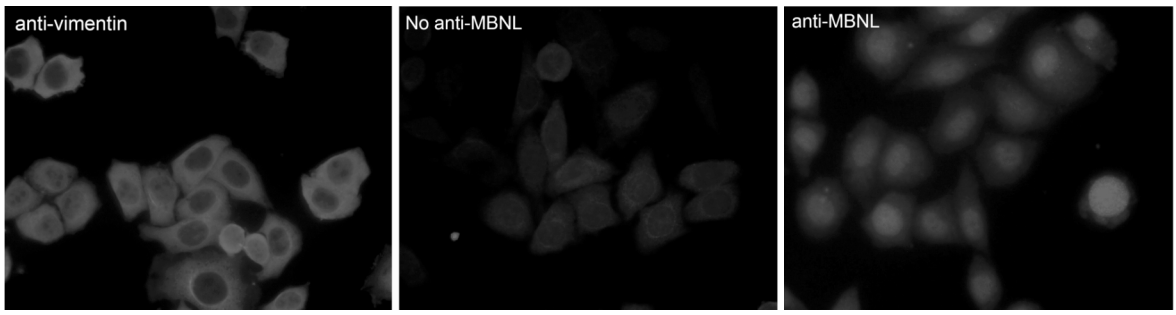


Figure 40 Nuclear MBNL staining in HeLa cells. Fluorescent micrographs showing MBNL staining primarily in the nucleus of HeLa cells. Anti-vimentin was used as a positive control for the immunocytochemistry procedure.

Overall the amount of fluorescence is lower than expected. Since MBNL is most abundant in the nucleus this could be due to difficulty of antibody entry through the nuclear membrane, or an effect of the fixation method used. To take a different approach western transfer and hybridisation was used. HeLa cells were transfected with pLoxEGFP5; pLoxEGFP250 or transfection reagent alone and harvested after 24 hours. Nuclear and cytoplasmic fractions were collected. At the same time, total protein was extracted from 3T3 cells to establish the presence of MBNL in this cell-line. Protein samples were separated by SDS-PAGE (4-12%) before membrane transfer. Protein bands were detected using α-MBNL monoclonal (MCA-MBNL, Encor Biotechnology Inc), followed by HRP-conjugate-chemiluminescent detection. Preliminary results indicated that MBNL was present (Figure 41). Ideally the hybridisation should have been optimised to increase the level of signal, which was faint. No signal was obtained using 3T3 cell extracts (data not shown). This was not an unexpected result since the HeLa cell chemiluminescent signal was faint; even though the extracts were nuclear indicating that the hybridisation was possibly at fault.

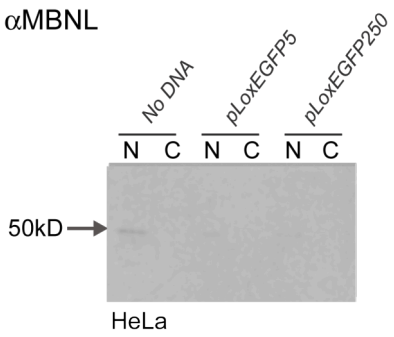


Figure 41 Western showing MBNL in HeLa nuclear and cytoplasmic fractions. MBNL was detected in the nuclear fraction.

4.3.5.3 Co-localisation between MBNL and foci

MBNL1, MBNL2 and MBNL3 have been shown to co-localise with nuclear foci of expanded repeats in DM1 and DM2 patient cells (Fardaei *et al.*, 2002). To determine whether MBNL was recruited to RNA foci in this system, double labelling was performed. In order to minimise antibody usage the cell-culture surface area was scaled down compared to the previous ICC, by growing HeLa cells in 8-well 81mm^2 chamber slides. The cells were transfected 24 hours after plating, with pLoxEGFP250 or pLoxEGFP5. The following day, subsets of each transfection were either processed singly for Cy3-$(CAG)_{10}$ *in situ* hybridisation or αMBNL1 (αEXP42) immunocytochemistry as positive controls for each procedure, or by double labelling with αMBNL1 immunocytochemistry followed by Cy3-$(CAG)_{10}$ *in situ* hybridisation. In the pLoxEGFP250 positive controls, foci were detected in samples with Cy3-$(CAG)_{10}$ *in situ* hybridisation, but no fluorescence was seen using αMBNL1 immunocytochemistry (data not shown). In the double-labelled pLoxEGFP250 sample, fluorescent microscopy revealed nuclear foci staining by Cy3-$(CAG)_{10}$ in the pLoxEGFP250 sample, but no MBNL staining by αMBNL1/AMCA immunocytochemistry (Figure 42). No foci were seen in any of the pLoxEGFP5 samples (data not shown).

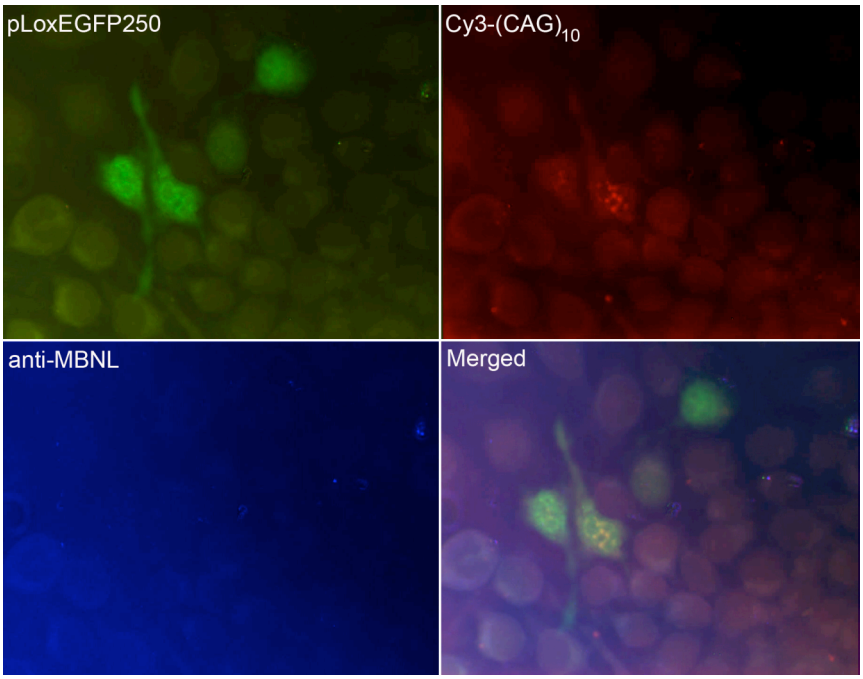


Figure 42 Failed attempt to visualise MBNL in foci Fluorescent micrographs. HeLa cells transfected with pLoxEGFP250 followed by αMBNL/AMCA ICC and Cy3-$(CAG)_{10}$ double labelling. Foci can be identified in the nucleus of EGFP positive cells, but no staining is seen using αMBNL, and therefore no co-localisation.

The experiment was repeated several times unsuccessfully: Cy3 positive foci did not show co-localised (or otherwise) staining with αMBNL1/AMCA. Another researcher had also encountered difficulties with the procedure (Houseley, personal communication). Antibody hybridisation appears to have failed since there is no general nuclear staining as seen in the HeLa cells during the ICC test (Figure 40). There could be a number of technical reasons for this. The antibody may have bound, but become disassociated during the *in situ* portion of the experiment -caused by formamide in the hybridisation buffer. Also, the foci do not appear as sharp when compared to *in situ* hybridisation done in a single labelling experiment (Figure 32). This could be due to partial degradation of RNA within foci during the antibody staining, from RNAses within the polyclonal antibody serum. Degradation of RNA in this way could also lead to the release of bound MBNL proteins. Some antibodies are not suited to ICC, and require high concentrations compared to western blots, but the antibody aliquot was insufficient for use at higher concentrations.

There were obstacles that hindered analysis making it unclear as to whether endogenous MBNL does not co-localise with foci or if the double labelling procedure itself is at fault. The experimental design was flawed due to the absence of a positive control for the antibody staining of foci within the nucleus. It was also observed that the rate of transfection was reduced when using chamber slides, perhaps due to the formation of a deep meniscus. This caused cells to settle around the edge, and may also have interfered with complex dispersion lowering the number of target cells for study. The choice of secondary fluorophore, AMCA, is not ideal since the fluorescent signal from the lower, blue end of the spectrum is not as strong as the middle range, but since EGFP utilises the middle range and Cy3 from the far end, there was no other choice using the microscope equipment we had available. Far-red fluorophores such as Cy5 are too faint under fluorescent microscopy. Dyes with similar emission frequencies would now be a viable option since the confocal we have can visualise infrared, and can also distinguish between different emission spectra *e.g.* EGFP *vs.* FITC.

To address some of the issues with the experimental design, a plasmid expressing MBNL1/GFP fusion protein under the control of the CMV promoter (a kind gift from Marion Hamshere (Fardaei *et al.*, 2001)) was used to help visualise endogenous MBNL in cells expressing expanded CUG arrays. Although the use of the same reporter protein in two different constructs in the same transfection was likely to confuse analysis, it could resolve some doubts about the procedure. From the experiments performed so far, it was not clear whether endogenous MBNL does not co-localise with foci or if the double

labelling procedure itself was at fault. In our experiments pLoxEGFP250 constructs did not produce EGFP (*i.e.* green) foci. If MBNL1/GFP fusion protein does produce EGFP foci in the presence of pLoxEGFP250, these could be used as a positive marker for α-MBNL1 immunocytochemistry to establish whether the double labelling procedure is working. It would also confirm the integrity of the α-MBNL1 used in the previous experiments.

HeLa cells were seeded into chamber slides and transfected singly with MBNL1/GFP plasmid, or co-transfected with MBNL1/GFP plasmid + pLoxEGFP250 or MBNL1/GFP plasmid + pLoxEGFP5. After 24 hours, immunocytochemistry was performed with αMBNL1/AMCA followed by *in situ* hybridisation with Cy3-$(CAG)_{10}$. No foci were detected in the MBNL1/GFP sample (Figure 43A) or the co-transfected MBNL1/GFP plasmid + pLoxEGFP5 (data not shown). Cy3 positive foci were detected in the co-transfected MBNL1/GFP plasmid + pLoxEGFP250 sample which co-localised both to EGFP foci and αMBNL1/AMCA (Figure 43A). This suggests that MBNL1/GFP fusion protein is recruited to the RNA foci. It also confirms the integrity of the MBNL1 antibody. It is strange that the anti-MBNL staining in the co-transfected 250 repeat cells is perinuclear in this particular cell (Figure 43A), and that no Cy3 foci-positive cells were seen showing co-localised anti-MBNL staining within the nucleus. In the MBNL1/GFP control there is less general fluorescence within the nucleus using the antibody/AMCA combination compared to EGFP fluorescence (Figure 43A), which suggests that entry of the antibody was hindered, only detecting a small proportion of the MBNL present. However, this result cannot be representative, in this transfection only a few EGFP positive cells per well were observed, perhaps because the transfection was inefficient due to the use of chamber slides. Also, each EGFP positive cell is not guaranteed to include both plasmids resulting in very few cells of interest. So because the small scale was causing a problem, and it looked as if EGFP from MBNL1/GFP may be co-localising with foci, the transfections were repeated on a larger scale using coverslips, but without ICC.

HeLa cells were transfected singly with MBNL1/GFP plasmid, pLoxEGFP250 or pLoxEGFP5. Co transfections were performed with MBNL1/GFP plasmid and pLoxEGFP250 or MBNL1/GFP plasmid and pLoxEGFP5. The transfectants were then subjected to *in situ* hybridisation with Cy3-$(CAG)_{10}$. All cells on the coverslip were assessed. Neither Cy3 foci nor EGFP foci were detected in MBNL1/GFP; pLoxEGFP5 or MBNL1/GFP plasmid and pLoxEGFP5 transfections (data not shown). Cy3 foci were detected in pLoxEGFP250, and MBNL1/GFP plasmid and pLoxEGFP250 co-transfections, indicating the presence of RNA foci (Figure 43B+C). EGFP foci were only

seen in the MBNL1/GFP plasmid + pLoxEGFP250 co-transfections indicating that the presence of both plasmids is required in order to form EGFP foci. Where EGFP foci are present, the positions co-localise with the Cy3-labelled RNA foci in the nucleus (Figure 43B Merged), indicating that both MBNL and CUG repeats are present within the same structure. It is difficult to form a conclusion with absolute certainty however because of the use of the same reporter protein to locate both MBNL and CUG-repeat expression. Ideally the experiment should be repeated, substituting an alternative fluorescent protein fusion such as EBFP (blue), or ECFP (cyan) for EGFP in the MBNL construct.

It was mentioned earlier, that although the EGFP transcript is retained in the nucleus in the form of foci, clearly enough transcript escapes into the cytoplasm to produce detectable levels of EGFP. The reason for this is not known. It could be that the DNA sequence surrounding the repeat region within the construct may be important. The woodchuck post-transcriptional element for instance has been shown to induce nuclear export of mutant DM1 transcripts when placed 3' to the repeats (Mastroyiannopoulos *et al.*, 2005). This escape could also be explained if MBNL is required to form foci, and is the limiting factor in their formation. Excess EGFP transcripts produced from a strong promoter may deplete MBNL stores leaving sufficient free of bound MBNL, to be exported to the cytoplasm to generate high enough levels of EGFP for visualisation. There are two pieces of evidence to support this: Firstly it is borne out by the difference in fluorescence levels between constructs containing 5 and 250 repeats: The 250 repeat transfectants exhibit lower fluorescence when transfected with the equivalent amount of DNA (chapter 5, Figure 45), probably because some of the expanded transcripts are retained within the nucleus, bound to MBNL. Secondly, during co-transfection with pLoxEGFP250 and MBNL1/GFP, preliminary results suggest that the foci are larger than those formed from pLoxEGFP250 and endogenous MBNL (Figure 43B white *vs.* pink arrows). Since both constructs utilise EGFP as a reporter, in a co-transfection it is only possible to hint at which cells contain which plasmids. The expression pattern of EGFP from MBNL1/GFP is mostly concentrated within the nucleus (Figure 43A), whereas that from pLoxEGFP250 is evenly spread throughout the cell (Figure 43C). Using this information it is possible to map the cells (Figure 43D). Foci from this experiment were studied in greater detail using animated z-stacks taken using confocal microscopy. Here MBNL1/GFP foci are large suggesting that most MBNL is sequestered into foci, since increasing the amount of MBNL increases the size of the foci (Figure 44). It was interesting to note that the foci were not present within nucleoli, which appear as voids in DAPI-stained nuclei.

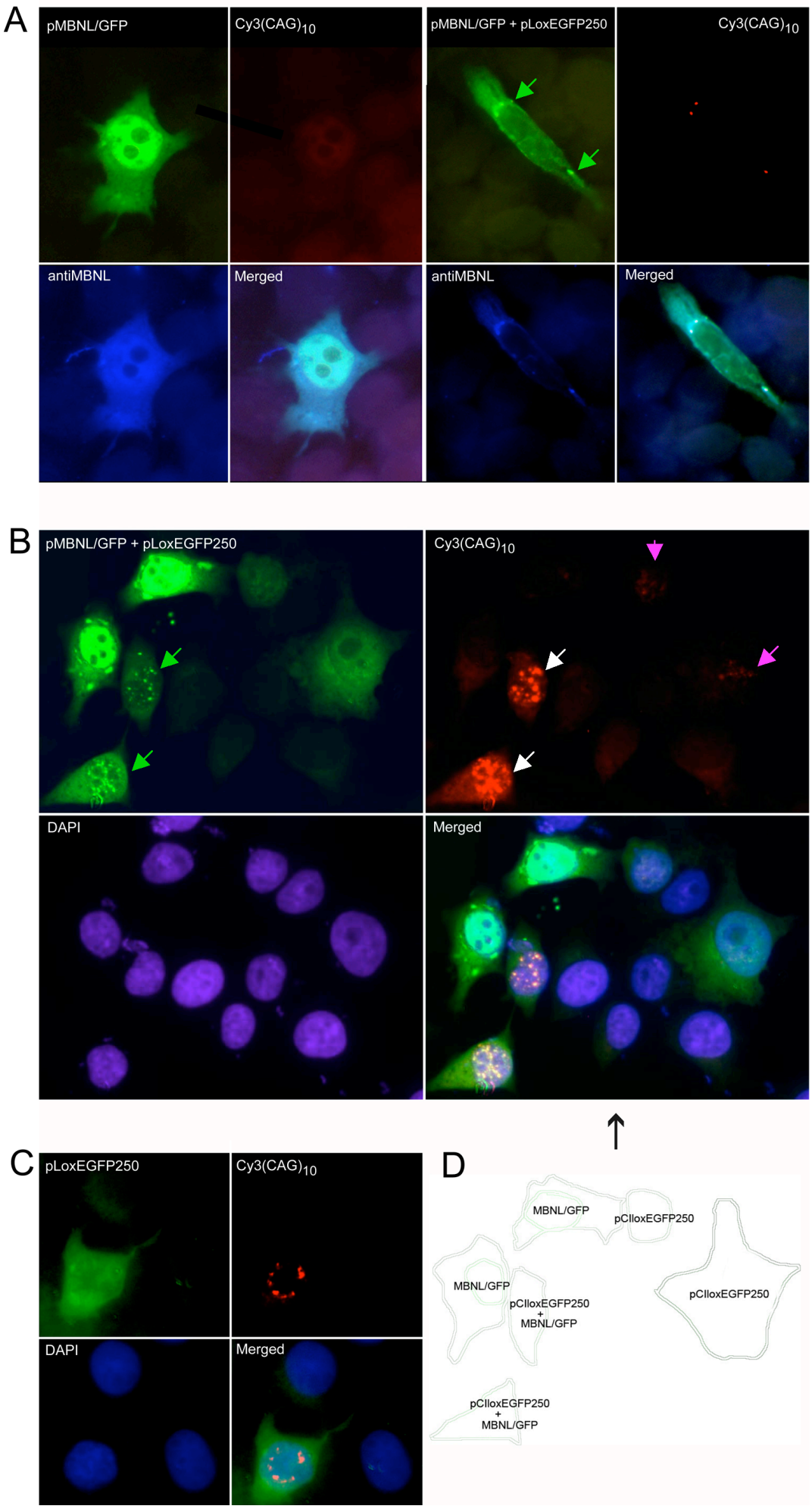

A
pMBNL/GFP
Cy3(CAG)10
pMBNL/GFP + pLoxEGFP250
Cy3(CAG)10
antiMBNL
Merged
antiMBNL
Merged
B
pMBNL/GFP + pLoxEGFP250
Cy3(CAG)10
DAPI
Merged
C
pLoxEGFP250
Cy3(CAG)10
DAPI
Merged
D
MBNL/GFP
pCIloxEGFP250
MBNL/GFP
pCIloxEGFP250
+
MBNL/GFP
pCIloxEGFP250
pCIloxEGFP250
+
MBNL/GFP

Figure 43 Fluorescent micrograph showing co-localisation of MBNL1/GFP with foci. A MBNL1/GFP and MBNL1/GFP + pLoxEGFP250 transfected HeLa cells with Cy3-$(CAG)_{10}$ *in situ* hybridisation and αMBNL1/AMCA immunocytochemistry. EGFP foci (green arrows) are formed in the presence of both MBNL and repeat plasmids, which co-localise to Cy3-$(CAG)_{10}$ foci within the nucleus. Note that in the left hand panel, EGFP fluorescence is much higher than the blue AMCA fluorescence within the nucleus indicating antibody entry is hindered. **B** MBNL1/GFP + pLoxEGFP250 or pLoxEGFP250 (panel C) transfected HeLa cells with Cy3-$(CAG)_{10}$ *in situ* hybridisation. EGFP foci (green arrows) co-localise to Cy3 foci (white arrows) indicating co-localisation of MBNL1/GFP and expanded array foci. Note that the RNA foci are larger in MBNL-GFP + pLoxEGFP250 (white arrows) compared to pLoxEGFP250 alone (pink arrows) see map D indicating that the size of foci may be dependent on the amount of MBNL present. **D** Schematic map of the probable plasmid content within cells, based on the distribution pattern of EGFP expression comparing MBNL1/GFP alone (*e.g.* in A), with pLoxEGFP250 alone (*e.g.* in C).

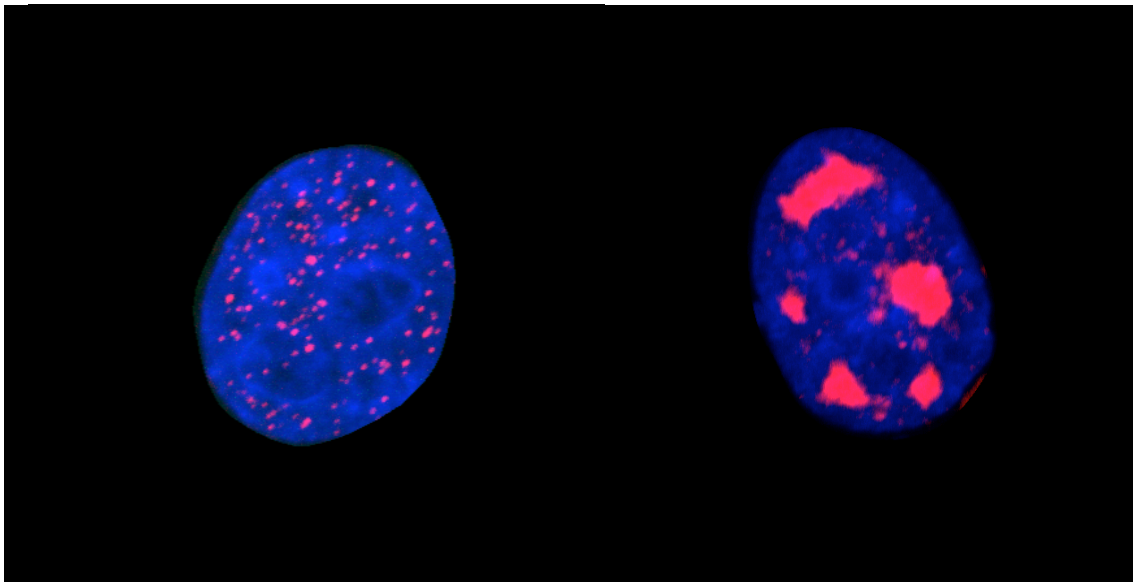


Figure 44 (movie) The effect of MBNL1/GFP on foci size. 3D animated confocal Z-stack. Hela cells transfected with pLoxEGFP250 alone (left) or with pLoxEGFP250 and MBNL1/GFP (right) after *in situ* hybridisation with Cy3-$(CAG)_{10}$. Foci are increased in size and reduced in number in cells expressing pLoxEGFP250 and MBNL1/GFP. DAPI (nucleus) and Cy3 (foci) merged. See attached CD.

4.4 Discussion

We set out to determine whether the constructs generated to create an *in vivo* mouse model of myotonic dystrophy type 1, could be utilised to produce an *in vitro* model of DM1 pathogenesis to study the earliest affected genes. Two approaches were investigated: pCre activation of pStopEGFP5 or pStopEGFP250 stable cell lines in HeLa and Cos7 cells; and direct transfection of HeLa; Cos7 and 3T3 cell-lines with the constitutively expressing pLoxEGFP5 or pLoxEGFP250. It is not clear why the former approach failed as many clones were generated (10 each of pStopEGFP5 and pStopEGFP5), and in different cell-

types (HeLa; Cos7 and *DmtD162* kidney). PCR analysis throughout the construct confirmed its presence within the cell clones. It is possible that transfection of pCre used to activate the stable lines, resulted in over expression of the recombinase in comparison to the low copy number of the integrated transgene. This might have interfered with transcription of the reporter gene, by physically blocking RNA polymerase by continual *loxP* binding. The latter constitutive approach however was successful for reporter expression, and chosen for validation. Nuclear foci are formed in both DM1 and DM2 patient cells and co-localise with MBNL protein (Fardaei *et al.*, 2002), this pathogenic marker is used as a diagnostic feature of myotonic dystrophy (Bonifazi *et al.*, 2006). In our cell-culture system, foci were identified in the nucleus 24 hours post transfection using pLoxEGFP250, but not detected using pLoxEGFP5. Foci co-localised with MBNL1/GFP fusion protein, and endogenous MBNL in perinuclear foci using αMBNL/AMCA and Cy3-$(CAG)_{10}$ double labelling. Interestingly, foci tended to be larger in Cos7 cells, and smaller and more well defined in HeLa and 3T3 cells, and those foci formed in the presence of excess MBNL (in the form of MBNL1/GFP) were larger than foci formed without additional MBNL. These observations were incidental and not substantiated, but could indicate that MBNL protein is limiting in foci formation, possibly preventing transcript exit into the cytoplasm up until that limit is reached.

4.4.1 Cell-lines

Our model is generic in terms of the cell-types affected in myotonic dystrophy. It is not biased by the specific function of a particular tissue and as such may be used to study the genesis of global effects, likely to be a good target for therapeutic intervention, and which may be important in the study of the congenital form of the disease.

The original intention was to use ES cells as the basis of the cell-culture model. Pleuripotent embryonic stem cells can be differentiated into almost any cell lineage (O'Shea, 1999; Nagy *et al.*, 2006), including those tissues characteristically affected in myotonic dystrophy such as skeletal (Dekel *et al.*, 1992) and cardiac muscle (Sachinidis *et al.*, 2003). An inducible model able to differentiate into the tissue of choice would be an invaluable and versatile tool to study and delineate the multisystemic facets of myotonic dystrophy throughout development. Unfortunately, due to questions raised over the pleuripotency of the clones generated in chapter 3, it was not deemed feasible to generate an ES cell-based model at this moment in time, but would be worth doing in the future.

Our model has been tested using a variety of cell-types from different organisms each positive for foci formation. As the project evolved, published research started to indicate that missplicing was fundamental to the disease process. As a result of this Cos7 (green African monkey kidney) cells were rejected for the model due to the poorly characterised genome sequence important for primer design in alternative splicing analysis. Both HeLa (human epithelial) and 3T3 (murine fibroblast) cells however are well-established general laboratory cell-lines, and produce a high rate of transfection with DNA. The genome project recently completed for both human and mouse can be utilised in conjunction with these cell lines to unite previously published data: Murine culture system with mouse model data and the human with patient data. The immediacy using transient constitutive transfection, allows study of the early pathogenesis and also overcomes confounding issues such as somatic expansion, and positional effects seen in other stable *in vivo* systems.

4.4.2 MBNL

At the time these experiments were designed and carried out, the focus of pathogenesis was on CUG-BP1. It wasn't known that the balance of two splicing regulators was pivotal to the pathogenesis of DM. In hindsight, although it has not been established that missplicing in DM is limited to MBNL1 and CUG-BP1 disequilibrium, ideally we need to determine that the levels of MBNL1 and CUG-BP1 in cells chosen for the model are equivalent to those cell-types affected in the disease to enable the same changes in splicing.

It was interesting to note that foci formed with MBNL1/GFP were large. Increasing the amount of MBNL1 available increases their size, suggesting that most MBNL1 is sequestered into foci. Is MBNL1 required for foci formation? Research by Housley *et al.* indicates that, in *Drosophila*, although muscleblind (the MBNL1 orthologue) does co-localise with foci, muscleblind is not required for foci formation. In some cells, induced expression of $(CUG)_{162}$ resulted in the formation of CUG positive foci where muscleblind could not be detected using standard immunohistochemical methods (Houseley *et al.*, 2005). This could be confirmed using existing mouse models to generate MBNL knockout-repeat expressing mice. Lack of foci would indicate a prerequisite of MBNL for foci formation. If this was so, as has been reported by MBNL1 siRNA knockdown experiments in DM1 myoblasts (Dansithong *et al.*, 2005), then MBNL could play a protective role in DM1. It has recently been shown that foci formation is separable from missplicing. Both expanded CAG and CUG repeats formed MBNL-associated nuclear foci, but mis-regulated splicing was only seen with CUG repeat expression (Ho *et al.*, 2005). This questions the

hypothesis that sequestration of MBNL into foci causes the shift of CUG-BP1/MBNL1-regulated splicing toward embryonic variants. More recently, involvement of the RNAi mechanism has been implicated. Long CNG repeat hairpins are a target for dicer, from which it generates gene-silencing siRNAs to that target. In the case of DM1, dicer cleaves the mutant repeat transcript into many short 21nt CUG repeats (Krol *et al.*, 2007). Recently Mahadevan *et al.* created an inducible mouse model where low level expanded repeat expression leads to foci formation, but a lack of phenotypic effects. The control cohort over-expressing 5 CUG repeats does not form nuclear foci but surprisingly exhibited myotonia, heart conduction defects and RNA missplicing (Mahadevan *et al.*, 2006). Perhaps short cytoplasmic CUG repeats are toxic, leading to repeat-associated CUG-BP1 within the cytoplasm, which in turn would raise levels of CUG-BP1 (Timchenko *et al.*, 2001) causing missplicing. It would be interesting therefore to investigate the role of short CUG repeats in the cytoplasm. It may be that MBNL1 is protective up to a limit, until sequestration leads to an irrevocable imbalance between splicing regulators, allowing unimpeded activity of CUG-BP1. It would explain why long repeats are more pathogenic since larger repeats would deplete MBNL1 more rapidly resulting in earlier onset and increased severity. Additionally, foci have been shown to be formed by as little as 57 repeats, lengths that are not pathogenic in people (Amack *et al.*, 1999).

From published research, it is clear that both reduced levels of MBNL, and increased levels of CUG-BP1 lead to aberrant splicing of target genes, but it is also clear that there is much to discover, that the mechanism of pathogenesis is not so straightforward.

5 The effects of expanded CUG-repeat expression on mRNA steady state levels and splicing patterns

5.1 Synopsis

Myotonic dystrophy type 1 is a multisystemic disease in which the missplicing of genes can be directly related to important clinical features, such as the *chloride channel* to myotonia, or the *insulin receptor* to diabetes. Many symptoms nevertheless have yet resisted clear explanation: cataracts, testicular atrophy, heart block, anaesthetic sensitivity and mental retardation for example. Here we have used an *in vitro* cell culture model to discover further genes possibly aberrantly spliced in myotonic dystrophy, to attempt to clarify some of the unexplained symptoms. Affymetrix whole and exon transcript arrays were used to determine differential transcript levels and alternative exon usage between murine fibroblast 3T3 cells transfected with either 5 (pLoxEGFP5) or 250 (pLoxEGFP250) CUG repeats. For the exon arrays 0 (pLoxEGFP) repeats were also compared. Using whole genome arrays, 6 genes were down-regulated and 128 up-regulated. With exon arrays, 58 genes showed alternative exon usage. Six genes were selected for further bioinformatics analysis: *MtmR4*, which has possible neuromuscular involvement; *Kcnk4*, *Narg1*, *Ttyh1* and *Bptf*, potentially related to brain development; and *Cacna1c*, a promising candidate for heart conductance defects and sudden death.

5.2 Background

In myotonic dystrophy, the expanded repeat transcript becomes trapped within nuclear foci. Recruitment and sequestration of the splicing regulator MBNL1 to these foci, is thought to alter the dynamic equilibrium between itself and another splicing regulator CUG-BP1, resulting in the production of developmentally inappropriate splice isoforms. The first gene discovered to be mis-spliced was *cardiac troponin T* (c*TNT*) (Philips *et al.*, 1998), originally identified as a target of CUG-BP1, whose activity and concentration within the cell is altered in DM1 (Timchenko *et al.*, 2001). Cardiac troponin T is the tropomyosin-binding subunit of the troponin complex, regulating muscle contraction in response to alterations in intracellular calcium ion concentration. It is normally expressed in adult heart, and embryonically within skeletal muscle, but in DM1 is also expressed in

adult skeletal muscle (Philips *et al.*, 1998; Ladd *et al.*, 2005). Mutations in the gene are associated with cardiomyopathy (Thierfelder *et al.*, 1994), which features in DM1 but is not always present (Harper, 2001). A clear link as to how missplicing in this gene could result in cardiac conduction system defects however, has not yet been established. It wasn't until splicing of the *chloride channel 1* gene (*CLCN1*) was found to be abnormal, resulting in the loss of chloride channel 1 (CLCN1) function and producing the definitive myotonia, that it was generally believed that aberrant splicing could also be responsible for the other symptoms of DM1 (Charlet *et al.*, 2002; Mankodi *et al.*, 2002). Missplicing of the *CLCN1* gene in DM1 skeletal muscle tissue compared to unaffected adult expression patterns results in the inclusion of exons normally reserved for the embryo, as well as the use of cryptic acceptor sites causing premature chain termination during translation (Charlet *et al.*, 2002). Since then, other missplicing events have been identified, accounting for further phenotypic effects. Exclusion of an adult-specific exon in the insulin receptor transcript has been associated with the specific type of insulin resistant diabetes seen in DM1 (Savkur *et al.*, 2001). The *microtubule-associated protein tau* (*MAPT*) transcript undergoes complex, regulated alternative splicing, giving rise to several mRNA species differentially expressed in the nervous system. In DM1, the proportion of splice variants is disrupted (Sergeant *et al.*, 2001; Leroy *et al.*, 2006). Deletions in this gene have been associated with developmental delay and learning disability (Shaw-Smith *et al.*, 2006), which is seen in congenital DM1. *Myotubularin related protein 1* (*MTMR1*), also aberrantly spliced (Buj-Bello *et al.*, 2002), contains the consensus sequence for the tyrosine phosphatase activity of myotubularin (MTM). Mutations in *MTM* have been identified as being responsible for X-linked myotubular myopathy (Buj-Bello *et al.*, 1999), and mice deficient for the gene develop a generalized and progressive myopathy starting at around 4 weeks of age, with amyotrophy and accumulation of central nuclei in skeletal muscle fibers leading to death at 6-14 weeks (Buj-Bello *et al.*, 2002). There may be a link therefore between the progressive muscle wasting and myopathy, and *MTMR*1 mis-processing in DM1. Aberrant splicing is also witnessed in transcripts of two major proteins of the sarcoplasmic reticulum, the *ryanodine receptor 1* (*RyR1*) and *sarcoplasmic/endoplasmic reticulum Ca2+-ATPase* (*SERCA*) 1 or 2. The foetal variant, ASI(-) of RyR1 which lacks residue 3481-3485, and SERCA1b which differs at the C-terminus, were significantly increased in skeletal muscles from DM1 patients and the HSA^{LR} DM1 transgenic mouse model (Kimura *et al.*, 2005). In addition, a novel variant of *SERCA*2 was significantly decreased in DM1 patients (Kimura *et al.*, 2005). Chemically increased Ca^{2+} in mouse skeletal muscle induced some DM-like symptoms –myotonia, balding and cataracts (Takahashi *et al.*, 1999), so inappropriately spliced *RyR*1 and *SERCA*1 mRNAs might contribute to impaired Ca2+ homeostasis in

DM1 muscle, and increased calcium levels seen in cultured DM1 cells (Benders *et al.*, 1996).

Many missplicing events linked to symptoms have now been identified in myotonic dystrophy, but the list is not exhaustive. There are many symptoms not explained, some of which are more distressing for patients and their families to live with such as somnolence and dementia. Here we utilise the cell-culture model –consisting of murine 3T3 cells expressing either 0, 5 or 250 repeats within the 3'UTR of the EGFP gene (chapter four), to attempt to identify genes mis-spliced early in the pathogenesis of the disease, which are most likely to make good therapeutic targets. The objective is to identify further genes mis-spliced as a direct result of expanded CUG repeat expression, using whole transcript and exon microarray analysis. Exon arrays became available during the course of the experiments, and are the most suitable chips to use to identify changes in exon inclusion and exclusion since probesets span all exons within the transcript. In whole transcript arrays most probsets are situated only at the 3' end of the gene. Using these arrays it was thought possible to identify splice variants due to altered mRNA levels resulting from nonsense mediated decay, a mechanism activated when missplicing results in a premature termination codon situated in an exon other than the 3' terminal exon –as observed in DM1 with the *chloride channel 1* transcript (Nagamitsu *et al.*, 2000; Charlet *et al.*, 2002)– as well as those differing in the 3' termini, although none have yet been reported in DM.

5.3 Microarray analysis

Our model is based on a transient rather than a stable transfection system, in order to identify the earliest pathogenic events. Even though experimental parameters can be adjusted to optimise the rate of transfection, only a proportion of cells will contain the construct. We decided therefore to increase the proportion of construct-positive cells by separating the EGFP positive and non-fluorescent fractions from each transfection using fluorescent activated cell sorting (FACS). For each transfection, this would result in two pools of cells differing only in the expression of *EGFP* and also provide a tight internal negative control to limit technical variation between treated 'EGFP on' and untreated 'EGFP off' fractions. RNA isolated from the 'on' and 'off' pools of each of the pLoxEGFP5 and pLoxEGFP250 transfected 3T3 cells gave four samples which could then be assessed for differential gene expression by microarray hybridisation (Table 10).

Table 10 EGFP 'on' and 'off' fractions used for differential gene expression analysis
Differences between 1 and 3 result from expanded array expression. Although samples 2 and 4 appear to be duplication, they are required to control for the fidelity of cell sorting.

	Sample	Interpretation
1	pLoxEGFP250 'on'	Test sample
2	pLoxEGFP250 'off'	Internal negative control
3	pLoxEGFP5 'on'	Negative test sample
4	pLoxEGFP5 'off'	Internal negative control

5.3.1 Parameter selection

5.3.1.1 Dynamics of foci formation

It was first necessary to establish the optimal time-point for RNA isolation after transfection –a point that would reflect the patient situation where aggregates of the mutant transcript are formed and retained within the nucleus in discrete foci, and at the same time express the EGFP reporter at a sufficient level for cell separation to increase the proportion of cells harbouring the construct. A time course experiment was set up to correlate foci formation and EGFP detection. Murine 3T3 cells were plated into 52 Petri dishes containing coverslips. The next day 13 dishes were each transfected with pLoxEGFP (0 repeats); pLoxEGFP5 (5 repeats) or pLoxEGFP250 (250 repeats) constructs, or reagent alone. At time zero, and then every four hours for a total of 48 hours, a dish for each transfection was removed and the live cells photographed using fluorescence microscopy. The coverslips were then removed, fixed and stored in PBS until all time-points had been gathered. The remaining cells in the dishes after coverslip removal were trypsinised and stored in RNAlater solution, to be used at a later date out with this thesis to study the effect of expanded repeat tracts on the dynamics of gene missplicing. The coverslip samples were then processed in one batch for Cy3-$(CAG)_{10}$ *in situ* hybridisation. Faint EGFP fluorescence was observed at 8 hrs post transfection and peaked in intensity between 20 and 32 hours (Figure 45; data not shown for reagent alone).

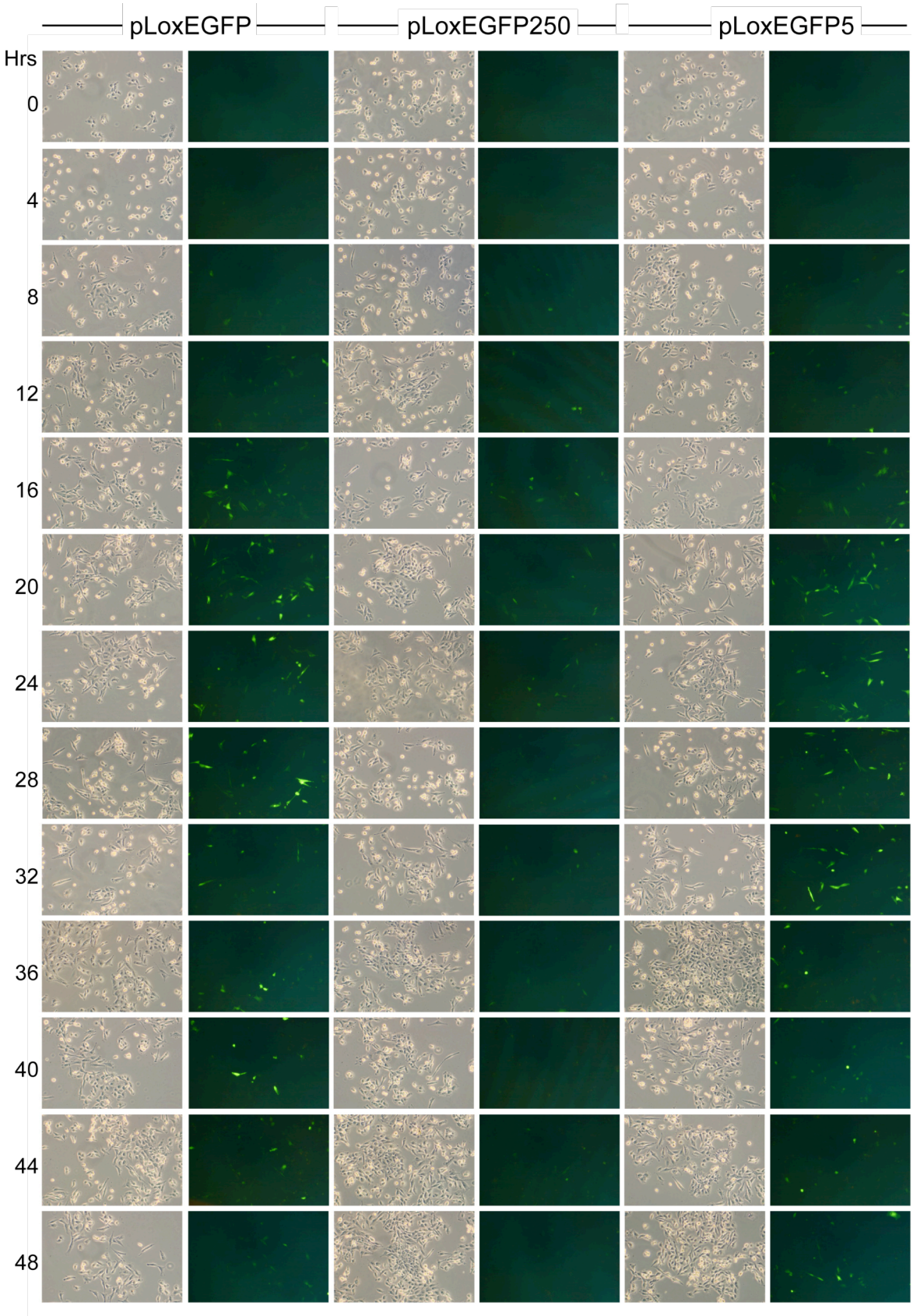
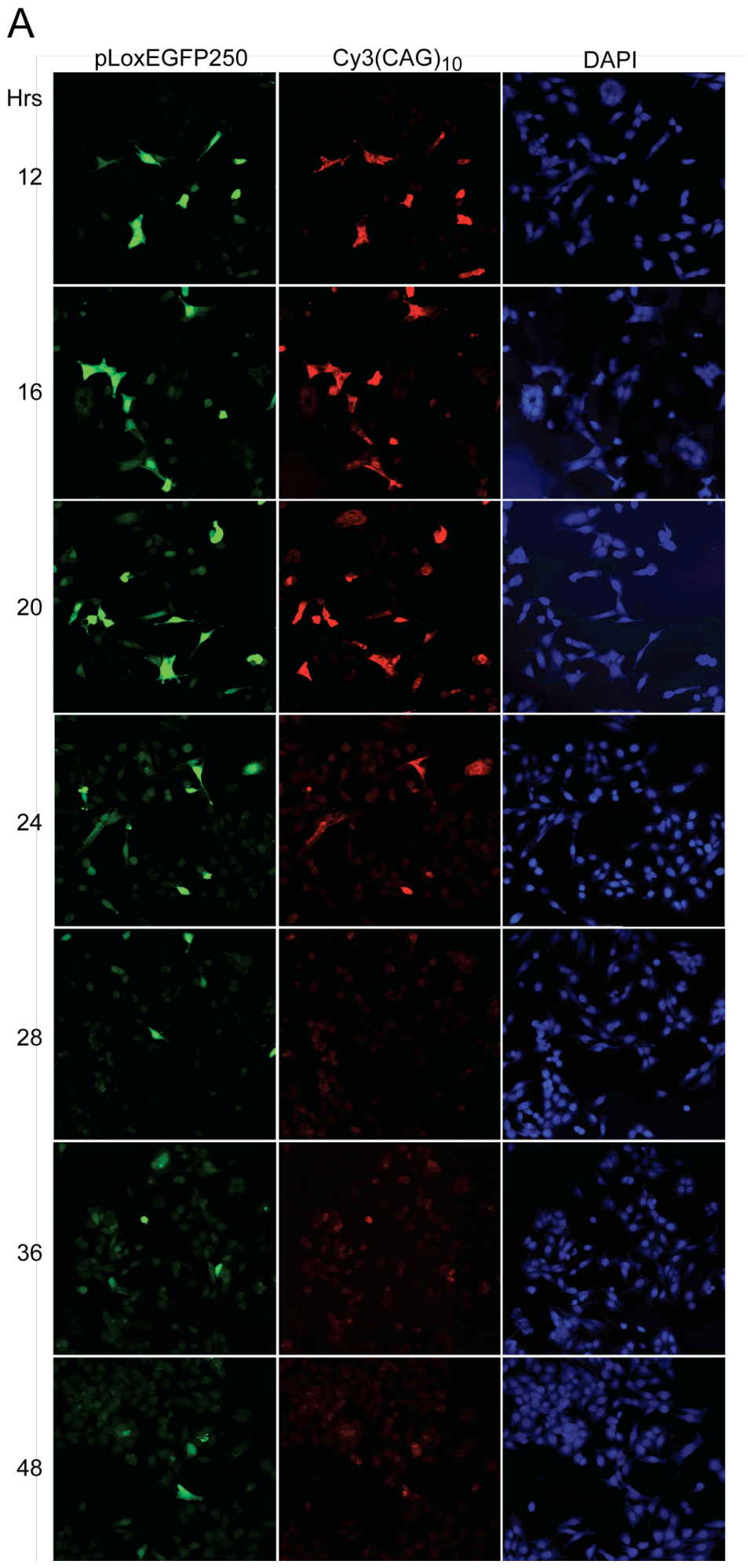


Figure 45 Fluorescent micrographs charting EGFP fluorescence over 48 hrs. Live cell micrographs taken using brightfield and fluorescence microscopy. 3T3 cells were transfected with pLoxEGFP5 or pLoxEGFP250 constructs and samples photographed at 4-hour intervals over 48 hours. EGFP fluorescence can be detected at 8 hours post transfection and peaked between 20 and 32 hours. The intensity of EGFP fluorescence in the pLoxEGFP250 transfected sample did not reach as high a level as the pLoxEGFP negative and pLoxEGFP5 normal control samples, probably because of transcript retention within the nucleus.

No foci or Cy3 staining were detected in pLoxEGFP (0 repeats, negative control), and pLoxEGFP5 (normal control) transfections (data not shown). With construct pLoxEGFP250, Cy3-$(CAG)_{10}$ *in situ* hybridisation illustrates the dynamics of foci formation within the cell over time. RNA was first detected faintly at 8hrs (data not shown). Of notable interest was the high amount of general fluorescence within the cytoplasm as the RNA levels increased and peaked between 16 and 20 hours. Over time the RNA became more discreet within foci, both within the nucleus and the cytoplasm, finally primarily in the nucleus from 28hrs (Figure 46A). It was not clear whether cytoplasmic RNA returned to the nucleus and became concentrated into foci, or if the foci were initially present but masked by the abundance of RNA, later revealed after RNA degradation perhaps, but in a previous parallel experiment using HeLa cells, general staining and some indistinct foci were apparent within the nucleus at 8hrs (Figure 47). It would be interesting to repeat this experiment using 1hr intervals between 6 and 12 hours post transfection, to clarify the presence of foci and pattern of general staining in the cell; whether cytoplasmic as well as nuclear, and whether foci are formed as a primary or secondary event. Cytoplasmic foci have not been reported in the literature, but have been observed by other researchers (L.Timchenco, personal communication) and were apparent in DM1 fibroblast patient cell-lines (a kind gift from J. D. Brook) (Figure 46D). However, it has been proposed that the presence of cytoplasmic foci is a phenotype of some, but not all, patient cell-lines (S. Reddy, personal communication).

The CMV promoter ensured high-level expression of the transcripts, which was confirmed by bright EGFP fluorescence. However, Houseley *et al.* observed bright general staining in cells without foci, expressing *GFP*-tagged CUG expansions using a single-copy *GFP* Riboprobe (Houseley *et al.*, 2005), so it was surprising that we did not see general staining in pLoxEGFP5 cells above background, especially within the cytoplasm (Figure 46B). This however, may have been an artefact due to the stringency of hybridisation since the Cy3-$(CAG)_{10}$ probe (30 nucleotides) would only bind to 15 nucleotides of the $(CAG)_5$ repeat, and could have failed to hybridise, or have been washed away during processing.


A
pLoxEGFP250
Cy3(CAG)10
DAPI
Hrs
12
16
20
24
28
36
48

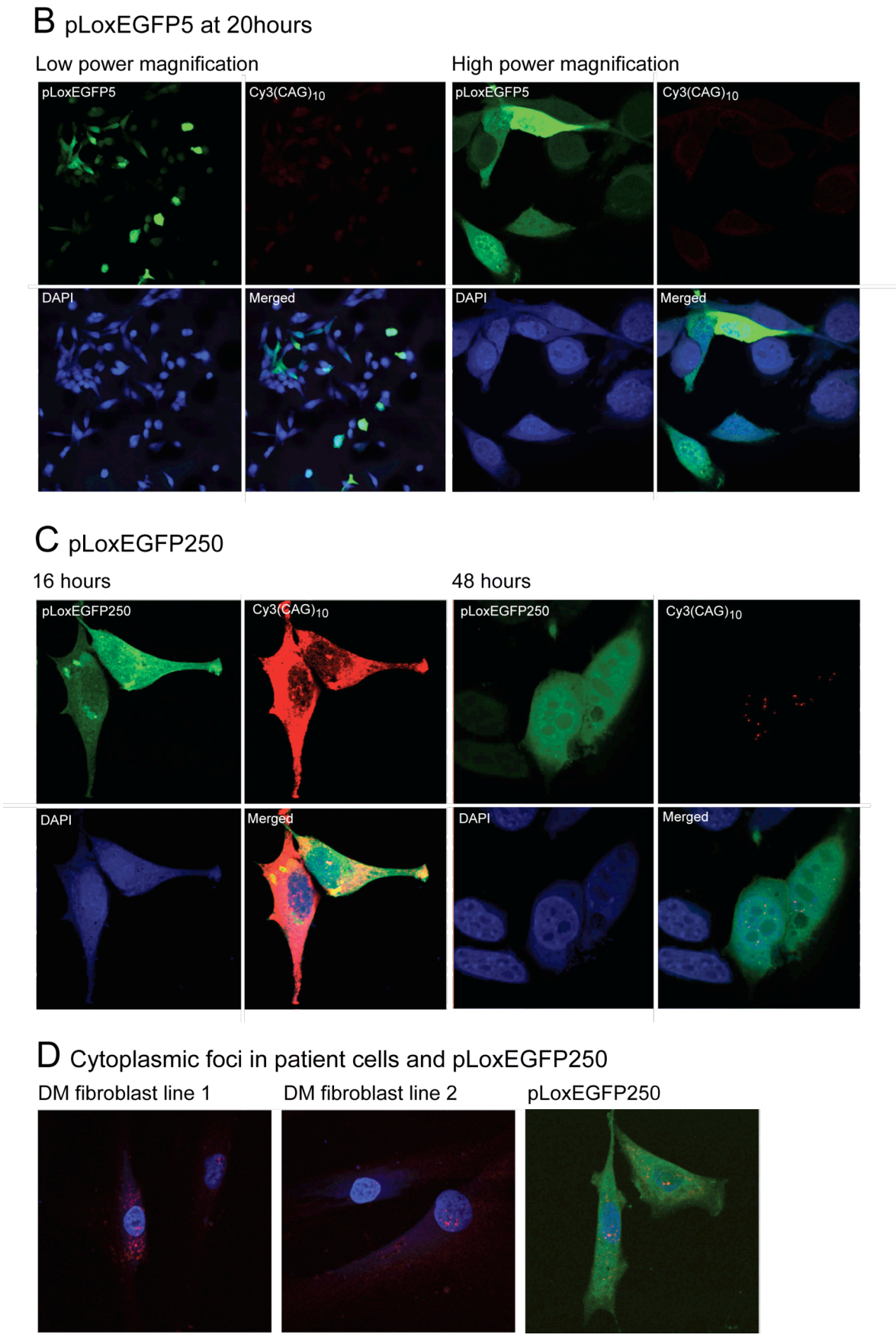


Figure 46 The dynamics of foci formation. Fluorescent micrographs. Cy3-$(CAG)_{10}$ *in situ* hybridisation on 3T3 cells transfected with pLoxEGFP250; pLoxEGFP5 or pLoxEGFP, sampled at 4 hourly intervals. **A.** The presence of EGFP fluorescence and foci formation over time. In 3T3/pLoxEGFP250 cells, EGFP fluorescence can be visualised faintly from 8 hours and peaks between 24 and 32 hours. Cy3-$(CAG)_{10}$ staining is predominantly cytoplasmic as well as nuclear at 16 and 20 hours post transfection, but becomes more discreet over time with foci mainly nuclear

from 28 hours post transfection. **B.** Low (X16) and high (X63) power images of 3T3 cells transfected with pLoxEGFP5 20 hours post transfection, showing that no Cy3-$(CAG)_{10}$ staining is present in the normal control. **C.** High power (X63) images of 3T3/pLoxEGFP250 transfections at 16 hours showing intense Cy3-$(CAG)_{10}$ staining in the cytoplasm. By 48 hours the staining is limited to discrete foci predominantly within the nucleus. **D.** Cy3-$(CAG)_{10}$ staining in patient DM1 fibroblast cell-lines (left and middle) reveals cytoplasmic foci. These have also been observed using pLoxEGFP250 in 3T3 cells (right).

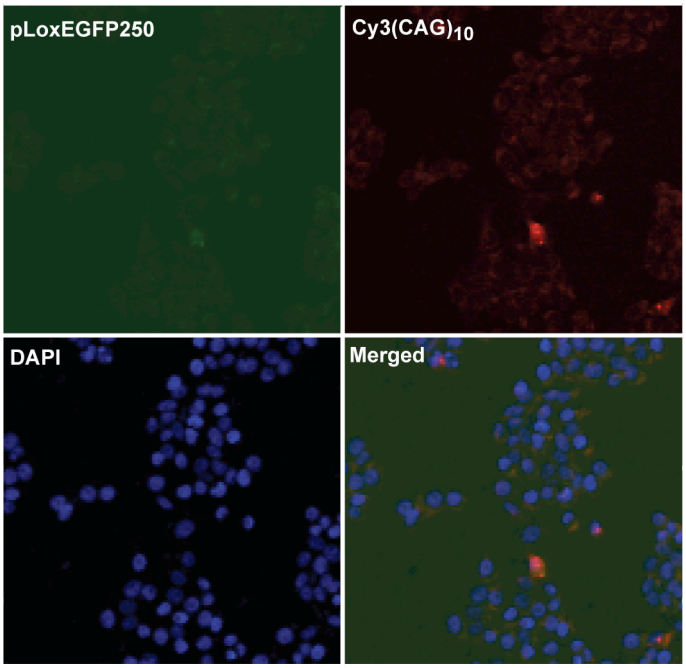


Figure 47 Foci are apparent in HeLa cells at 8hrs post transfection. Fluorescent micrograph. HeLa cells transfected with pLoxEGFP250 show Cy3-$(CAG)_{10}$ staining within the nucleus and diffuse foci at 8 hrs post transfection.

Not all EGFP positive cells transfected with pLoxEGFP250 exhibited foci (Figure 46A). This could have been due to plasmid heterogeneity as previously described (chapter 3), or the confocal plane may not have passed through the focus. Foci are spread throughout the nucleoplasm avoiding the nucleoli (chapter 4 Figure 44 3D movies). Therefore, when taking a representative image, the focal plane will not pass through all foci, and could miss all of them. To assess the presence of nuclear and cytoplasmic foci within EGFP positive cells, foci were identified in individual cells by focusing through the whole thickness of the cell dorsally to ventrally and the presence noted (Figure 48B). All EGFP positive cells on each slide were counted. This resulted in a variable number (14-54) of cells assessed at each time point, but was enough to give an indication of foci and EGFP correlation.

The 48hr time-point was selected to prepare RNA for microarray analysis since the Cy3-$(CAG)_{10}$ staining pattern most closely reflected the situation in patient cells, where RNA is concentrated mostly into a few larger nuclear foci (Figure 46A and C).

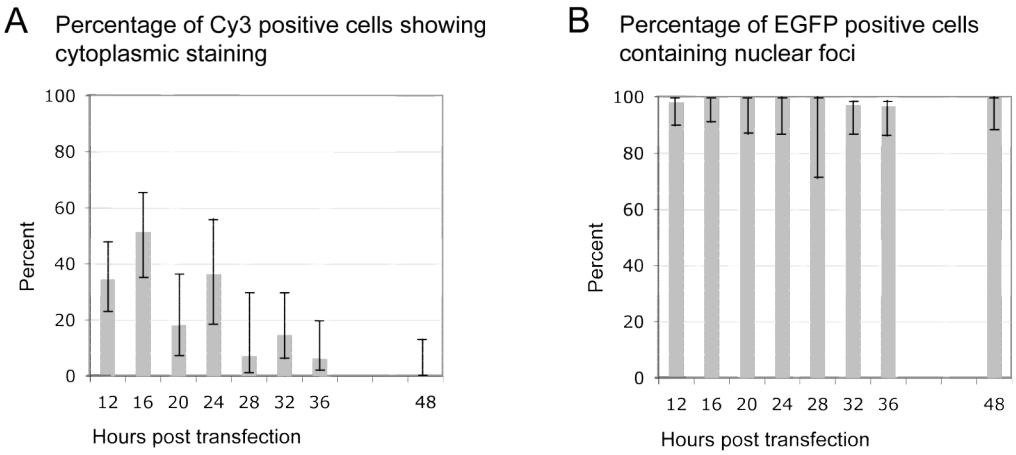


Figure 48 Cytoplasmic CUG repeats and the presence of nuclear foci within EGFP positive cells charted over time post transfection. A Bright general cytoplasmic staining and cells with defined cytoplasmic foci were counted as positive, in cells positive for nuclear foci. **B** Greater than 97 percent of EGFP positive cells contained nuclear foci at each time interval. Error bars show the 95% confidence interval. Time points 0, 4, 8, 40 and 44 hrs were not determined.

Unfortunately the intensity of EGFP fluorescence at this time-point was reduced, probably because of the sequestration of the RNA within the nucleus, also the proportion of EGFP positive cells was low probably due to faster growth of untransfected cells, making the acquisition of fluorescent cells by FACS more prolonged. Pilot tests were carried out to ensure that the cells could survive the separation procedure and that the RNA made from them was of good quality (addressed in 5.3.1.2.3).

5.3.1.2 Fluorescent activated cell sorting (FACS)

The use of the FACS separation system FACSCalibur (BD Biosciences), was kindly provided by Prof. J.Mottram (Wellcome Centre for Molecular Parasitology, Anderson College, Glasgow).

Firstly, parameters for cell-separation were established.

5.3.1.2.1 Cell density

The pattern of gene expression varies during different phases of cell growth, so in order to maintain consistency it was important that the cells would be actively growing at harvest. The exponential phase is considered to be up to 95% confluence. Pervious experiments indicated that the transfection of cells with densities below 50% resulted in cell death (data not shown), so cells were plated at different densities in large tissue culture flasks and allowed to grow overnight, and cultures closest to 50% confluence were selected for

transfection. This would allow the maximum exponential growth post transfection. Cells were then photographed after 24 and 48 hrs.

5.3.1.2.2 Gates and purity

Cells were sorted during FACS using gates set with control samples. This was first tested using fluorescent beads. Voltage gates were optimised for each bead colour separately and used to sort and collect green beads from an approximately equal mixture of non-fluorescent, fluorescent green and fluorescent red beads in PBS. For sorting, three modes are available: Single cell, where only single positive beads are collected; Recovery, where all events containing a positive bead are collected; and exclusion, where all negative beads are rejected. Single cell sorting was not used since it was expected that the yield would be low. Using the recovery mode, the proportion of green beads was enriched from 56% to 69%. In exclusion mode, where only pure positive fractions are collected, 98% of the beads recovered were green (Figure 49). Therefore exclusion mode was chosen for future cell sorting.

Following these successful attempts at sorting, the gates were then established for separating EGFP positive cells from non-expressing cells. For the control samples 3T3 cells were transfected with either regent alone or pLoxEGFP5, and harvested after 24 hours. After separation optimisation, the reagent alone sample produced a single 'negative' grouping. The pLoxEGFP5-transfected sample produced two groupings: negative and EGFP positive, since not all cells in the sample would be expressing EGFP, thus allowing identification of the EGFP positive fraction (Figure 50). Gates were drawn around the two fractions and the fluorescence intensity within each gate (region 1 and 2) plotted for each sample. The shape of the curves within region 1 is similar for each sample, indicating that there are few contaminating EGFP positive cells within the untransfected cell gate. Correspondingly, when events collected using region 2 are plotted, few cells were counted from the negative sample indicating little untransfected-cell contamination within the EGFP positive gate (Figure 50). The separation procedure was lengthy. In the previous experiment, setting the gates, collection and concentration of 1 X 10^5 EGFP positive and 1 X 10^5 negative events took on average 5 hours.

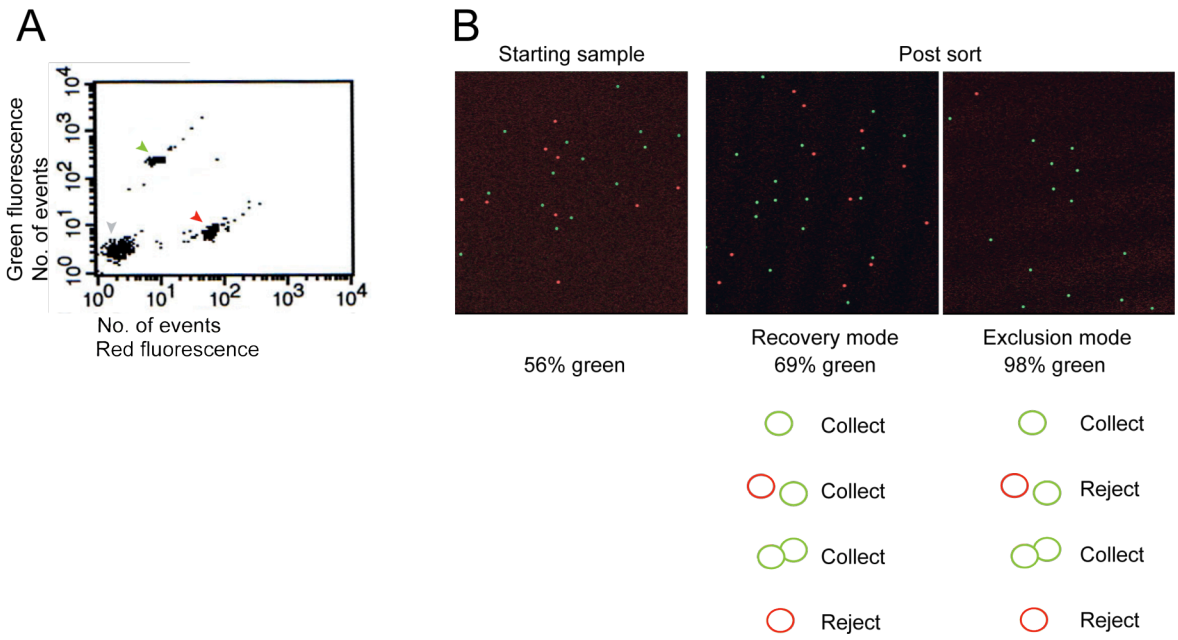


Figure 49 Calibration beads shown before and after sorting. A. Facscalibur acquisition dot-plot showing the differences in the bead mixture based on emission spectra during passage through the 488 nm laser. Each crossing of the laser beam is termed an event, which can include more than one bead. Separation parameters were optimised for a mixture of non-fluorescent, red fluorescent and green fluorescent beads. Each grouping corresponds to clear (grey arrow); FITC (green arrow) and PE (red arrow) beads. Sort gates were then drawn around the regions to be collected (not shown). **B.** The bead mixture before and after sorting and recovery using two different sort modes. In recovery mode, every positive event is collected, even if a contaminating bead is present. This results in high recovery of positive events, but lower purity. In exclusion mode only events containing positive beads are collected, mixtures are rejected. This results in a pure but lower yield of positive events.

The 3T3 transfections were repeated, again using reagent-alone or pLoxEGFP5. Both samples were used to refine the gate settings. The EGFP positive fraction was sorted and collected from the pLoxEGFP5 sample. After sorting, the proportion of EGFP-positive cells detectable by fluorescence microscopy was much higher (63%; 28 of 44 cells) than in the unsorted sample (26.5%; 21 of 79 cells), and EGFP fluorescence could be detected in most but not all cells (Figure 51). The reason why some cells appeared non-fluorescent during microscopy was unclear, but could have been because fluorescence detection during FACS is more sensitive than the eye or camera, or due to the position of the gates, or plasmid loss from the cell. As a precaution the gate drawn for collection of the negative sample was tightened.

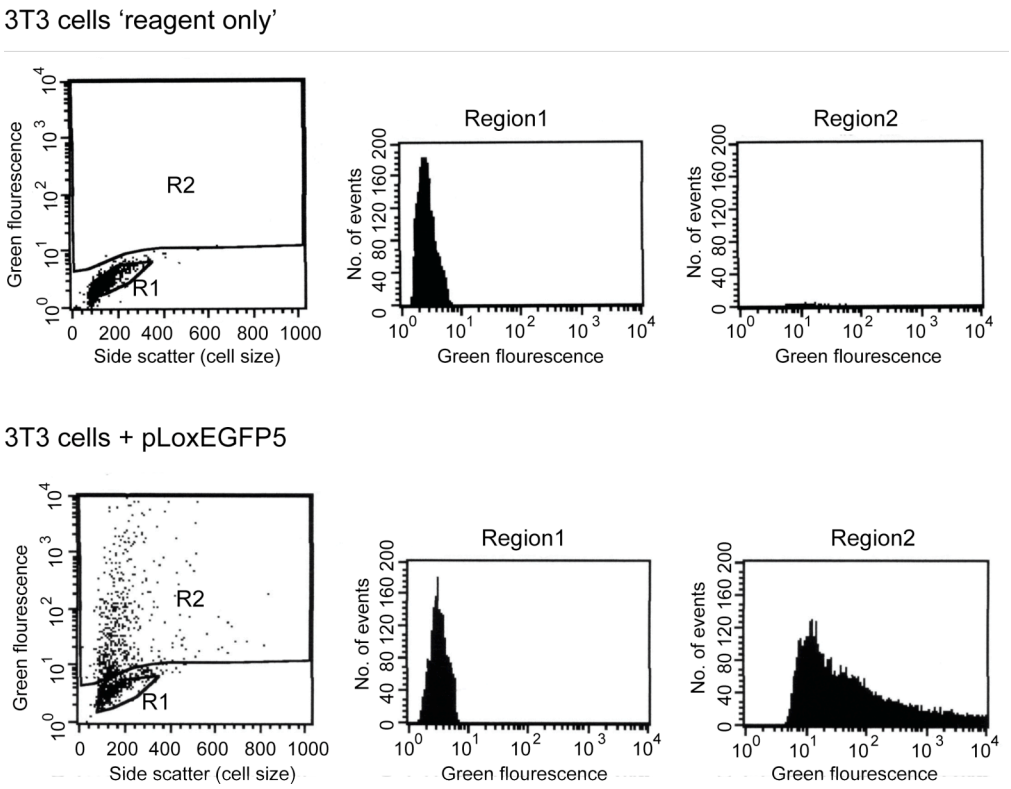


Figure 50 Establishment of sort gates. Reagent-only and pLoxEGFP5-transfected 3T3 cells were used to establish separation settings. Acquisition dot plots (left) show a single cell-group in the reagent-only sample corresponding to untransfected cells. In the pLoxEGFP5-transfected sample two groups are apparent, corresponding to untransfected and EGFP-positive fractions. Gates were drawn around the two groups –region 1 and 2 – and the number and fluorescence intensity of events plotted in each (histograms middle and right). Region 1 histograms (middle) show similar area curves indicating few contaminating EGFP positive cells within the negative gate. Region 2 histograms (right) show few negative cells (top) contaminating the EGFP positive gate.

5.3.1.2.3 Cell viability and RNA integrity

It was of concern that the cells used during the sorting –and those recovered would have been stored in PBS for a considerable amount of time. To ensure that cells collected using the cell-sorter could survive the procedure, pLoxEGFP5-transfected cells were saved at each stage of the sorting in the previous experiment, plated into tissue culture dishes and incubated overnight. The cells were photographed using brightfield and fluorescent microscopy from all stages of the experiment; starting sample; post trypsin harvest; recovered sorted cells and unsorted post sort (starting sample after the experiment) and appeared healthy, with no bacterial or fungal contamination observed (Figure 51). Leftover starting sample was also tested for viability using trypan blue reagent, a diazo dye used to stain cells with disrupted outer membranes. Cells with an intact membrane exclude the dye, and dead cells are coloured blue. Cells were found to be a healthy with few taking up the pigment (Figure 52). In conclusion, the sorting procedure and the maintenance of the

sample in PBS on ice for the extended period of 5 hours was not found to be detrimental to cell viability.

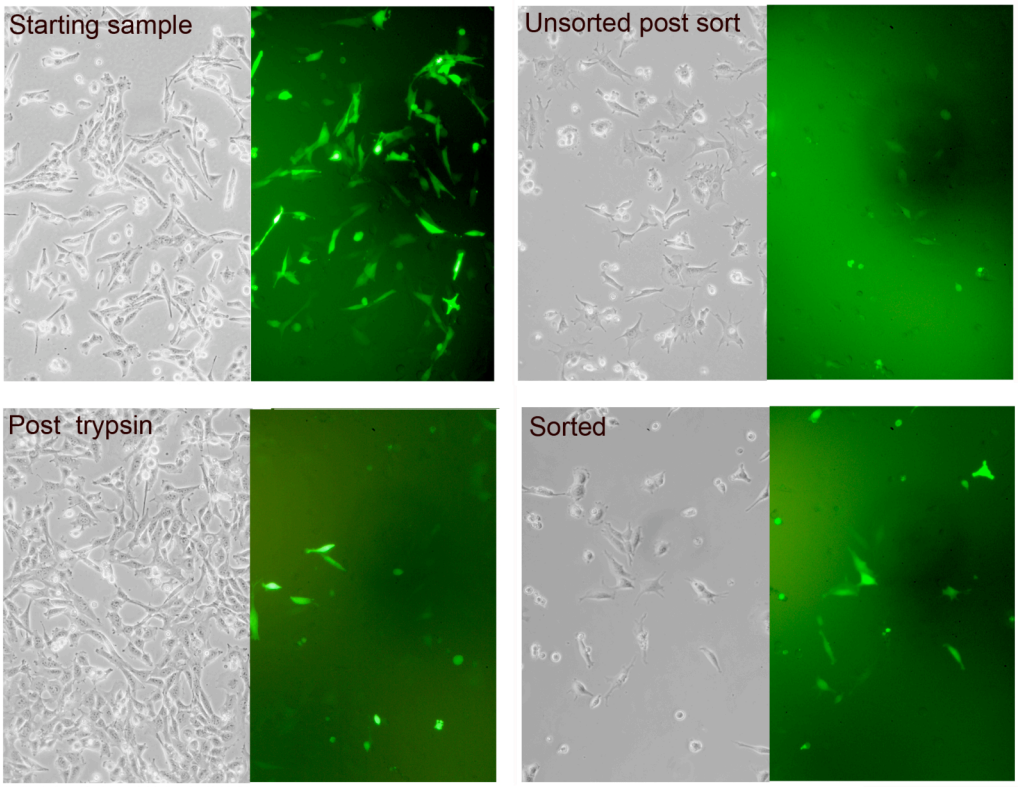


Figure 51 Cell viability after the sorting procedure Brightfield and fluorescent micrographs. pLoxEGFP5 transfected 3T3 cells were used to test the sorting procedure and subsequent cell viability. The starting sample was photographed. Samples were taken at the stages indicated, plated onto tissue culture dishes and incubated overnight. The sorting procedure was not found to be detrimental to cell viability. The proportion of EGFP positive cells was increased in the sorted compared to the unsorted and post-trypsin samples.

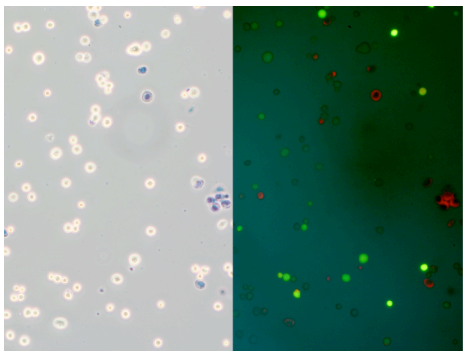


Figure 52 Cell viability after storage on ice. Left, brightfield and right, fluorescent micrographs. 3T3 Cells from the post-trypsin harvest were stored on ice for 5 hours and tested for viability using trypan blue reagent. A few cells took up the dye but most cells did not, indicating the majority had an intact membrane and were still alive.

The separation procedure was prolonged, allowing the collection of one positive and one negative fraction from one transfection per day. Since three replicates of each sample were required for the analysis, the collection of samples would span many weeks. It was

necessary to store sorted cells to allow all replicates collected to be processed at the same time, limiting technical variations, whilst maintaining quality. To confirm the integrity of RNA during storage of the samples, RNA was extracted from unsorted and sorted cells stored in RNA-later solution. RNA-later is a proprietary solution formulated for the storage of cell and tissue samples without compromising the integrity of the RNA prior to extraction. Mouse 3T3 cells were transfected with reagent alone; pLoxEGFP5; or pLoxEGFP250. After 24 hours, cells were harvested and stored in RNAlater for 11 weeks at -20ºC. The pLoxEGFP5 and pLoxEGFP250 transfections were repeated, EGFP positive and negative fractions sorted and 5 X 10^5 events collected from each fraction as described previously. The samples were stored for one week in RNAlater at -20ºC. The cells from all samples were collected as previously described, and RNA extracted using RNeasy mini kit (Qiagen), the method recommended by the Molecular Biology Support Unit for microarray analysis. RNA was visualised using guanidinium hydrochloride agarose gel electrophoresis, and assayed using the Bioanalyser (Agilent Technologies). The RNA stored before sorting was of the same quality as RNA prepared from live 3T3 cells, but the RNA extracted form sorted cells showed signs of degradation (Figure 53A+B). This could have been due to temperature rises in the sample during sorting, so to counteract this, the sorting protocol was modified such that the source sample was aliquotted and stored on ice, each aliquot discarded after 30 minutes sampling. The buffers were also kept on ice and topped up more frequently. When recovering the sample it proved difficult to form a pellet and a significant amount of cellular material remained bound to the wall of the tube. The number of cells (events) recovered from each FACS is small, so no losses could be tolerated. For this reason, we decided to store collected cells in RNA lysis buffer, the first solution used in the RNA extraction, at -80ºC since this would reduce the sample loss to zero. After these measures were adopted, RNA integrity was no longer compromised (Figure 53C).

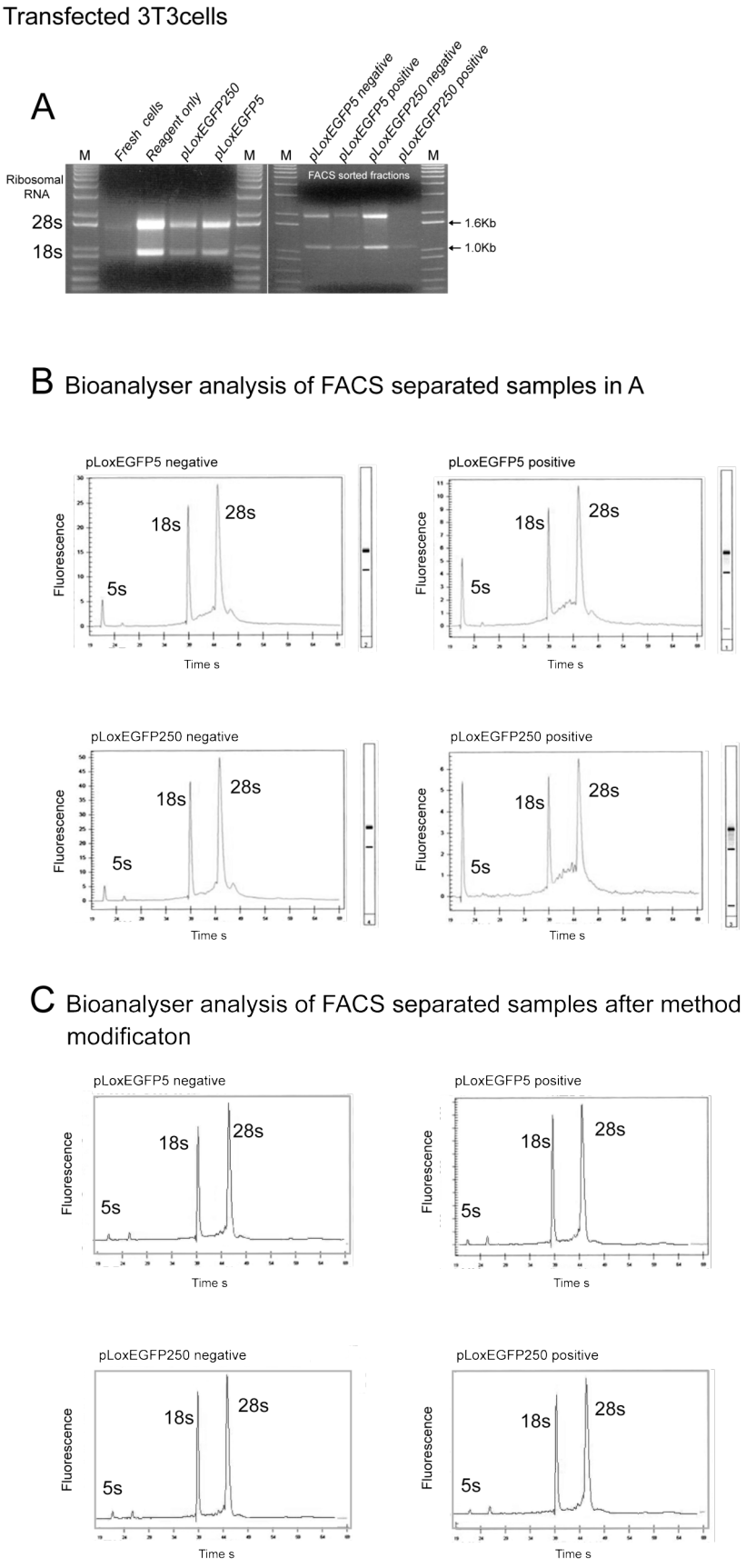


Figure 53 RNA integrity following storage in RNAlater solution. A. Guanidinium thiocyanite agarose gel electrophoresis. The samples on the left hand gel had been stored in RNAlater for 11 weeks without affecting the quality of the RNA extracted. RNA was compared to that extracted from live cells as a control. The samples on the right hand gel were first sorted using FACS, then stored in RNAlater solution for 3 weeks. The sample quality of the positive fractions is reduced since the intensity of the ribosomal RNA band at 28s is not higher than that of the 18s band. **B.** Bioanalyser results. A proportionately large 5s peak compared to the 18s and 28s peaks, and the deviation

from the baseline between the 18s and 28s peaks indicated sample degradation within the positive cell-fractions. **C.** After minor modification of sample handling during sorting, the RNA isolated was of better quality using the same criteria of ribosomal RNA comparison. Note the proportionately small 5s peak and the low baseline between 18 and 28s peaks, indicating a lack of degradation.

5.3.2 Transcript expression array

The Affymetrix "GeneChip Mouse Genome 430 2.0 Array" allows for the analysis of 39,000 transcripts in a single array. Each transcript is represented by multiple probe sets (four adjacent oligonucleotides) allowing several measurements to be made to accurately evaluate individual transcript levels. Whilst coverage of the mouse genome may not be absolutely complete, it is comprehensively represented on the array since the probe sets were derived from sequences selected from GenBank®, dbEST, and RefSeq, which are further refined by comparison with the draft assembly of the mouse genome (Whitehead Institute for Genome Research MGSC, April 2002). In this particular GeneChip, the bias for the probeset position is towards the 3' end of the transcript, which is not ideal for the analysis of alternatively spliced isoforms. It should however reveal any missplicing that results in a premature termination codon, since this would lead to nonsense mediated decay and reduced levels of transcript (Nagamitsu *et al.*, 2000), as is the case with the *chloride channel* in DM1 (Charlet *et al.*, 2002).

5.3.2.1 Sample preparation for whole transcript arrays

The samples were prepared in triplicate, in parallel, such that three chips were hybridised for each sample group pLoxEGFP5 'on'; pLoxEGFP5 'off'; pLoxEGFP250 'on' and pLoxEGFP250 'off' –twelve chips in total. The mouse 3T3 cells used for transfections were plated at diluting densities some weeks before starting the experiment such that cells plated for each transfection had the same passage number, even though the transfections were staggered over two consecutive days for 3 weeks. Mouse 3T3 cells were plated to give 50% confluence after 24 hours growth, then transfected with pLoxEGFP5 or pLoxEGFP250. After 48 hours, the cells were harvested using trypsin and washed in ice cold PBS, and kept cool throughout processing. EGFP positive and EGFP negative fractions were sorted and collected from transfected samples using a FACSCalibur cell-sorter, with the parameters established in section 5.3.1.2. On any particular day, only one negative and one positive fraction was collected from a single transfection, so after sorting, cells were concentrated by centrifugation, resuspended in RNA lysis buffer and stored at -80ºC until all samples had been collected. The transfections and cell sorting were repeated

until three replicates each of pLoxEGFP5 and pLoxEGFP250 EGFP positive and EGFP negative fractions had been gathered. Total RNA was prepared, and transferred to MBSU for quality assessment and to carry out the microarray hybridisation procedure.

RNA samples were labelled using the Affymetrix one-cycle gene chip target labelling protocol, which does not involve amplification of the sample. After hybridisation, signals were first normalised using robust microchip average (RMA (Irizarry *et al.*, 2003)), then scored for differential expression between pLoxEGFP5 *vs.* pLoxEGFP250 transfected samples using Rank Products (Breitling *et al.*, 2004). Expression differences were also ordered according to Iterative Group Analysis (iGA), which statistically identifies functional classes of genes with significant changes (Breitling *et al.*, 2004). These genes may be members of a pathway for instance. Not all members of the group need to be changed and not by a large amount.

5.3.2.2 Results of analysis

The comparison of pLoxEGFP5- *vs.* pLoxEGFP250- should give an indication of the fidelity of separation since each sample should contain only the untransfected fraction of non-fluorescent cells, and so there should be no difference in gene expression levels. According to the rank product analysis however, with a false discovery rate of less than 5%, in the pLoxEGFP250- sample, 106 genes are up-regulated and 10 are down-regulated, relative to pLoxEGFP5-. It is possible therefore that the separation was not absolute, or that some pLoxEGFP250 transfected cells did not fluoresce perhaps because of transcript retention, thus contaminating the negative sample with cells expressing expanded arrays. Likewise, there should be no difference in transcript expression between pLoxEGFP250+ *vs.* pLoxEGFP5- and pLoxEGFP250+ *vs*. pLoxEGFP250-, since again, both pLoxEGFP5- and pLoxEGFP250- samples should contain only untransfected cells. In fact the differences between these two are large (312 up and 79 down-regulated compared to 49 up and 5 down-regulated respectively, (Table 11: Tests 2 and 5)), compared to the differential expression *within* pLoxEGFP250+ *vs*. pLoxEGFP250- (Table 11, Test 3), indicating that the nature of pLoxEGFP250+ and pLoxEGFP250- is similar. Down-regulated genes identified within pLoxEGFP250- *vs.* pLoxEGFP5- were compared to those listed in pLoxEGFP250+ *vs.* pLoxEGFP5- (Table 11, Tests 5 and 2). If there was contamination of pLoxEGFP250+ within the pLoxEGFP250- sample then one would expect the same genes to be listed. Seven of the 10 genes were found within the 5% cut off, so indicating probable contamination of the pLoxEGFP250- sample with pLoxEGFP250+. The down-regulated gene lists were then compared within the pLoxEGFP5+ *vs*. pLoxEGFP250+ and

pLoxEGFP5+ *vs*. pLoxEGFP250- analyses (Table 11, Tests 1 and 6). Again if there was contamination of pLoxEGFP250- with pLoxEGFP250+, the same genes would be expected in the lists. At the same 5% limit, 5 of 6 genes were also listed within the 42 up-regulated genes (*N.B.* up-regulated genes were compared because the base-line was reversed in the pLoxEGFP5+ *vs.* pLoxEGFP250- (test 6, Table 11) comparison).

Table 11 Up and down regulation of transcripts

Test *vs.* baseline	Up	Down	Test
1. 250+ *vs.* 5+	128	6	Extended CUG repeats
2. 250+ *vs.* 5-	312	79	Extended CUG repeats and construct related
3. 250+ *vs.* 250-	49	5	Extended CUG repeats and construct related
4. 5+ *vs.* 5-	130	90	Normal length repeats and construct related
5. 250- *vs.* 5-	106	10	Separation control
6. 5+ *vs.* 250-	42	66	Normal length repeats, construct related and separation control

Note–False discovery rate limit ≤ P=0.05.

To detect structure in the relationship between our samples, a principle components analysis (PCA) was carried out (at the MBSU). Briefly, PCA is a type of factor analysis, which uses a mathematically defined transformation of a given multivariable data set. This simplifies the data so that the greatest variance lies on the first coordinate – termed the principle component. It reduces the dimensionality of the data in effect, by combining variables until only the defining components remain, and is commonly used with microarray data sets to look at the relationship between samples. The results of the analysis show the relationships between the pairs of separated 'EGFP on' and 'EGFP off' samples, for each replicate of pLoxEGFP5 and pLoxEGFP250. In general the relationship between EGFP off samples (-5(1); -5(2); -5(3); -250(2) and -250(3)) and EGFP on samples (+5(1); +5(2); +5(3); +250(2) and +250(3)), irrespective of whether they were derived from 5 or 250 CUG repeat expression, shifts upwards and to the right for negative samples. The exception to this is the -250(1) and +250(1) pair, where the shift is downwards (Figure 54A). To test whether the samples had been mislabelled, residual RNA unused in the

microarray analysis was subjected to semi-quantitative EGFP RT-PCR. Unfortunately, the -250(1) sample was used up during array hybridisation, but comparison of +250(1) EGFP PCR product levels with other EGFP+ RNA fractions should indicate whether the +250(1) sample belonged with the positive or negative group. RT-PCR was carried out using the remaining RNA samples. To ensure that the levels of product seen after amplification were a quantification of template in the starting sample, the number of cycles required to generate linearity of the PCR product was established using control RNA. Total RNA from untransfected, and pLoxEGFP-transfected 3T3 cells was mixed in varying proportions: 0; 1; 5; 10; 20; 50 and 100% pLoxEGFP-transfected, and used to determine the optimal number of cycles for amplification (Figure 30 Bottom).

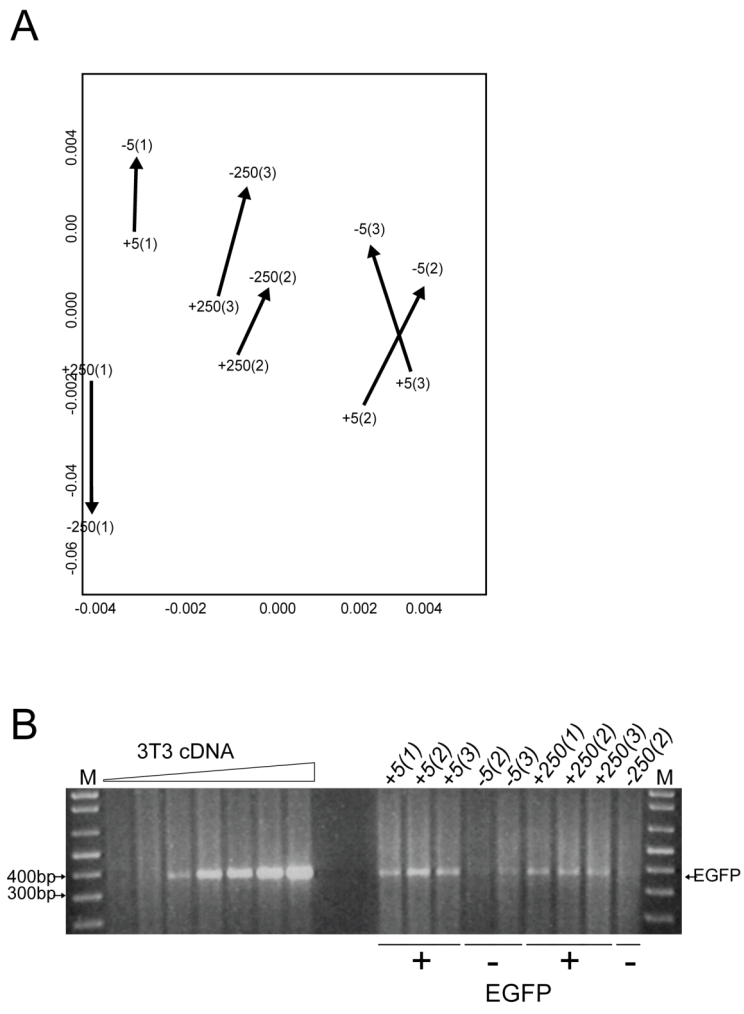


Figure 54 Samples +250(1) and -250(1) were not accidentally switched A. Principle component analysis. Mathematical algorithm plotting the relationship between the FACS-separated sample pairs suggested +250(1) and -250(1) may have been switched. **B.** Semi-quantitive EGFP RT-PCR indicates that the levels of PCR product amplified from +250(1) were equivalent to the other EGFP+ fractions, therefore a switch was unlikely.

For both control and test samples 500 ng of RNA was reverse-transcribed, and 2/5 of the reaction product used for the EGFP PCR analysis. Fifteen cycles were used for the amplification (Figure 54B). Note that it was not possible to include a "no RT" control due to the very limited amounts of sample. In fact, the samples would be expected to be positive for DNA since they were not DNase treated (small amounts of plasmid DNA such as pLoxEGFP5/250 used for the transfections tend to co-isolate with RNA during preparation). This was of little consequence here since we were looking at the level of separation. If the product arose from plasmid DNA rather than transcript, the separation level was still a valid measure. The amount of EGFP product amplified from the +250(1) sample was found to be equivalent to the other EGFP positive samples, which suggested the samples had not been accidentally switched.

Whilst this data set was incomplete since some samples were used up during array hybridisation, there were enough to indicate the level of separation, and at the molecular level it was clear that separation was not totally successful. As predicted from the microarray results, EGFP negative samples are contaminated with EGFP transcripts across all samples, whether pLoxEGFP5 or pLoxEGFP250 derived. In conclusion, since the separation had failed, reliable data was restricted to the pLoxEGFP250+ *vs.* pLoxEGFP5+ comparison (Test 1 Table 11) since both these fractions must contain the construct.

In this comparison 128 genes were up-regulated (Table 12) and 6 down-regulated (Table 13). Genes subjected to non-sense mediated decay would appear to be down-regulated in this analysis, as would those with exon exclusion in alternatively spliced 3' ends. Of the 6 down-regulated genes 3 were unannotated RIKEN DNA bank cDNAs, the remainder comprised *Hes1* (*hairy and enhancer of split 1*); *D3Ertd300e* (*DNA segment, Chr 3, ERATO Doi 300, expressed*) and *Tfrc* (*transferrin receptor*). *Hes1* is involved in neuronal differentiation (Bae *et al.*, 2000) and also tenuously thought to be a negative regulator of myogenesis due to its 'helix-loop-helix' domain, and its role in maintenance of the undifferentiated state (Kageyama *et al.*, 2000). At the time of this analysis, nothing was known about the function of the *D3Ertd300e* gene product (also known as p38 interacting protein), since then researchers have found it is critically required for the regulation of gastrulation in the mouse. Zohn *et al.* created a mouse model with a splicing defect in this gene, which exhibits exencephaly and spina bifida. These mutations are not completely penetrant, probably because low levels of transcripts are processed normally. The same researchers showed that mutant embryos homozygous for a more severe *p38IP* allele are necrotic, developmentally delayed, exhibit misshapen head folds and exencephaly, and fail

to form somites or form only a few anterior somites (Zohn *et al.*, 2006). Some aspects of this phenotype are reminiscent of the congenital form of DM1 such as misshapen head and impaired muscle development, so in hindsight this gene should have been selected for further characterisation.

Iterative Group Analysis (iGA) predicted the involvement of iron homeostasis (*transferrin receptor*), and nucleolus organisation and biogenesis pathways. Nucleoli, the sites of ribosome biosynthesis, enlarge in response to large protein production requirements. Due to the nuclear retention of EGFP transcripts in the pLoxEGFP250 sample, a high protein production requirement would not be expected compared to pLoxEGFP5 transfected cells where the transcript is not thought to be retained in the nucleus. Supporting this, during transfection experiments, EGFP fluorescence from pLoxEGFP250 transfected cells was generally found to be of a lower intensity than in pLoxEGFP5 transfected cells (Figure 45), therefore differences relating to this pathway may be due to artefact. Up-regulated genes (Table 12) were also assessed to determine whether mutations in these genes could lead to DM-like symptoms, but no connections could be made. IGA analysis was used to facilitate interpretation by prediction of the putatively affected pathways, most of which were found to be cell-growth and metabolism related, such as gelatinase A activity; glucuronate and glycospingolipid metabolism; lysosome organisation and biogenesis, and response to sterol depletion; all of which could relate to the effects of excessive EGFP protein production in pLoxEGFP5 transfected cells. It was stated earlier that an increase in the number of genes down-regulated upon CUG repeat expression could be due to nonsense mediated decay. This could also result from decreased mRNA stability. Houseley *et al*. found that over-expression of an MBNL1/GFP fusion protein in *Drosophila* already expressing CUG repeats, resulted in increased stability of the CUG repeat transcript (Houseley *et al*. 2005). CUGBP1 has deadenylase activity, if MBNL1 aided mRNA stability in other transcripts, it follows then that depletion by CUG repeat expression may appear as an increase in down-regulated genes. Conversely, for those genes up-regulated, the stability of the mRNA could be increased.

After extensive analysis of the genes listed using published literature, connections between the genes listed and DM symptoms seemed tentative. By this time Affymetrix had developed mouse exon arrays, which were ideally suited for the discovery of alternative splice forms, so no genes were selected for further characterisation from this analysis.

Table 12 Genes up-regulated; pLoxEGFP250+ *vs.* pLoxEGFP5+.

FDR	Fold Change	Gene Symbol	Gene Title
0	2.22	*Dbp*	D site albumin promoter binding protein
0	2.46	*Pla2g1b*	Soluble PLA2-Ib precursor (Pla2g1b)
0	2.37	*Ogn*	osteoglycin
0	2.23	*Ogn*	osteoglycin
0	2.15	*Tmsb4x*	thymosin, beta 4, X chromosome
0	2.11	*C3*	complement component 3
0	2.07	*C1s*	complement component 1, s subcomponent
0	2.57	*Cyr61*	cysteine rich protein 61
0	2.07	*Lcn2*	lipocalin 2
0.1	2.17	*D530037H12Rik*	RIKEN cDNA D530037H12 gene
0.09	1.98	*Efemp1*	epidermal growth factor-containing fibulin-like extracellular matrix protein 1
0.08	2.31	*Cyr61*	cysteine rich protein 61
0.15	1.92	*Cbr2*	carbonyl reductase 2
0.21	1.95	*Ddx3y*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked
0.2	1.9	*Trp53inp1*	transformation related protein 53 inducible nuclear protein 1
0.19	1.87	*Nt5e*	5' nucleotidase, ecto
0.18	1.83	*Matn2*	matrilin 2
0.17	2.08	---	Avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog (Maf), mRNA
0.21	2.12	*Sepp1*	selenoprotein P, plasma, 1
0.2	1.87	*Svs5*	seminal vesicle secretion 5
0.19	1.86	*Dcn*	decorin
0.18	1.91	*Maf*	avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog
0.17	1.35	*Car9*	carbonic anhydrase 9
0.17	1.8	*Tgfbi*	transforming growth factor, beta induced
0.2	1.56	*Fos*	FBJ osteosarcoma oncogene
0.19	1.84	*Rbmy1a1*	RNA binding motif protein, Y chromosome, family 1, member A1
0.19	1.78	*Tgfbi*	transforming growth factor, beta induced
0.36	1.82	*Sqrdl*	sulfide quinone reductase-like (yeast)
0.48	1.88	*Fzd10*	frizzled homolog 10 (Drosophila)
0.47	2.06	*Abca1*	ATP-binding cassette, sub-family A (ABC1), member 1
0.65	1.74	*Zc3h6*	zinc finger CCCH type containing 6
0.62	1.74	*Nfkbiz*	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, zeta
0.61	1.83	*1110032E23Rik*	RIKEN cDNA 1110032E23 gene
0.79	1.69	*Matn2*	matrilin 2
0.8	1.68	*Tgfbi*	transforming growth factor, beta induced
0.81	1.72	*Capn6*	calpain 6
0.81	1.69	*Ugt1a2 /// Ugt1a6a /// Ugt1a10 /// Ugt1a7c /// Ugt1a5 /// Ugt1a9 /// Ugt1a6b /// Ugt1a1*	UDP glucuronosyltransferase 1 family, polypeptide A2 /// UDP glucuronosyltransferase 1 family, polypeptide A6A /// UDP glycosyltransferase 1 family, polypeptide A10 /// UDP glucuronosyltransferase 1 family, polypeptide A7C /// UDP glucuronosyltransferase 1 family, polypeptide A5 /// UDP glucuronosyltransferase 1 family, polypeptide A9 /// UDP glucuronosyltransferase 1 family, polypeptide A6B /// UDP glucuronosyltransferase 1 family, polypeptide A1
0.82	1.26	*Adm*	adrenomedullin
0.85	1.74	*Enpp2*	ectonucleotide pyrophosphatase/phosphodiesterase 2
0.88	1.68	---	---
1.02	1.73	*Steap2*	PREDICTED: six transmembrane epithelial antigen of prostate 2 [Mus musculus], mRNA sequence
1	1.39	*Ccng2*	cyclin G2
1.02	1.66	*2310047A01Rik*	RIKEN cDNA 2310047A01 gene
1.16	1.86	*Aldh3a1*	aldehyde dehydrogenase family 3, subfamily A1
1.13	1.68	*Aldh6a1*	aldehyde dehydrogenase family 6, subfamily A1
1.15	1.65	*Ctso*	cathepsin O

1.13	1.65	*Ppil6*	peptidylprolyl isomerase (cyclophilin)-like 6
1.15	1.64	*Alcam*	activated leukocyte cell adhesion molecule
1.29	1.66	*Rab6b*	RAB6B, member RAS oncogene family
1.3	1.47	*Eno2*	enolase 2, gamma neuronal
1.29	1.67	*Mtss1*	metastasis suppressor 1
1.27	1.64	*---*	0 day neonate eyeball cDNA, RIKEN full-length enriched library, clone:E130004J04 product:unclassifiable, full insert sequence
1.34	1.62	*Islr*	immunoglobulin superfamily containing leucine-rich repeat
1.35	1.64	*Luc7l2*	LUC7-like 2 (S. cerevisiae)
1.45	1.65	*1810011O10Rik*	RIKEN cDNA 1810011O10 gene
1.46	1.74	*Car6*	carbonic anhydrase 6
1.44	1.69	*Ly6c*	lymphocyte antigen 6 complex, locus C
1.41	1.68	*Cdsn*	corneodesmosin
1.41	1.62	*H1f0*	H1 histone family, member 0
1.4	1.61	*Ii*	Ia-associated invariant chain
1.48	1.62	*Pink1*	PTEN induced putative kinase 1
1.48	1.61	*Tomm7*	translocase of outer mitochondrial membrane 7 homolog (yeast)
1.46	1.61	*Mapk12*	mitogen-activated protein kinase 12
1.48	1.65	*Pdlim5*	PDZ and LIM domain 5, mRNA (cDNA clone MGC:46824 IMAGE:4457868)
1.49	1.57	*Tmem53*	transmembrane protein 53
1.47	1.63	*Ly6a*	lymphocyte antigen 6 complex, locus A
1.52	1.49	*Lss*	lanosterol synthase
1.56	1.68	*Mgp*	matrix Gla protein
1.72	1.62	*Gchfr*	GTP cyclohydrolase I feedback regulator
1.7	1.59	*Pla2g1b*	phospholipase A2, group IB, pancreas
1.96	1.59	*BC010787*	cDNA sequence BC010787
1.97	1.59	*Cyba*	cytochrome b-245, alpha polypeptide
1.99	1.59	*Pltp*	phospholipid transfer protein
2.03	1.62	*Sdpr*	serum deprivation response
2.03	1.59	*Hexb*	hexosaminidase B
2.07	1.58	*Hdac11*	histone deacetylase 11
2.04	1.6	*Sulf2*	sulfatase 2
2.03	1.6	*Haghl*	hydroxyacylglutathione hydrolase-like
2.01	1.58	*Dcxr*	dicarbonyl L-xylulose reductase
2.09	1.33	*Ccng2*	cyclin G2
2.1	1.57	*Wdr34*	WD repeat domain 34
2.07	1.57	*LOC545323*	similar to neurobeachin-like 1
2.22	1.56	*Erbb2ip*	Erbb2 interacting protein
2.3	1.56	*Dzip1*	DAZ interacting protein 1
2.33	1.57	*Cldn10*	Claudin 10 (Cldn10), transcript variant 1, mRNA
2.34	1.57	*Tgfbi*	transforming growth factor, beta induced
2.36	1.58	*Igfbp3*	insulin-like growth factor binding protein 3
2.34	1.59		Transcribed locus, moderately similar to XP_484812.1 PREDICTED: hypothetical protein XP_484812 [Mus musculus]
2.36	1.56	*Irf2bp2*	interferon regulatory factor 2 binding protein 2
2.6	1.58	*Eif2s3y*	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked
2.58	1.36	*AI451617*	expressed sequence AI451617
2.58	1.61	*BC049816*	cDNA sequence BC049816
2.56	1.54	*Hrasls3*	HRAS like suppressor 3
2.7	1.55	*Alcam*	activated leukocyte cell adhesion molecule
2.72	1.6	*Sdpr*	serum deprivation response
2.77	1.67	*Mmp2*	matrix metallopeptidase 2
2.75	1.69	*0910001A06Rik*	RIKEN cDNA 0910001A06 gene (0910001A06Rik), mRNA
2.79	1.54	*Neu1*	neuraminidase 1
2.76	1.14	*Acss2*	acyl-CoA synthetase short-chain family member 2
2.74	1.61	*2210419I08Rik*	RIKEN cDNA 2210419I08 gene
2.9	1.68	*Aspn*	asporin
3.01	1.52	*Ranbp2*	RAN binding protein 2
2.99	1.54	*Maf*	avian musculoaponeurotic fibrosarcoma (v-maf) AS42

			oncogene homolog
3.24	1.56	*1500004A08Rik*	RIKEN cDNA 1500004A08 gene
3.21	1.52	---	---
3.56	1.79	---	---
3.54	1.55	*Xdh*	xanthine dehydrogenase
3.52	1.42	*6330406I15Rik*	RIKEN cDNA 6330406I15 gene
3.76	1.52	*Foxc1*	forkhead box C1
3.75	1.51	*Tgm2*	transglutaminase 2, C polypeptide
3.76	1.52	*Galntl2*	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 2
3.84	1.53	*AI325464*	Expressed sequence AI325464, mRNA (cDNA clone MGC:30834 IMAGE:4006813)
3.92	1.13	*Adm*	adrenomedullin
4.08	1.52	*Gstm1*	glutathione S-transferase, mu 1
4.09	1.51	*Psap*	prosaposin
4.22	1.53	*Calu*	calumenin
4.21	1.64	*Mmp2*	matrix metallopeptidase 2
4.19	1.52	*Hsd3b7*	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 7
4.42	1.51	*Rnase4*	ribonuclease, RNase A family 4

FDR, False discovery rate; **---**, unannotated cDNA or EST sequences.

Table 13 Genes downregulated; pLoxEGFP250+ *vs.* pLoxEGFP5+.

FDR	Fold Change	Gene Symbol	Gene Title
0	-2.69	---	---
3.5	-1.88	*Hes1*	hairy and enhancer of split 1 (Drosophila)
3	-1.57	*D3Ertd300e*	DNA segment, Chr 3, ERATO Doi 300, expressed
3.25	-1.57	*Tfrc*	transferrin receptor
3	-1.75	*AI662270*	expressed sequence AI662270
3.83	-1.57	*1810054D07Rik*	RIKEN cDNA 1810054D07 gene

FDR, False discovery rate; **---**, unannotated cDNA or EST sequences.

Table 14 Genes upregulated; pLoxEGFP250+ *vs.* pLoxEGFP5-.

FDR	Fold Change	*Gene Symbol*	Gene Title
0	6.29	*1110032E23Rik*	RIKEN cDNA 1110032E23 gene
0	4.99	*Tmsb4x*	thymosin, beta 4, X chromosome
0	4.17	*Car6*	carbonic anhydrase 6
0	4.15	*Sdpr*	serum deprivation response
0	3.8	*Sdpr*	serum deprivation response
0	3.74	*Sdpr*	serum deprivation response
0	3.74	*Ddx3y*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked
0	3.51	*Sqrdl*	sulfide quinone reductase-like (yeast)
0	3.15	*Alcam*	activated leukocyte cell adhesion molecule
0	3.01	*Dcn*	decorin
0	3.02	*Gsta1 /// Gsta2*	glutathione S-transferase, alpha 1 (Ya) /// glutathione S-transferase, alpha 2 (Yc2)
0	3.12	*Efemp1*	epidermal growth factor-containing fibulin-like extracellular matrix protein 1
0	2.94	*Rbmy1a1*	RNA binding motif protein, Y chromosome, family 1, member A1

0	2.9	*2210419I08Rik*	RIKEN cDNA 2210419I08 gene
0	2.89	*Alcam*	activated leukocyte cell adhesion molecule
0	2.88	*Maf*	avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog
0	2.66	*9430093N24Rik*	MKIAA1983 protein
0	2.59	*Alcam*	activated leukocyte cell adhesion molecule
0	2.55	*Mapk12*	mitogen-activated protein kinase 12
0	2.58	*Cyba*	cytochrome b-245, alpha polypeptide
0	2.49	*Ccl2*	chemokine (C-C motif) ligand 2
0	2.55	*Alcam*	activated leukocyte cell adhesion molecule
0	2.44	*---*	---
0	2.45	*Gatm*	glycine amidinotransferase (L-arginine:glycine amidinotransferase)
0	2.4	*Grem1*	gremlin 1
0	2.41	*Mtss1*	metastasis suppressor 1
0	2.42	*Gpnmb*	glycoprotein (transmembrane) nmb
0	2.36	*Ccl7*	chemokine (C-C motif) ligand 7
0	2.32	*Eif2s3y*	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked
0	2.52	*Fyb*	FYN binding protein
0	2.3	*H60*	PREDICTED: histocompatibility 60 [Mus musculus], mRNA sequence
0	2.3	*Rorb*	RAR-related orphan receptor beta
0	2.26	*Igf2r*	insulin-like growth factor 2 receptor
0.03	2.25	*Cdsn*	corneodesmosin
0.03	2.2	*Ampd3*	AMP deaminase 3
0.03	2.15	*Ppil6*	peptidylprolyl isomerase (cyclophilin)-like 6
0.03	2.21	*BC034507*	cDNA sequence BC034507
0.03	2.46	*Pla2g1b*	Soluble PLA2-Ib precursor (Pla2g1b)
0.03	2.15	*Ociad2*	OCIA domain containing 2
0.03	2.24	*Kctd4*	potassium channel tetramerisation domain containing 4
0.02	2.1	*Ass1*	argininosuccinate synthetase 1
0.02	2.15	*Mtss1*	metastasis suppressor 1
0.02	2.22	*3110004L20Rik*	RIKEN cDNA 3110004L20 gene
0.02	2.09	*Gchfr*	GTP cyclohydrolase I feedback regulator
0.02	2.15	*Enpp2*	ectonucleotide pyrophosphatase/phosphodiesterase 2
0.02	2.03	*Ctsb*	cathepsin B
0.04	2.03	*Gadd45a*	growth arrest and DNA-damage-inducible 45 alpha
0.04	2.01	*Svs5*	seminal vesicle secretion 5
0.04	2	*Dusp3*	dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)
0.04	2.08	*BC034507*	cDNA sequence BC034507
0.04	1.98	*Tnfaip2*	tumor necrosis factor, alpha-induced protein 2
0.04	1.97	*Nqo1*	NAD(P)H dehydrogenase, quinone 1
0.04	1.97	*Tgfbi*	transforming growth factor, beta induced
0.06	1.98	*H1f0*	H1 histone family, member 0
0.07	2.05	*Plf /// Plf2 /// Mrpplf3*	proliferin /// proliferin 2 /// mitogen regulated protein, proliferin 3
0.07	1.94	*U90926*	cDNA sequence U90926
0.07	2.04	*BC049816*	cDNA sequence BC049816
0.09	1.98	*Erbb3*	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)
0.12	2.02	*Cth*	cystathionase (cystathionine gamma-lyase)
0.12	2.32	*Cyr61*	cysteine rich protein 61
0.11	1.96	*Igf2r*	insulin-like growth factor 2 receptor
0.11	1.93	*Dpysl3*	dihydropyrimidinase-like 3
0.11	2.19	*Hcn1*	hyperpolarization-activated, cyclic nucleotide-gated K+ 1
0.11	1.95	*9430041P20Rik*	RIKEN cDNA 9430041P20 gene
0.11	1.91	*Ugt1a2*	UDP glucuronosyltransferase 1 family, polypeptide A2 /// UDP
		Ugt1a6a	glucuronosyltransferase 1 family, polypeptide A6A /// UDP
		Ugt1a10	glycosyltransferase 1 family, polypeptide A10 /// UDP
		Ugt1a7c	glucuronosyltransferase 1 family, polypeptide A7C /// UDP
		Ugt1a5	glucuronosyltransferase 1 family, polypeptide A5 /// UDP
		Ugt1a9	glucuronosyltransferase 1 family, polypeptide A9 /// UDP

		Ugt1a6b	glucuronosyltransferase 1 family, polypeptide A6B /// UDP
		Ugt1a1	glucuronosyltransferase 1 family, polypeptide A1
0.12	2.21	---	Avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog (Maf), mRNA
0.13	1.95	*Malat1*	metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)
0.16	2.36	*Cyr61*	cysteine rich protein 61
0.17	1.92	*Tgfbi*	transforming growth factor, beta induced
0.19	1.88	*Tgm2*	transglutaminase 2, C polypeptide
0.2	1.86	*Hdac11*	histone deacetylase 11
0.19	1.93	*Pik3r5*	phosphoinositide-3-kinase, regulatory subunit 5, p101
0.21	1.82	*Hsd3b7*	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 7
0.2	1.85	*Ntn1*	netrin 1
0.23	1.85	*Ctsb*	cathepsin B
0.22	1.83	*Tgfbi*	transforming growth factor, beta induced
0.22	1.88	*Emb*	embigin
0.22	1.8	*Gnpda2*	glucosamine-6-phosphate deaminase 2
0.23	1.87	---	---
0.22	1.82	*Masp1*	mannan-binding lectin serine peptidase 1
0.23	1.82	*Aldh2*	aldehyde dehydrogenase 2, mitochondrial
0.23	1.8	*Hexb*	hexosaminidase B
0.23	1.84	*Trp53inp1*	transformation related protein 53 inducible nuclear protein 1
0.23	1.83	*Wdr34*	WD repeat domain 34
0.24	1.92	*C3*	complement component 3
	1.78	*Ugt1a2 ///*	UDP glucuronosyltransferase 1 family, polypeptide A2 /// UDP
0.24		*Ugt1a6a ///*	glucuronosyltransferase 1 family, polypeptide A6A /// UDP
		Ugt1a10 ///	glycosyltransferase 1 family, polypeptide A10 /// UDP
		Ugt1a7c ///	glucuronosyltransferase 1 family, polypeptide A7C /// UDP
		Ugt1a5 ///	glucuronosyltransferase 1 family, polypeptide A5 /// UDP
		Ugt1a9 ///	glucuronosyltransferase 1 family, polypeptide A9 /// UDP
		Ugt1a6b ///	glucuronosyltransferase 1 family, polypeptide A6B /// UDP
		Ugt1a1	glucuronosyltransferase 1 family, polypeptide A1
0.24	1.79	*2810013C04Rik*	RIKEN cDNA 2810013C04 gene
0.24	1.8	*2610003J06Rik*	RIKEN cDNA 2610003J06 gene
0.25	1.85	*Sulf2*	sulfatase 2
0.24	1.99	*H3f3b*	H3 histone, family 3B
0.27	1.89	*1110013L07Rik*	RIKEN cDNA 1110013L07 gene
0.29	1.77	*Gpc1*	glypican 1
0.32	1.86	*Parp12*	poly (ADP-ribose) polymerase family, member 12
0.32	1.87	*C1s*	complement component 1, s subcomponent
0.34	1.78	*Vegfc*	vascular endothelial growth factor C
0.41	1.78	*Evi2a*	ecotropic viral integration site 2a
0.42	1.75	*Neu1*	neuraminidase 1
0.42	1.76	*Vegfc*	vascular endothelial growth factor C
0.43	1.73	*Nt5e*	5' nucleotidase, ecto
0.47	1.77	*Hs6st2*	heparan sulfate 6-O-sulfotransferase 2
0.48	1.73	*Gsto2*	glutathione S-transferase omega 2
0.47	1.74	*Grina*	glutamate receptor, ionotropic, N-methyl D-asparate-associated protein 1 (glutamate binding)
0.47	1.72	*Cxcl5*	chemokine (C-X-C motif) ligand 5
0.48	1.72	*Htra1*	HtrA serine peptidase 1
0.49	1.74	*Vat1*	vesicle amine transport protein 1 homolog (T californica)
0.5	1.73	*Gstm1*	glutathione S-transferase, mu 1
0.51	1.71	*H1f0*	H1 histone family, member 0
0.51	1.71	*Cryzl1*	crystallin, zeta (quinone reductase)-like 1
0.5	1.7	*Rassf5*	Ras association (RalGDS/AF-6) domain family 5
0.5	1.73	*Jarid1d*	jumonji, AT rich interactive domain 1D (Rbp2 like)
0.53	1.7	*Mgst2*	microsomal glutathione S-transferase 2
0.54	1.71	*Ddit3*	DNA-damage inducible transcript 3
0.56	1.7	*Psap*	prosaposin
0.57	1.87	*AI256396*	EST AI256396
0.57	1.77	*Xdh*	xanthine dehydrogenase
0.59	1.76	*Cbr2*	carbonyl reductase 2

0.59	1.79	*1500005K14Rik*	RIKEN cDNA 1500005K14 gene
0.58	2.04	*Sepp1*	selenoprotein P, plasma, 1
0.59	1.69	*Them2*	thioesterase superfamily member 2
0.59	1.7	*Psap*	prosaposin
0.6	1.72	*Tgm2*	transglutaminase 2, C polypeptide
0.59	1.7	*Ctsl*	cathepsin L
0.6	1.82	*B230342M21Rik*	RIKEN cDNA B230342M21 gene
0.61	1.83	*Gdpd1*	glycerophosphodiester phosphodiesterase domain containing 1
0.61	1.76	*Igfbp3*	insulin-like growth factor binding protein 3
0.63	1.68	*Ii*	Ia-associated invariant chain
0.63	1.68	*Htra1*	HtrA serine peptidase 1
0.68	1.67	*Gas2*	Growth arrest specific 2, mRNA (cDNA clone MGC:18565 IMAGE:4237356)
0.72	1.74	*B430320C24Rik*	RIKEN cDNA B430320C24 gene
0.72	1.7	*D430039N05Rik*	RIKEN cDNA D430039N05 gene
0.71	1.7	*Ctso*	cathepsin O
0.72	1.72	*Rapgef3*	Rap guanine nucleotide exchange factor (GEF) 3
0.71	2.04	*Abca1*	ATP-binding cassette, sub-family A (ABC1), member 1
0.75	1.66	*Pdlim1*	PDZ and LIM domain 1 (elfin)
0.79	1.66	*Rnf157*	ring finger protein 157
0.79	1.69	*Gstm1*	glutathione S-transferase, mu 1
0.85	1.73	*Igfbp3*	insulin-like growth factor binding protein 3
0.85	1.96	---	---
0.84	1.7	*Pscdbp*	pleckstrin homology, Sec7 and coiled-coil domains, binding protein
0.84	1.69	*Ctsb*	cathepsin B
0.84	1.64	---	---
0.85	1.73	*Ly6a*	lymphocyte antigen 6 complex, locus A
0.84	1.68	*Tgfbi*	transforming growth factor, beta induced
0.86	1.7	*Ctsb*	cathepsin B
0.88	1.67	*Vegfc*	vascular endothelial growth factor C
0.96	1.66	*Cnp1*	cyclic nucleotide phosphodiesterase 1
0.98	1.64	*Gnpda2*	glucosamine-6-phosphate deaminase 2
1.01	1.63	*Napb*	N-ethylmaleimide sensitive fusion protein attachment protein beta
1.05	1.62	*Atp6ap2*	ATPase, H+ transporting, lysosomal accessory protein 2
1.05	1.63	*Laptm4b*	lysosomal-associated protein transmembrane 4B
1.06	1.68	*Itgb7*	integrin beta 7
1.05	1.61	*Bphl*	biphenyl hydrolase-like (serine hydrolase, breast epithelial mucin-associated antigen)
1.06	1.81	*Fzd10*	frizzled homolog 10 (Drosophila)
1.05	1.69	*Phex*	phosphate regulating gene with homologies to endopeptidases on the X chromosome (hypophosphatemia, vitamin D resistant rickets)
1.05	1.63	*A930001N09Rik*	RIKEN cDNA A930001N09 gene (A930001N09Rik), mRNA
1.05	1.64	*Ptgs1*	prostaglandin-endoperoxide synthase 1
1.06	1.65	*Klhl24*	kelch-like 24 (Drosophila)
1.06	1.65	*Sh2d5*	SH2 domain containing 5
1.06	1.66	*Tgm2*	transglutaminase 2, C polypeptide
1.06	1.62	*P2rx4*	purinergic receptor P2X, ligand-gated ion channel 4
1.06	1.63	*Plxnd1*	plexin D1
1.07	1.62	*Malat1*	metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)
1.09	1.61	*Hrsp12*	heat-responsive protein 12
1.1	1.62	*Dcxr*	dicarbonyl L-xylulose reductase
1.1	1.64	*Trim44*	tripartite motif-containing 44
1.11	1.65	*Rassf5*	Ras association (RalGDS/AF-6) domain family 5
1.16	1.59	*Klhl24*	kelch-like 24 (Drosophila)
1.15	1.68	*Dhrs6*	dehydrogenase/reductase (SDR family) member 6
1.15	1.26	*Adm*	adrenomedullin
1.17	1.6	*Kif21a*	kinesin family member 21A
1.19	1.62	*Cst6*	cystatin E/M

1.2	1.19	*Adm*	adrenomedullin
1.22	1.6	*Bbs7*	Bardet-Biedl syndrome 7
1.26	1.59	*Scara5*	scavenger receptor class A, member 5 (putative)
1.27	1.59	*A630026N12Rik*	RIKEN cDNA A630026N12 gene
1.39	1.58	*Ugt1a2 ///*	UDP glucuronosyltransferase 1 family, polypeptide A2 /// UDP
		Ugt1a6a ///	glucuronosyltransferase 1 family, polypeptide A6A /// UDP
		Ugt1a10 ///	glycosyltransferase 1 family, polypeptide A10 /// UDP
		Ugt1a7c ///	glucuronosyltransferase 1 family, polypeptide A7C /// UDP
		Ugt1a5 ///	glucuronosyltransferase 1 family, polypeptide A5 /// UDP
		Ugt1a9 ///	glucuronosyltransferase 1 family, polypeptide A9 /// UDP
		Ugt1a6b ///	glucuronosyltransferase 1 family, polypeptide A6B /// UDP
		Ugt1a1	glucuronosyltransferase 1 family, polypeptide A1
1.38	1.59	*Hspa4l*	heat shock 70kDa protein 4 like
1.4	1.65	*Trappc6a*	trafficking protein particle complex 6A
1.4	1.65	*Emb*	embigin
1.39	1.61	*Itgb2*	integrin beta 2
1.39	1.64	*Tcn2*	transcobalamin 2
1.41	1.59	*Enah*	Enabled homolog (Drosophila) (Enah), mRNA
1.4	1.63	*Phyhd1 /// Lrrc8*	phytanoyl-CoA dioxygenase domain containing 1 /// leucine rich repeat containing 8
1.45	1.62	*Klhl24*	kelch-like 24 (Drosophila)
1.45	1.66	*Capn6*	calpain 6
1.45	1.65	*1810011O10Rik*	RIKEN cDNA 1810011O10 gene
1.5	1.62	*5730457F11Rik*	RIKEN cDNA 5730457F11 gene
1.49	1.59	*Msln*	mesothelin
1.51	1.64	*Uhrf2*	ubiquitin-like, containing PHD and RING finger domains 2
1.54	1.59	*0610031J06Rik*	RIKEN cDNA 0610031J06 gene
1.55	1.62	*Sirt5*	sirtuin 5 (silent mating type information regulation 2 homolog) 5 (S. cerevisiae)
1.58	1.57	*6230421P05Rik*	RIKEN cDNA 6230421P05 gene
1.59	1.57	*Bhlhb5*	basic helix-loop-helix domain containing, class B5
1.65	1.56	*Ghitm*	growth hormone inducible transmembrane protein
1.67	1.63	*2810003C17Rik*	RIKEN cDNA 2810003C17 gene
1.7	1.62	*Gsta4*	glutathione S-transferase, alpha 4
1.73	1.56	*Ndrg4*	N-myc downstream regulated gene 4
1.72	1.57	*Dbp*	D site albumin promoter binding protein
1.72	1.58	*Creg1*	cellular repressor of E1A-stimulated genes 1
1.73	1.56	*Atp6ap2*	ATPase, H+ transporting, lysosomal accessory protein 2
1.78	1.61	*1110046J04Rik*	RIKEN cDNA 1110046J04 gene
1.78	1.57	*MGI:2446326*	suprabasin
1.88	1.57	*5730469M10Rik*	RIKEN cDNA 5730469M10 gene
1.87	1.59	*Jarid1b*	jumonji, AT rich interactive domain 1B (Rbp2 like)
1.93	1.72	*LOC231914*	similar to mKIAA0060 protein
1.95	1.6	*---*	Expressed sequence R75581 (R75581), mRNA
1.94	1.58	*9130227C08Rik*	RIKEN cDNA 9130227C08 gene
1.93	1.64	*Hist1h2bc*	histone 1, H2bc
1.96	1.3	*Ccng2*	cyclin G2
1.96	1.59	*Hba-a1*	hemoglobin alpha, adult chain 1
2.04	1.59	*Matn2*	matrilin 2
2.08	1.57	*D330001F17Rik*	RIKEN cDNA D330001F17 gene
2.14	1.53	*Stch*	stress 70 protein chaperone, microsome-associated, human homolog
2.19	1.55	*Sema3c*	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C
2.22	1.57	*Atp6v0b*	ATPase, H+ transporting, V0 subunit B
2.25	1.58	*Sema3a*	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3A
2.24	1.61	*Phyh*	phytanoyl-CoA hydroxylase
2.33	1.6	*Grina*	glutamate receptor, ionotropic, N-methyl D-asparate-associated protein 1 (glutamate binding)
2.37	1.62	*Foxf1a*	forkhead box F1a
2.42	1.54	*Malat1*	metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)
2.53	1.67	*Enpp2*	ectonucleotide pyrophosphatase/phosphodiesterase 2

2.58	1.54	*Aldh6a1*	aldehyde dehydrogenase family 6, subfamily A1
2.59	1.53	*Hexb*	hexosaminidase B
2.6	1.55	*Dpp7*	dipeptidylpeptidase 7
2.61	1.53	---	Adult male medulla oblongata cDNA, RIKEN full-length enriched library, clone:6330544N02 product:unclassifiable, full insert sequence
2.61	1.57	---	Transcribed locus
2.61	1.52	*Tspyl4*	TSPY-like 4
2.67	1.63	*Ahr*	aryl-hydrocarbon receptor
2.71	1.55	*2510042P03Rik*	RIKEN cDNA 2510042P03 gene
2.7	1.54	*9130022K13Rik*	RIKEN cDNA 9130022K13 gene
2.69	1.54	*Laptm4b*	lysosomal-associated protein transmembrane 4B
2.71	1.51	*Igfbp6*	insulin-like growth factor binding protein 6
2.8	1.51	*Abhd4*	abhydrolase domain containing 4
2.8	1.5	*4921509J17Rik*	RIKEN cDNA 4921509J17 gene
2.82	1.79	*Aldh3a1*	aldehyde dehydrogenase family 3, subfamily A1
2.83	1.51	*1110003E01Rik*	RIKEN cDNA 1110003E01 gene
2.94	1.53	*Atp6ap2*	ATPase, H+ transporting, lysosomal accessory protein 2
3.02	1.29	*Ccng2*	cyclin G2
3.09	1.53	*D5Ertd593e*	DNA segment, Chr 5, ERATO Doi 593, expressed
3.1	1.61	*Dpysl3*	dihydropyrimidinase-like 3
3.11	1.51	*Serpinb6a*	Serine (or cysteine) peptidase inhibitor, clade B, member 6a, mRNA (cDNA clone MGC:6042 IMAGE:3481963)
3.18	1.53	*Parvb*	Parvin, beta (Parvb), mRNA
3.2	1.54	*Zhx1*	zinc fingers and homeoboxes protein 1
3.24	1.54	*Creg1*	cellular repressor of E1A-stimulated genes 1
3.25	1.6	*Vat1*	vesicle amine transport protein 1 homolog (T californica)
3.25	1.53	*MGI:2446326*	suprabasin
3.26	1.53	---	---
3.43	1.51	*5133401H06Rik*	RIKEN cDNA 5133401H06 gene
3.46	1.49	*Gba*	glucosidase, beta, acid
3.56	1.56	*1200016E24Rik* /// *A130040M12Rik* /// *E430024C06Rik* /// *LOC433071*	RIKEN cDNA 1200016E24 gene /// RIKEN cDNA A130040M12 gene /// RIKEN cDNA E430024C06 gene /// hypothetical gene supported by AK004796; BC040222
3.61	1.49	*U2af1-rs1*	U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) 1, related sequence 1
3.59	1.53	*Zc3h6*	zinc finger CCCH type containing 6
3.58	1.51	*5031439G07Rik*	RIKEN cDNA 5031439G07 gene
3.6	1.5	*Igfbp4*	insulin-like growth factor binding protein 4
3.59	1.51	*Igfbp4*	insulin-like growth factor binding protein 4
3.59	1.5	*Igfbp4*	insulin-like growth factor binding protein 4
3.58	1.51	*Ghitm*	growth hormone inducible transmembrane protein
3.66	1.5	*Acpl2*	acid phosphatase-like 2
3.79	1.56	*2010309G21Rik* /// *LOC207685* /// *LOC547243*	RIKEN cDNA 2010309G21 gene /// hypothetical protein LOC207685 /// similar to Ig lambda-2 chain
3.78	1.53	*1810011O10Rik*	RIKEN cDNA 1810011O10 gene
3.79	1.49	---	---
3.81	1.49	*Tpp1*	tripeptidyl peptidase I
3.99	1.55	*Rapgef3*	Rap guanine nucleotide exchange factor (GEF) 3
4.01	1.48	*Echdc3*	enoyl Coenzyme A hydratase domain containing 3
4.06	1.49	*Il1rl1*	interleukin 1 receptor-like 1
4.05	1.52	*Irg1*	immunoresponsive gene 1
4.03	1.47	*Irf2bp2*	interferon regulatory factor 2 binding protein 2
4.06	1.49	*Tob1*	transducer of ErbB-2.1
4.06	1.52	*6030443O07Rik*	RIKEN cDNA 6030443O07 gene
4.06	1.47	*Gla*	galactosidase, alpha
4.12	1.49	*Dpp7*	dipeptidylpeptidase 7
4.12	1.48	*1110003E01Rik*	RIKEN cDNA 1110003E01 gene
4.12	1.5	*Ndrg4*	N-myc downstream regulated gene 4
4.28	1.48	*Rnase4*	ribonuclease, RNase A family 4
4.29	1.49	---	---
4.28	1.48		

4.27	1.34	*Hist1h1c*	histone 1, H1c
4.48	1.51	*Steap2*	PREDICTED: six transmembrane epithelial antigen of prostate 2 [Mus musculus], mRNA sequence
4.49	1.47	*Bcl2l11*	BCL2-like 11 (apoptosis facilitator)
4.48	1.47	*Decr1*	2,4-dienoyl CoA reductase 1, mitochondrial
4.5	1.48	---	Transcribed locus
4.5	1.53	*Arrdc3*	arrestin domain containing 3
4.5	1.51	*Gstm1 /// LOC433943 /// LOC436518 /// LOC547390*	glutathione S-transferase, mu 1 /// similar to Glutathione S-transferase Mu 1 (GST class-mu 1) (Glutathione S-transferase GT8.7) (pmGT10) (GST 1-1) /// similar to Glutathione S-transferase Mu 1 (GST class-mu 1) (Glutathione S-transferase GT8.7) (pmGT10) (GST 1-1) /// similar to Glutathione S-transferase Mu 1 (GST class-mu 1) (Glutathione S-transferase GT8.7) (pmGT10) (GST 1-1)
4.49	1.57	*Uty /// LOC546404 /// LOC546411*	ubiquitously transcribed tetratricopeptide repeat gene, Y chromosome /// similar to male-specific histocompatibility antigen H-YDb /// similar to male-specific histocompatibility antigen H-YDb
4.59	1.47	*Capg*	capping protein (actin filament), gelsolin-like
4.58	1.47	*LOC546143*	hypothetical protein LOC546143
4.59	1.5	*B930075F07*	hypothetical protein B930075F07
4.6	1.49	*Atp6v0b*	ATPase, H+ transporting, V0 subunit B
4.62	1.49	*2310047A01Rik*	RIKEN cDNA 2310047A01 gene
4.65	1.47	*Hspa4l*	heat shock 70kDa protein 4 like
4.63	1.49	*2410006H16Rik*	RIKEN cDNA 2410006H16 gene
4.62	1.47	*1700022C21Rik*	RIKEN cDNA 1700022C21 gene
4.69	1.5	*Gdpd1*	glycerophosphodiester phosphodiesterase domain containing 1
4.89	1.56	*Saa3*	serum amyloid A 3
4.87	1.54	*Zfand2a*	zinc finger, AN1-type domain 2A
4.86	1.36	*Eno2*	enolase 2, gamma neuronal
4.85	1.46	*Gba*	glucosidase, beta, acid
4.84	1.49	*Foxg1*	forkhead box G1
4.82	1.47	*Ccnt2*	cyclin T2
4.82	1.31	*Ddit4*	DNA-damage-inducible transcript 4
4.84	1.49	---	---
4.85	1.5	*Gsto1*	glutathione S-transferase omega 1
4.85	1.46	*Atg10*	autophagy-related 10 (yeast)
4.86	1.47	*P2rx4*	purinergic receptor P2X, ligand-gated ion channel 4
4.85	1.49	*Mitf*	microphthalmia-associated transcription factor
4.87	1.48	*Ldh2*	lactate dehydrogenase 2, B chain
4.85	1.61	*2310043N10Rik*	RIKEN cDNA 2310043N10 gene
4.89	1.45	*Dstn*	Destrin (Dstn), mRNA
4.89	1.45	*B230315F11Rik*	RIKEN cDNA B230315F11 gene
4.88	1.5	*1810015C04Rik*	RIKEN cDNA 1810015C04 gene
4.87	1.45	*Scpep1*	serine carboxypeptidase 1
4.96	1.48	*Dusp1*	dual specificity phosphatase 1

Table 15 Genes downregulated; pLoxEGFP250+ *vs.* pLoxEGFP5-.

FDR	Fold Change	Gene Symbol	Gene Title
0	-2.28	*Has2*	hyaluronan synthase 2
0	-2.06	*6530401D17Rik*	RIKEN cDNA 6530401D17 gene
0	-2.02	*5930437A14Rik*	RIKEN cDNA 5930437A14 gene
0	-1.93	*Has2*	hyaluronan synthase 2
0	-1.89	*X83313*	EST X83313
0	-1.93	*Dlk1*	delta-like 1 homolog (Drosophila)
0	-1.87	*Hivep3*	human immunodeficiency virus type I enhancer binding protein 3
0	-1.96	*2010012C16Rik*	RIKEN cDNA 2010012C16 gene

0	-2.09	*AI662270*	expressed sequence AI662270
0.1	-1.75	*Grem2*	gremlin 2 homolog, cysteine knot superfamily (Xenopus laevis)
0.09	-1.76	*B230120H23Rik*	RIKEN cDNA B230120H23 gene
0.08	-1.94	*AI662270*	expressed sequence AI662270
0.08	-1.78	*Dcp2*	DCP2 decapping enzyme homolog (S. cerevisiae)
0.07	-1.84	*Sp100*	nuclear antigen Sp100
0.33	-1.82	*Oas1a*	2'-5' oligoadenylate synthetase 1A
0.31	-1.75	*Asb4*	ankyrin repeat and SOCS box-containing protein 4
0.35	-1.89	*Egr3*	early growth response 3
0.61	-1.72	*Tyki*	thymidylate kinase family LPS-inducible member
0.58	-1.66	*Kbtbd8*	kelch repeat and BTB (POZ) domain containing 8
0.9	-1.72	*Igf2bp3*	insulin-like growth factor 2, binding protein 3
1.14	-1.71	*Mlf1*	myeloid leukemia factor 1
1.09	-1.65	---	---
1.04	-1.7	---	---
1.08	-1.69	*Sp100*	nuclear antigen Sp100
1.04	-1.51	*Dusp9*	dual specificity phosphatase 9
1.08	-1.6	*Isg20l1*	interferon stimulated exonuclease gene 20-like 1
1.22	-1.64	*5730596K20Rik*	RIKEN cDNA 5730596K20 gene
1.18	-1.7	*Lphn2*	latrophilin 2
1.17	-1.69	*Adamtsl1*	ADAMTS-like 1
1.2	-1.67	*Odz3*	odd Oz/ten-m homolog 3 (Drosophila)
1.16	-1.62	*Adam12*	a disintegrin and metallopeptidase domain 12 (meltrin alpha)
1.19	-1.61	*Thbs1*	thrombospondin 1
1.18	-1.64	*Tmem47*	transmembrane protein 47
1.26	-1.58	---	RNA transcript from U17 small nucleolar RNA host gene
1.26	-1.86	*Iigp1*	interferon inducible GTPase 1
1.28	-1.72	*Iigp1*	interferon inducible GTPase 1
1.24	-1.57	*Mcm5*	minichromosome maintenance deficient 5, cell division cycle 46 (S. cerevisiae)
1.21	-1.6	*Samd9l*	sterile alpha motif domain containing 9-like
1.26	-1.64	*Cacna1c*	Calcium channel, voltage-dependent, L type, alpha 1C subunit (Cacna1c), mRNA
1.27	-1.57	*Egfl9*	EGF-like-domain, multiple 9
1.29	-1.91	*Serpine1*	serine (or cysteine) peptidase inhibitor, clade E, member 1
1.4	-1.57	*Thbs1*	thrombospondin 1
1.53	-1.63	*8030463A06Rik*	RIKEN cDNA 8030463A06 gene (8030463A06Rik), mRNA
1.52	-1.55		PREDICTED: similar to protease [Mus musculus], mRNA sequence
1.51	-1.62	*Iigp2*	interferon inducible GTPase 2
1.54	-1.62	*Tnc*	tenascin C
1.51	-1.56	*Cdh11*	cadherin 11
1.67	-1.61	*Id2*	inhibitor of DNA binding 2
1.86	-1.55	---	---
1.92	-1.53	*Hnrpdl*	heterogeneous nuclear ribonucleoprotein D-like
2.29	-1.52	*Ddx49*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 49
2.48	-1.52	*2310001H13Rik*	RIKEN cDNA 2310001H13 gene
2.53	-1.59	*2900073C17Rik*	RIKEN cDNA 2900073C17 gene
2.52	-1.57	---	Transcribed locus
2.49	-1.59	*Tgtp*	T-cell specific GTPase
2.8	-1.52	---	---
2.88	-1.5	*Tbrg4*	transforming growth factor beta regulated gene 4
2.91	-1.51	*Flnc*	filamin C, gamma (actin binding protein 280)
2.95	-1.64	---	Transcribed locus
3.1	-1.58	*Ror1*	receptor tyrosine kinase-like orphan receptor 1
3.05	-1.53	*Oasl2*	2'-5' oligoadenylate synthetase-like 2
3.47	-1.47	*D19Bwg1357e*	DNA segment, Chr 19, Brigham & Women's Genetics 1357 expressed
3.63	-1.49	*5033421C21Rik*	RIKEN cDNA 5033421C21 gene
3.73	-1.54	*Slfn8*	schlafen 8
3.83	-1.51	*Thbs1*	thrombospondin 1
3.83	-1.49	*Mybl2*	myeloblastosis oncogene-like 2
3.99	-1.43	*Pa2g4*	proliferation-associated 2G4

4.24	-1.49	*Angpt1*	angiopoietin 1
4.3	-1.48	*MGI:1933403*	type 1 tumor necrosis factor receptor shedding aminopeptidase regulator
4.31	-1.48	*Cdc6*	cell division cycle 6 homolog (S. cerevisiae)
4.25	-1.61	*Igtp*	interferon gamma induced GTPase
4.38	-1.57	*Sp100-rs*	similar to component of Sp100-rs
4.36	-1.56	*Psip1*	PC4 and SFRS1 interacting protein 1
4.85	-1.55	*Ifi44*	interferon-induced protein 44
4.85	-1.26	*AA408556*	expressed sequence AA408556
4.96	-1.49	*C1ql3*	C1q-like 3
4.99	-1.6	*Hmgcs1*	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA (cDNA clone MGC:36620 IMAGE:5347038)
4.94	-1.41	*Usp18*	ubiquitin specific peptidase 18
4.94	-1.63	*Hey1*	hairy/enhancer-of-split related with YRPW motif 1

FDR, False discovery rate; **---**, unannotated cDNA or EST sequences.

Table 16 Genes up-regulated; pLoxEGFP5+ *vs.* pLoxEGFP5-.

FDR	Fold Change	Gene Symbol	Gene Title
0	3.59	---	---
0	3.44	*1110032E23Rik*	RIKEN cDNA 1110032E23 gene
0	3.13	*Gsta1 /// Gsta2*	glutathione S-transferase, alpha 1 (Ya) /// glutathione S-transferase, alpha 2 (Yc2)
0	2.56	*Sdpr*	serum deprivation response
0	2.52	*Sdpr*	serum deprivation response
0	2.38	*Sdpr*	serum deprivation response
0	2.4	*Car6*	carbonic anhydrase 6
0	2.32	*Tmsb4x*	thymosin, beta 4, X chromosome
0	2.21	*Alcam*	activated leukocyte cell adhesion molecule
0	1.94	*Gadd45a*	growth arrest and DNA-damage-inducible 45 alpha
0	1.97	*9430093N24Rik*	MKIAA1983 protein
0	1.88	*Dpysl3*	dihydropyrimidinase-like 3
0	1.86	*Nqo1*	NAD(P)H dehydrogenase, quinone 1
0	1.91	*Ddx3y*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked
0	1.93	*Sqrdl*	sulfide quinone reductase-like (yeast)
0	1.8	*Ccl2*	chemokine (C-C motif) ligand 2
0	1.85	*LOC433022*	hypothetical LOC433022
0	1.96	*D3Ertd300e*	DNA segment, Chr 3, ERATO Doi 300, expressed
0	1.79	*Grem1*	gremlin 1
0	1.75	*Plf /// Plf2 /// Mrpplf3*	proliferin /// proliferin 2 /// mitogen regulated protein, proliferin 3
0	1.74	*Fyb*	FYN binding protein
0	1.8	*2210419I08Rik*	RIKEN cDNA 2210419I08 gene
0	1.77	*Alcam*	activated leukocyte cell adhesion molecule
0	1.73	*Ptprc*	Alternatively spliced Ly-5 glycoprotein mRNA, 5' end
0	1.73	*Gatm*	glycine amidinotransferase (L-arginine:glycine amidinotransferase)
0	1.75	*Dub1*	deubiquitinating enzyme 1
0	1.7	*Lynx1*	Ly6/neurotoxin 1
0	1.68	*Ass1*	argininosuccinate synthetase 1
0	1.71	*U90926*	cDNA sequence U90926
0.03	1.76	*Alcam*	activated leukocyte cell adhesion molecule
0.03	1.62	*2010309G21Rik /// LOC207685 /// LOC547243*	RIKEN cDNA 2010309G21 gene /// hypothetical protein LOC207685 /// similar to Ig lambda-2 chain
0.03	1.66	*H60*	PREDICTED: histocompatibility 60 [Mus musculus], mRNA sequence
0.03	1.72	*Tiparp*	TCDD-inducible poly(ADP-ribose) polymerase
0.06	1.64	*LOC433022*	hypothetical LOC433022

0.06	1.62	*Cyba*	cytochrome b-245, alpha polypeptide
0.06	1.64	*A130040M12Rik*	RIKEN cDNA A130040M12 gene
0.05	1.69	*Hist1h2bc*	histone 1, H2bc
0.11	1.72	*Malat1*	metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)
0.13	1.7	*Parp12*	poly (ADP-ribose) polymerase family, member 12
0.12	1.64	*0610040B09Rik*	RIKEN cDNA 0610040B09 gene
0.12	1.68	*Evi2a*	ecotropic viral integration site 2a
0.12	1.6	*Gpnmb*	glycoprotein (transmembrane) nmb
0.12	1.62	*2810013C04Rik*	RIKEN cDNA 2810013C04 gene
0.11	1.59	*Gsta4*	glutathione S-transferase, alpha 4
0.11	1.7	*Kctd4*	potassium channel tetramerisation domain containing 4
0.11	1.62	*Dcn*	decorin
0.11	1.59	*Igf2r*	insulin-like growth factor 2 receptor
0.1	1.63	---	---
0.1	1.6	*Dusp4*	dual specificity phosphatase 4
0.1	1.78	*Rorb*	RAR-related orphan receptor beta
0.1	1.67	*Sfrs2ip*	splicing factor, arginine/serine-rich 2, interacting protein
0.12	1.58	*Mtss1*	metastasis suppressor 1
0.15	1.59	*Rbmy1a1*	RNA binding motif protein, Y chromosome, family 1, member A1
0.15	1.66	*Mep1a*	meprin 1 alpha
0.18	1.63	*3110004L20Rik*	RIKEN cDNA 3110004L20 gene
0.2	1.57	*Ccl7*	chemokine (C-C motif) ligand 7
0.23	1.59	*Mapk12*	mitogen-activated protein kinase 12
0.22	1.56		
0.34	1.58	---	Transcribed locus
0.33	1.58	*Calr3*	calreticulin 3
0.34	1.55	*Seh1l*	SEH1-like (S. cerevisiae
0.4	1.52	*D11Ertd730e*	DNA segment, Chr 11, ERATO Doi 730, expressed
0.41	1.56	*Traf1*	Tnf receptor-associated factor 1
0.44	1.51	---	Transcribed locus
0.46	1.51	*9130227C08Rik*	RIKEN cDNA 9130227C08 gene
0.47	1.52	*E430024C06Rik*	RIKEN cDNA E430024C06 gene
0.46	1.52	*B430320C24Rik*	RIKEN cDNA B430320C24 gene
0.46	1.5	*Myd116*	myeloid differentiation primary response gene 116
0.54	1.53	*Edn1*	endothelin 1
0.56	1.52	*Atf3*	activating transcription factor 3
0.62	1.51	*Gpc1*	glypican 1
0.61	1.51	*BC034507*	cDNA sequence BC034507
0.7	1.49	*Ociad2*	OCIA domain containing 2
		LOC432995 ///	hypothetical gene supported by BC047216 /// hypothetical
0.72	1.52	*LOC436333*	gene supported by BC047216
0.71	1.49	*Vegfc*	vascular endothelial growth factor C
0.76	1.48	*2810003C17Rik*	RIKEN cDNA 2810003C17 gene
0.82	1.48	*Erbb3*	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)
0.81	1.5	*Gadd45b*	growth arrest and DNA-damage-inducible 45 beta
0.81	1.48	*Ampd3*	AMP deaminase 3
0.85	1.51	*Zfp622*	zinc finger protein 622
0.91	1.54	*BC034507*	cDNA sequence BC034507
0.95	1.51	*Hcn1*	hyperpolarization-activated, cyclic nucleotide-gated K+ 1
0.95	1.48	*Sqstm1*	sequestosome 1
0.94	1.47	*Atp6v0b*	ATPase, H+ transporting, V0 subunit B
0.94	1.5	*Foxf1a*	forkhead box F1a
0.98	1.51	*Maf*	avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog
0.99	1.48	*Dazl*	deleted in azoospermia-like
0.98	1.53	*Trim44*	tripartite motif-containing 44
1.04	1.47	*3110023G01Rik*	RIKEN cDNA 3110023G01 gene
1.07	1.48	*Hmox1*	heme oxygenase (decycling) 1
1.05	1.64	*Alcam*	activated leukocyte cell adhesion molecule
1.16	1.5	*Napb*	N-ethylmaleimide sensitive fusion protein attachment protein beta

1.18	1.47	*Eif2s3y*	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked
1.24	1.45	*1200016E24Rik /// A130040M12Rik /// E430024C06Rik /// LOC433071*	RIKEN cDNA 1200016E24 gene /// RIKEN cDNA A130040M12 gene /// RIKEN cDNA E430024C06 gene /// hypothetical gene supported by AK004796; BC040222
1.25	1.47	*Ppifs*	peptidylprolyl isomerase F, opposite strand transcription unit
1.24	1.46	*Ahr*	aryl-hydrocarbon receptor
1.39	1.47	*Zfand2a*	zinc finger, AN1-type domain 2A
1.39	1.57	*Efemp1*	epidermal growth factor-containing fibulin-like extracellular matrix protein 1
1.41	1.44	*Vegfc*	vascular endothelial growth factor C
1.44	1.45	---	---
1.47	1.45	*9430041P20Rik*	RIKEN cDNA 9430041P20 gene
1.63	1.43	*Pik3r5*	phosphoinositide-3-kinase, regulatory subunit 5, p101
1.68	1.45	*Mtss1*	metastasis suppressor 1
1.84	1.44	*Hs6st2*	heparan sulfate 6-O-sulfotransferase 2
1.9	1.47	*LOC231914*	similar to mKIAA0060 protein
1.93	1.47	*2900093K20Rik*	RIKEN cDNA 2900093K20 gene
2.36	1.42	*1110013L07Rik*	RIKEN cDNA 1110013L07 gene
2.41	1.42	*Csf1r*	colony stimulating factor 1 receptor
2.5	1.43	*1110051M20Rik*	RIKEN cDNA 1110051M20 gene
2.61	1.44	---	Transcribed locus
2.69	1.43	*Dusp3*	dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)
2.72	1.41	*Rbm3*	RNA binding motif protein 3
2.84	1.41	*Bhlhb5*	basic helix-loop-helix domain containing, class B5
2.87	1.41	*4930535I16Rik*	RIKEN cDNA 4930535I16 gene
3.29	1.41	*Malat1*	metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)
3.28	1.4	*Stx3*	syntaxin 3
3.96	1.4	---	---
4.17	1.39	*Lmo1*	LIM domain only 1
4.39	1.4	*9130008F23Rik*	RIKEN cDNA 9130008F23 gene

FDR, False discovery rate; **---**, unannotated cDNA or EST sequences.

Table 17 Genes down-regulated; pLoxEGFP5+ *vs.* pLoxEGFP5-.

FDR	Fold Change	Gene Symbol	Gene Title
0	-1.81	*Dlk1*	delta-like 1 homolog (Drosophila)
0	-1.75	*Cacna1c*	Calcium channel, voltage-dependent, L type, alpha 1C subunit (Cacna1c), mRNA
0	-1.78	*6530401D17Rik*	RIKEN cDNA 6530401D17 gene
0	-1.71	*Tomm7*	translocase of outer mitochondrial membrane 7 homolog (yeast)
0	-1.7	*Sc4mol*	Sterol-C4-methyl oxidase-like, mRNA (cDNA clone MGC:11745 IMAGE:3152545)
0	-1.68	*Lcn2*	lipocalin 2
0	-1.78	*Car9*	carbonic anhydrase 9
0	-1.74	*Hivep3*	human immunodeficiency virus type I enhancer binding protein 3
0	-1.65	*Odz3*	odd Oz/ten-m homolog 3 (Drosophila)
0	-1.68	*Asb4*	ankyrin repeat and SOCS box-containing protein 4
0	-1.69	*Ogn*	osteoglycin
0	-1.66	*Haghl*	hydroxyacylglutathione hydrolase-like
0.31	-1.58	*Gpc6*	glypican 6
0.29	-1.62	*Fos*	FBJ osteosarcoma oncogene
0.27	-1.61	*Lancl2*	LanC (bacterial lantibiotic synthetase component C)-like 2

0.25	-1.64	*5930437A14Rik*	RIKEN cDNA 5930437A14 gene
0.29	-1.6	---	---
0.28	-1.66	*Col3a1*	procollagen, type III, alpha 1
0.89	-1.57	*AI325464*	Expressed sequence AI325464, mRNA (cDNA clone MGC:30834 IMAGE:4006813)
0.95	-1.48	*Igf2bp3*	insulin-like growth factor 2, binding protein 3
1	-1.6	*X83313*	EST X83313
1.05	-1.57	*Nfix*	nuclear factor I/X
1	-1.55	---	Transcribed locus
1.08	-1.55	*Hnrpab*	heterogeneous nuclear ribonucleoprotein A/B
1.16	-1.53	*Ror1*	receptor tyrosine kinase-like orphan receptor 1
1.12	-1.55	*Psip1*	PC4 and SFRS1 interacting protein 1
1.11	-1.55	*Apbb2*	Amyloid beta (A4) precursor protein-binding, family B, member 2, mRNA (cDNA clone MGC:100042 IMAGE:30550087)
1.11	-1.54	*Has2*	hyaluronan synthase 2
1.07	-1.54	---	Transcribed locus
1.13	-1.59	*Rtn4*	RTN4 (Rtn4) mRNA, complete cds, alternatively spliced
1.1	-1.56	*Arf6*	ADP-ribosylation factor 6
1.06	-1.6	*Ogn*	osteoglycin
1.09	-1.53	*Acss2*	acyl-CoA synthetase short-chain family member 2
1.09	-1.58	*Il13ra1*	interleukin 13 receptor, alpha 1
1.17	-1.61	*Nrn1*	neuritin 1
1.28	-1.56	*6330406I15Rik*	RIKEN cDNA 6330406I15 gene
1.32	-1.53	*Tgfb1i4*	TSC22-related inducible leucine zipper 1b (Tilz1b)
1.37	-1.52	*D630048P19Rik*	Optic atrophy 3 (human), mRNA (cDNA clone MGC:106414 IMAGE:6404450)
1.44	-1.56	*Pcbp3*	Poly(rC) binding protein 3 (Pcbp3), mRNA
1.77	-1.5	*D530037H12Rik*	RIKEN cDNA D530037H12 gene
1.78	-1.51	*8430408J07Rik*	RIKEN cDNA 8430408J07 gene
1.93	-1.5	*Il13ra1*	interleukin 13 receptor, alpha 1
2.21	-1.5	*Cp*	ceruloplasmin
2.18	-1.47	*Iigp1*	interferon inducible GTPase 1
2.38	-1.49	*5730555F13Rik*	RIKEN cDNA 5730555F13 gene
2.39	-1.5	*Cog1*	component of oligomeric golgi complex 1
2.68	-1.46	*Mlf1*	myeloid leukemia factor 1
2.67	-1.47	*Ctsb*	Cathepsin B, mRNA (cDNA clone MGC:6211 IMAGE:3500700)
2.61	-1.48	*Zfp289 /// LOC434076*	zinc finger protein 289 /// similar to zinc finger protein 289
2.76	-1.46	*Tmem45a*	transmembrane protein 45a
2.78	-1.49	*H2-T23*	histocompatibility 2, T region locus 23
2.9	-1.48	*Tead1*	TEA domain family member 1 (Tead1), mRNA
2.91	-1.48	*Rpap1*	RNA polymerase II associated protein 1
3.04	-1.47	*Cald1*	Caldesmon 1 (Cald1), mRNA
3.16	-1.47	*Arntl2*	aryl hydrocarbon receptor nuclear translocator-like 2
3.16	-1.46	*Dclre1b*	DNA cross-link repair 1B, PSO2 homolog (S. cerevisiae)
3.25	-1.48	*Asb4*	ankyrin repeat and SOCS box-containing protein 4
3.24	-1.47	*A730054J21Rik*	RIKEN cDNA A730054J21 gene
3.27	-1.41	*2610005L07Rik*	RIKEN cDNA 2610005L07 gene
3.25	-1.47	*6330406I15Rik*	RIKEN cDNA 6330406I15 gene
3.2	-1.46	*Dscr1l1*	Down syndrome critical region gene 1-like 1
3.18	-1.45	*5730596K20Rik*	RIKEN cDNA 5730596K20 gene
3.19	-1.48	*Cp*	ceruloplasmin
3.25	-1.45	*Rgs20*	regulator of G-protein signaling 20
3.2	-1.45	*C1ql3*	C1q-like 3
3.24	-1.46	*Cdh11*	cadherin 11
3.33	-1.45	*Col1a1*	procollagen, type I, alpha 1
3.34	-1.47	---	Transcribed locus
3.29	-1.47	*Cp*	ceruloplasmin
3.24	-1.47	*Cp*	ceruloplasmin
3.32	-1.44	*2010012C16Rik*	RIKEN cDNA 2010012C16 gene
3.47	-1.45	*5033421C21Rik*	RIKEN cDNA 5033421C21 gene
3.42	-1.46	*Chrdl1*	Chordin-like 1 (Chrdl1), mRNA
3.39	-1.46	---	Adult male medulla oblongata cDNA, RIKEN full-length

			enriched library, clone:6330531L09 product:unclassifiable, full insert sequence
3.41	-1.43	*LOC56628*	MHC (A.CA/J(H-2K-f) class I antigen
3.38	-1.41	*Phf20*	PHD finger protein 20
3.36	-1.44	*Ccnc*	cyclin C
3.37	-1.43	*C030045D06Rik*	RIKEN cDNA C030045D06 gene
3.52	-1.46	*Pdlim5*	PDZ and LIM domain 5, mRNA (cDNA clone MGC:46824 IMAGE:4457868)
3.55	-1.44	*Phf3*	PHD finger protein 3
3.79	-1.44	*Psmb9*	proteosome (prosome, macropain) subunit, beta type 9 (large multifunctional peptidase 2)
3.96	-1.43	*Adam12*	a disintegrin and metallopeptidase domain 12 (meltrin alpha)
4.52	-1.44	*Islr*	immunoglobulin superfamily containing leucine-rich repeat
4.5	-1.43	*Pcdhb21*	protocadherin beta 21
4.47	-1.45	*Fmod*	fibromodulin
4.52	-1.43	*Thy1*	thymus cell antigen 1, theta
4.72	-1.43	*Osmr*	oncostatin M receptor
4.76	-1.49	*AI451617*	expressed sequence AI451617
4.83	-1.42	*Scd2*	stearoyl-Coenzyme A desaturase 2
4.98	-1.42	*Adamtsl1*	ADAMTS-like 1

FDR, False discovery rate; **---**, unannotated cDNA or EST sequences.

5.3.3 Exon expression array

In the Affymetrix GeneChip® Mouse Exon 1.0 ST Array, probesets are interspersed evenly throughout the length of the transcript, with at least four probes per exon. At least one probeset is present per exon, and forty probes per transcript on average. The array is very comprehensive, designed to identify exon skipping; intron retention (by placement of probes within introns adjacent to exon-intron boundaries in genes known to undergo events of this type); mutually exclusive exon usage; alternative promoter usage; alternative polyadenylation and alternative splicing donor/acceptor sites with changes over 25 bp (the size of a probe). Exon sequences were compiled from existing cDNA-based information taken from REFSeq and GenBank® mRNAs, and ESTs from dbEST. Predicted exon sequences were included derived from information taken from GENSCAN (genes.mit.edu/GENSCAN.html); Ensembl: (www.ensembl.org/); Vega: (vega.sanger.ac.uk/); geneid and sgp: (www1.imim.es/software/geneid/index.html); TWINSCAN: (genes.cs.wustl.edu/); Exoniphy: (genome.ucsc.edu/cgi-bin/hgTrackUi?g=exoniphy&db=hg16); MicroRNA Registry: (www.sanger.ac.uk/Software/Rfam/mirna/); MITOMAP: (www.mitomap.org/) and Structural RNA Predictions: (genome.ucsc.edu/cgi-bin/hgTrackUi?g=rnaGene&db=hg16). It is unlikely that an alternative exon would not be included within these measures, but if it were not, the exon would not be identified. Cryptic exons such as those of the *chloride channel* in DM however, are likely to be under represented.

5.3.3.1 Sample preparation for exon arrays

Samples were prepared in triplicate, in parallel, such that three chips were hybridised for each sample group pLoxEGFP; pLoxEGFP5 and pLoxEGFP250 –nine chips in total. It has been noted that on occasion cells have been observed that contain foci but do not show EGFP fluorescence when viewed microscopically. Although this was well controlled for by sample choice in the whole transcript array, it was clear from the EGFP RT-PCR (Figure 54B) the quality of separation at the molecular level was poor. Also, recently published data by Mahadevan *et al.* 2006, hinted that over-expression of 5 CUG repeats may be pathogenic (Mahadevan *et al.*, 2006). For these reasons the basic design of the experiment was simplified to pLoxEGFP *vs.* pLoxEGFP5 *vs.* pLoxEGFP250, without cell sorting. This means that if pLoxEGFP5 over-expression is pathogenic, differences should be apparent in comparison to the zero repeat sample pLoxEGFP.

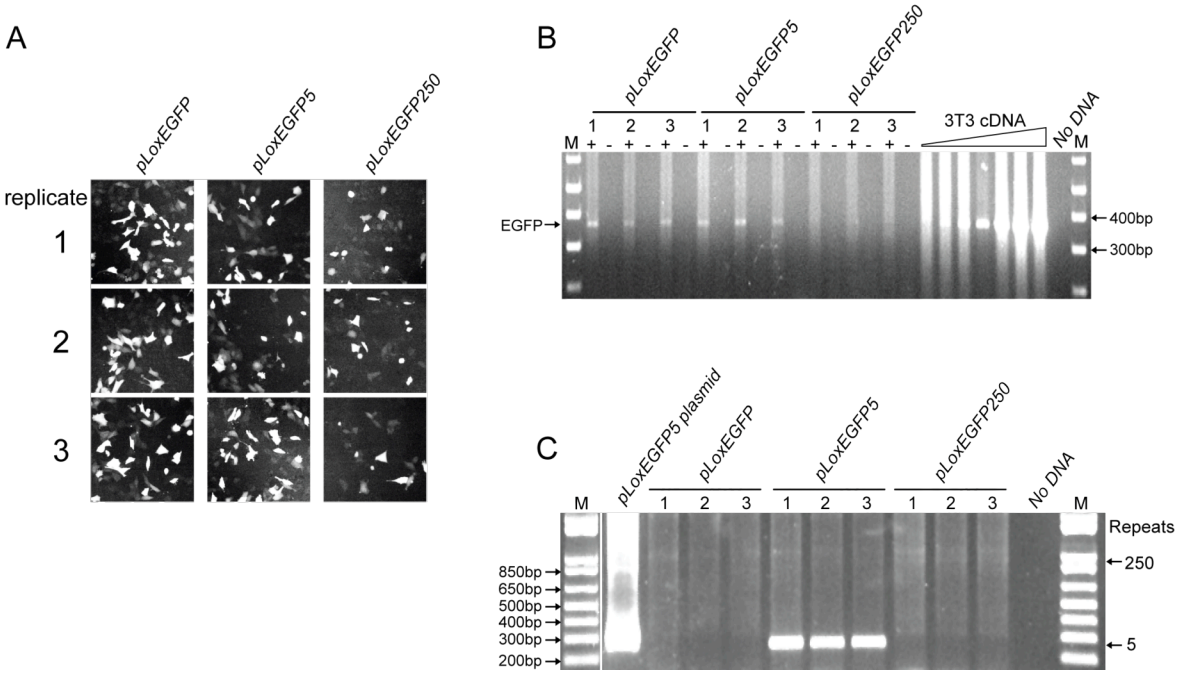


Figure 55 Construct expression in exon array samples. A Fluorescent micrographs. Exon array starting samples. Live 3T3 cells transfected with pLoxEGFP; pLoxEGFP5 or pLoxEGFP250 prior to harvest. Cells were 60% confluent. EGFP fluorescence was apparent in all samples. **B** EGFP RT-PCR. Each sample with and without reverse transcriptase. The EGFP product was present in all samples, faintly in pLoxEGFP250. **C** DM-H/DM-BR RT-PCR over the repeat region. No repeats are apparent in pLoxEGFP replicates; 5 repeats are present in the pLoxEGFP5 samples and a faint presence of 250 repeats can be seen in the pLoxEGFP250 samples. Note that in both EGFP and DM-H/DM-BR RT-PCRs the amount of pLoxEGFP250 products are much lower than pLoxEGFP5, even though EGFP fluorescence in the starting sample is almost as high as pLoxEGFP and pLoxEGFP5 in A.

Cells were transfected as described in 5.3.2.1 ‘Sample preparation for whole transcript arrays’ using pLoxEGFP; pLoxEGFP or pLoxEGFP250. To ensure the cells were always in the exponential phase, after 24 hours, they were harvested using trypsin, diluted 1:3 and incubated for a further 24 hours. Total RNA was prepared and a portion subjected to EGFP RT-PCR to confirm expression of the construct, and DM-H/DM-BR PCR to confirm the repeat length. The results show that the expected repeat length is present in each sample, and that EGFP is also transcribed in each (Figure 55).

The amount of construct expression in both RT-PCRs using the pLoxEGFP250 sample is low compared to pLoxEGFP5. The reasons for this are not clear since EGFP fluorescence is clearly present in the starting sample confocal images, although they are of a lower intensity (Figure 55A). It may be that the RNA within foci is not as easily isolated or reverse-transcribed. Hamshere *et al.* reported that expanded DMPK polyA+ alleles were underrepresented in nuclear polyA+ fractions, suggesting this may be the case (Hamshere *et al.*, 1997), although their method of isolation was based on phenol-chloroform extraction compared to the guanidinium based extraction used here. In the recent reversible mouse model generated by Mahadevan *et al.*, lines with expanded repeats also showed low expression of repeat arrays using quantitative real-time PCR. However, it is not clear which chemistry was used for RNA isolation (Langlois *et al.*, 2003; Mahadevan *et al.*, 2006).

5.3.3.2 Results of analysis

Genes known to be mis-spliced in DM do not always show a large difference or fold change between any two particular RNA isoforms (Kanadia *et al.*, 2006; Lin *et al.*, 2006). In some transcripts a new isoform has appeared, such as the 6b/7a/8 or 6/7a/7/8a/8 exon inclusion of *CLCN1*, expressed at a low level compared to the ‘normal’ 6/7/8 transcript (Figure 65, right). In DM, the proportions of a collection of alternative transcripts are often changed, with some forms only present in DM samples. Therefore the differential splicing analysis was ranked according to *p*-value of alternative splicing (the interaction between the two ANOVA factors -probeset level within the gene and transcript level between the samples with 0, 5 and 250 repeats), rather than a high fold change between transcript isoforms.

GeneChip differences were analysed by 2-way ANOVA analysis using Partek® Genomics Suite software, on three metaprobeset levels, core; extended and full. The core set comprises probes with confirmed cDNAs. The extended set also includes expressed

sequence tags (EST), and the full set comprises confirmed cDNAs; ESTs and predicted splice products. All three lists were studied during gene selection. Significance was limited to a *p*-value of ≤0.001 for alternative exon usage ANOVAs. Between the test samples: pLoxEGFP0; pLoxEGFP5 and pLoxEGFP250 repeats, 29; 54 and 58 genes showed alternative exon usage the core (Table 18), extended (Table 19) and full (Table 20) probesets respectively.

Table 18 Genes with differential exon usage using the core metaprobeset

gene_assignment	*p*-value Alternative splicing
NM_198017 // C430003P19Rik // RIKEN cDNA C430003P19 gene // 7 F3 // 109359 /// A	3.75E-11
NM_015744 // Enpp2 // ectonucleotide pyrophosphatase/phosphodiesterase 2 // 15 D	9.33E-08
NM_027135 // Sec24d // SEC24 related gene family, member D (S. cerevisiae) // 3	2.15E-06
BC029239 // Timm8b // translocase of inner mitochondrial membrane 8 homolog b (y	2.52E-06
NM_008444 // Kif3b // kinesin family member 3B // 2 86.0 cM // 16569 /// AK08154	2.81E-06
XM_975671 // LOC674573 // similar to Mitogen-activated protein kinase kinase kin	3.16E-06
XM_001000465 // 9930012K11Rik // RIKEN cDNA 9930012K11 gene // 14 D2 // 268759	4.55E-06
NM_010764 // Man2b1 // mannosidase 2, alpha B1 // 8 C2|8 37.0 cM // 17159 /// AK	4.58E-06
NM_028066 // F11 // coagulation factor XI // 8 B1.1|8 25.0 cM // 109821	7.81E-06
NM_178224 // Cbs // cystathionine beta-synthase // 17 A-C|17 17.4 cM // 12411 //	8.68E-06
NM_054048 // Rcor2 // REST corepressor 2 // 19 A // 104383 /// AK205817 // Rcor2	9.16E-06
NM_021536 // Rhot1 // ras homolog gene family, member T1 // 11 B5|11 47.26 cM //	9.90E-06
NM_027347 // Crsp3 // cofactor required for Sp1 transcriptional activation, subu	1.27E-05
NM_025982 // Tspan31 // tetraspanin 31 // 10 D3 // 67125 /// AK160269 // Tspan31	1.35E-05
NM_016774 // Atp5b // ATP synthase, H+ transporting mitochondrial F1 complex, be	1.48E-05
NM_146226 // Apeh // acylpeptide hydrolase // 9 F2 // 235606 /// BC034199 // Ape	2.17E-05
XM_897143 // LOC622510 // hypothetical LOC622510 // 2 C3 // 622510 /// XM_922075	2.51E-05
NM_021398 // Slc43a3 // solute carrier family 43, member 3 // // 58207 /// NM_0	2.97E-05
XM_994018 // 2600009E05Rik // RIKEN cDNA 2600009E05 gene // 2 F3 // 77006 /// XM	4.36E-05
NM_001001454 // Ttyh1 // tweety homolog 1 (Drosophila) // 7 A1 // 57776 /// NM_0	4.57E-05
XM_992228 // Epn2 // epsin 2 // 11 B2 // 13855 /// NM_010148 // Epn2 // epsin 2	4.73E-05
NM_023220 // 2010106G01Rik // RIKEN cDNA 2010106G01 gene // 2 F3 // 66552 /// BC	4.94E-05
NM_026127 // 4833420G17Rik // RIKEN cDNA 4833420G17 gene // 13 D2.3 // 67392 ///	5.73E-05
XM_988508 // Bptf/Falz // fetal Alzheimer antigen // 11 E1 // 207165 /// XM_903112 //	6.02E-05
NM_013680 // Syn1 // synapsin I // X A1-A4|X 6.2 cM // 20964	7.13E-05
NM_008782 // Pax5 // paired box gene 5 // 4 B1|4 20.7 cM // 18507	7.66E-05
NM_183315 // Ctxn1 // cortexin 1 // 8 A1.1|8 13.0 cM // 330695	7.87E-05
XM_990359 // Cbx1 // chromobox homolog 1 (Drosophila HP1 beta) // 11 D|11 58.0 c	9.36E-05
BC067070 // 4932415G12Rik // RIKEN cDNA 4932415G12 gene // 10 C2 // 67723	9.39E-05

Genes selected for further bioinformatics analysis are highlighted in red

Table 19 Genes with differential exon usage using the extended metaprobeset

gene_assignment	*p*-value Alternative splicing
NM_133916 // Eif3s9 // eukaryotic translation initiation factor 3, subunit 9 (et	1.41E-08
NM_133215 // Mtmr4 // myotubularin related protein 4 // 11 C // 170749	2.90E-08
NM_198017 // C430003P19Rik // RIKEN cDNA C430003P19 gene // 7 F3 // 109359 /// A	8.71E-08
NM_015744 // Enpp2 // ectonucleotide pyrophosphatase/phosphodiesterase 2 // 15 D	1.50E-07
XR_004056 // LOC382931 // similar to 3-phosphoglycerate dehydrogenase // 14 E2.1	1.80E-07
XM_983388 // Ppp2r5a // protein phosphatase 2, regulatory subunit B (B56), alpha	4.83E-07
NM_013854 // Abcf1 // ATP-binding cassette, sub-family F (GCN20), member 1 // 17	8.97E-07
NM_177615 // C78409 // expressed sequence C78409 // 10 D3 // 216441	9.90E-07
NM_007992 // Fbln2 // fibulin 2 // 6 D-E|6 37.2 cM // 14115 /// AK017870 // 5730	2.27E-06
NM_001039080 // Rbms2 // RNA binding motif, single stranded interacting protein	3.44E-06
XM_001000465 // 9930012K11Rik // RIKEN cDNA 9930012K11 gene // 14 D2 // 268759	5.23E-06
NM_008431 // Kcnk4 // potassium channel, subfamily K, member 4 // 19 A|19 4.5 cM	5.35E-06
NM_019426 // Atf7ip // activating transcription factor 7 interacting protein //	1.14E-05
NM_008782 // Pax5 // paired box gene 5 // 4 B1|4 20.7 cM // 18507	1.38E-05
XM_983442 // Kcnmb4 // potassium large conductance calcium-activated channel, su	1.43E-05
NM_007801 // Ctsh // cathepsin H // 9 E3.1|9 50.0 cM // 13036 /// BC080767 // Ct	1.55E-05
NM_025813 // Mfsd1 // major facilitator superfamily domain containing 1 // 3 E2	1.58E-05
NM_053089 // Narg1 // NMDA receptor-regulated gene 1 // 3 D // 74838 /// AK07804	1.63E-05
AK020953 // B230104C08Rik // RIKEN cDNA B230104C08 gene // 6 G3 // 77841	1.68E-05
NM_008149 // Gpam // glycerol-3-phosphate acyltransferase, mitochondrial // 19 D	1.86E-05
NM_021398 // Slc43a3 // solute carrier family 43, member 3 // // 58207 /// NM_0	2.67E-05
XM_001052047 // Lhx3 // LIM homeobox protein 3 // 2 A2-C1|2 16.0 cM // 16871 ///	2.68E-05
NM_178892 // Tiparp // TCDD-inducible poly(ADP-ribose) polymerase // 3 E1 // 999	3.14E-05
XM_001005692 // LOC673486 // similar to eukaryotic translation initiation factor	3.32E-05
NM_009392 // Tlx2 // T-cell leukemia, homeobox 2 // 6 C3-D1|6 35.5 cM // 21909	3.56E-05
BC029239 // Timm8b // translocase of inner mitochondrial membrane 8 homolog b (y	3.82E-05
NM_172339 // Snapc4 // small nuclear RNA activating complex, polypeptide 4 // 2	4.77E-05
XM_979043 // Gm1027 // gene model 1027, (NCBI) // 4 C7 // 381538 /// NM_028209 /	4.80E-05
XM_994018 // 2600009E05Rik // RIKEN cDNA 2600009E05 gene // 2 F3 // 77006 /// XM	5.13E-05
NM_030109 // Sf3b2 // splicing factor 3b, subunit 2 // 19 A // 319322 /// AK1890	5.14E-05
NM_011883 // Rnf13 // ring finger protein 13 // 3 D // 24017 /// AF037206 // Rnf	5.51E-05
XM_984379 // Prpf19 // PRP19/PSO4 pre-mRNA processing factor 19 homolog (S. cere	6.50E-05
NM_001001144 // Scap // SREBP cleavage activating protein // 9 F2 // 235623 ///	6.68E-05
NM_026476 // 2610101N10Rik // RIKEN cDNA 2610101N10 gene // 9 E4 // 67958 /// AK	6.76E-05
NM_145959 // D15Ertd621e // DNA segment, Chr 15, ERATO Doi 621, expressed // 15	6.79E-05
NM_013746 // Plekhb1 // pleckstrin homology domain containing, family B (evectin	7.23E-05
XM_986293 // Rhox12 // reproductive homeobox 12 // X A3.3 // 382282	7.68E-05
XM_992477 // Ak3l1 // adenylate kinase 3 alpha-like 1 // 4 C6|4 47.6 cM // 11639	8.22E-05
NM_028411 // Tmem138 // transmembrane protein 138 // 19 B // 72982 /// AK007197	8.40E-05
NM_023671 // Clns1a // chloride channel, nucleotide-sensitive, 1A // 7 E3|7 50.0	8.68E-05
XM_923609 // Ptpn23 // protein tyrosine phosphatase, non-receptor type 23 // 9 F	8.94E-05
NM_133781 // Cab39 // calcium binding protein 39 // 1 C5 // 12283 /// XM_976988	9.03E-05
NM_175004 // Ptrh2 // peptidyl-tRNA hydrolase 2 // 11 C // 217057 /// AK031620 /	9.27E-05
NM_001013026 // Ttf2 // transcription termination factor, RNA polymerase II // 3	9.80E-05
NM_028247 // Slc16a10 // solute carrier family 16 (monocarboxylic acid transport	1.25E-05
NM_054048 // Rcor2 // REST corepressor 2 // 19 A // 104383 /// AK205817 // Rcor2	1.25E-05
---	2.79E-05
NM_028066 // F11 // coagulation factor XI // 8 B1.1|8 25.0 cM // 109821	3.45E-05
NM_001038663 // Mapk1 // mitogen activated protein kinase 1 // 16 A3|16 9.82 cM	4.47E-05
NM_026127 // 4833420G17Rik // RIKEN cDNA 4833420G17 gene // 13 D2.3 // 67392 ///	7.35E-05
NM_211355 // 1110034C04Rik // RIKEN cDNA 1110034C04 gene // 12 E // 68734 /// AK	9.97E-05
XR_003366 // LOC384857 // similar to Keratin, type II cytoskeletal 8 (Cytokerati	1.26E-05
AK029589 // 4930405O22Rik // RIKEN cDNA 4930405O22 gene // // 414079 /// AK1604	1.29E-05
NM_019468 // G6pd2 // glucose-6-phosphate dehydrogenase 2 // 5 C3.1|5 39.0 cM //	2.87E-05

Genes selected for further bioinformatics analysis are highlighted in red

Table 20 Genes with differential exon usage using the full metaprobeset

Gene_assignment	*p*-value Alternative splicing
NM_177615 // C78409 // expressed sequence C78409 // 10 D3 // 216441	7.95E-07
NM_133215 // Mtmr4 // myotubularin related protein 4 // 11 C // 170749	6.48E-08
NM_198017 // C430003P19Rik // RIKEN cDNA C430003P19 gene // 7 F3 // 109359 /// A	8.96E-08
NM_133916 // Eif3s9 // eukaryotic translation initiation factor 3, subunit 9 (et	1.15E-07
NM_008008 // Fgf7 // fibroblast growth factor 7 // 2 F-G // 14178 /// AK037172 /	4.59E-07
---	1.30E-06
NM_008149 // Gpam // glycerol-3-phosphate acyltransferase, mitochondrial // 19 D	1.86E-06
---	2.88E-06
NM_013854 // Abcf1 // ATP-binding cassette, sub-family F (GCN20), member 1 // 17	3.84E-06
NM_173182 // Fndc3b // fibronectin type III domain containing 3B // 3 A3 // 7200	5.08E-06
NM_001039080 // Rbms2 // RNA binding motif, single stranded interacting protein	5.42E-06
NM_019426 // Atf7ip // activating transcription factor 7 interacting protein //	5.87E-06
NM_008431 // Kcnk4 // potassium channel, subfamily K, member 4 // 19 A|19 4.5 cM	6.38E-06
XM_994149 // Gcn1l1 // GCN1 general control of amino-acid synthesis 1-like 1 (ye	6.56E-06
NM_053089 // Narg1 // NMDA receptor-regulated gene 1 // 3 D // 74838 /// AK07804	9.64E-06
---	9.97E-06
---	1.06E-05
XM_983388 // Ppp2r5a // protein phosphatase 2, regulatory subunit B (B56), alpha	1.23E-05
AK029589 // 4930405O22Rik // RIKEN cDNA 4930405O22 gene // // 414079 /// AK1604	1.30E-05
NM_009781 // Cacna1c // calcium channel, voltage-dependent, L type, alpha 1C sub	1.48E-05
NM_007801 // Ctsh // cathepsin H // 9 E3.1|9 50.0 cM // 13036 /// BC080767 // Ct	1.63E-05
---	1.78E-05
AK020953 // B230104C08Rik // RIKEN cDNA B230104C08 gene // 6 G3 // 77841	2.13E-05
XM_001000465 // 9930012K11Rik // RIKEN cDNA 9930012K11 gene // 14 D2 // 268759 /	3.20E-05
NM_008591 // Met // met proto-oncogene // 6 4.0 cM // 17295 /// AK021346 // D730	3.39E-05
NM_001001144 // Scap // SREBP cleavage activating protein // 9 F2 // 235623 ///	3.46E-05
---	3.53E-05
NM_026127 // 4833420G17Rik // RIKEN cDNA 4833420G17 gene // 13 D2.3 // 67392 ///	3.62E-05
NM_178654 // Pkn2 // protein kinase N2 // 3 H1 // 109333 /// AK083670 // D030063	3.70E-05
NM_009392 // Tlx2 // T-cell leukemia, homeobox 2 // 6 C3-D1|6 35.5 cM // 21909	3.79E-05
XM_001052047 // Lhx3 // LIM homeobox protein 3 // 2 A2-C1|2 16.0 cM // 16871 ///	4.01E-05
NM_011883 // Rnf13 // ring finger protein 13 // 3 D // 24017 /// AF037206 // Rnf	4.02E-05
NM_007458 // Ap2a1 // adaptor protein complex AP-2, alpha 1 subunit // 7 B2 // 1	4.20E-05
XM_001005692 // LOC673486 // similar to eukaryotic translation initiation factor	4.38E-05
XM_994018 // 2600009E05Rik // RIKEN cDNA 2600009E05 gene // 2 F3 // 77006 /// XM	4.38E-05
NM_001033259 // D130073L02Rik // RIKEN cDNA D130073L02 gene // 10 B4 // 215999 /	4.82E-05
BC029239 // Timm8b // translocase of inner mitochondrial membrane 8 homolog b (y	4.92E-05
NM_011161 // Mapk11 // mitogen-activated protein kinase 11 // // 19094 /// NM_0	5.09E-05
NM_054048 // Rcor2 // REST corepressor 2 // 19 A // 104383 /// AK205817 // Rcor2	5.23E-05
XM_001001499 // Akap11 // A kinase (PRKA) anchor protein 11 // 14 D3 // 219181 /	5.65E-05
NM_026476 // 2610101N10Rik // RIKEN cDNA 2610101N10 gene // 9 E4 // 67958 /// AK	5.84E-05
XM_979043 // Gm1027 // gene model 1027, (NCBI) // 4 C7 // 381538 /// NM_028209 /	5.85E-05
NM_001005331 // Eif4g1 // eukaryotic translation initiation factor 4, gamma 1 //	5.85E-05
XM_984379 // Prpf19 // PRP19/PSO4 pre-mRNA processing factor 19 homolog (S. cere	6.23E-05
NM_030109 // Sf3b2 // splicing factor 3b, subunit 2 // 19 A // 319322 /// AK1890	6.36E-05
NM_021605 // Nek7 // NIMA (never in mitosis gene a)-related expressed kinase 7 /	6.46E-05
---	6.62E-05
---	6.64E-05
NM_028411 // Tmem138 // transmembrane protein 138 // 19 B // 72982 /// AK007197	6.67E-05
---	6.70E-05
NM_033075 // D17H6S56E-5 // DNA segment, Chr 17, human D6S56E 5 // 17 B1|17 19.0	6.75E-05
NM_019468 // G6pd2 // glucose-6-phosphate dehydrogenase 2 // 5 C3.1|5 39.0 cM //	7.60E-05
NM_175093 // Trib3 // tribbles homolog 3 (Drosophila) // 2 G3 // 228775 /// NM_1	7.90E-05
XM_986293 // Rhox12 // reproductive homeobox 12 // X A3.3 // 382282	8.27E-05
---	8.45E-05
NM_019710 // Smc1l1 // SMC (structural maintenance of chromosomes 1)-like 1 (S.	9.51E-05
NR_002854 // Dlx1as // distal-less homeobox 1, antisense // 2 C2|2 44.0 cM // 11	9.88E-05
XM_975459 // Trim55 // tripartite motif-containing 55 // 3 A2 // 381485/// XM9	9.93E-05

Genes selected for further bioinformatics analysis are highlighted in red

Genes were selected for further analysis by using web-based bioinformatics such as USCS (http://genome.ucsc.edu/cgi-bin/hgGateway) and MGI (http://gbrowse.informatics.jax.org/cgi-bin/gbrowse/mouse_current/) genome browser gateways. The browsers show alignments of a comprehensive set of sequences from public domain databases that include known and predicted DNA and mRNA sequences, both full length and expressed sequence tags. Often exon inclusion or exclusion was apparent within the alignments, and published alternative splice isoforms could be identified. Genes with known alternatively spliced exons were considered good candidates. In addition, NCBI OMIM database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM) and MGI (http://www.informatics.jax.org/) gene detail searches were used to identify candidate genes whose disruption resulted in symptoms related to those observed in DM. Six genes were selected (Table 21) and the function, known alternative splicing and possible relations to DM1 symptoms were summarised:

Table 21 Alternative splice ANOVA *p-values* using core, extended and full probeset data.

	Core	Extended	Full
MtmR4	4.99E-01	2.90E-08	6.48E-08
Kcnk4	2.38E-03	5.35E-06	6.38E-06
Narg1	9.19E01	1.63E-05	9.64E-06
Ttyh1	4.57E-05	6.96E-01	1.54-04
Bptf	6.02E-05	1.41E-01	2.32-01
Cacna1c	2.87E-02	1.54E-02	1.48E-05

Note–Significant values are shown in red

5.3.3.2.1 MtmR4

The myotubularin gene *MTM1*, is the archetypal member of a family of highly conserved protein-tyrosine phosphatase-like enzymes, of which there are at least ten members (*MTM1* and *MTMR1-9*). *MTM1* is mutated in X-linked myotubular myopathy, causing a severe congenital myopathy characterised by the presence of disorganized skeletal muscle with fibres that contain centrally located nuclei (Laporte *et al.*, 1996; Buj-Bello *et al.*, 1999).

Mutations in *MTMR2* result in the neurodegenerative disorder type 4B Charcot-Marie-Tooth disease (CMT4B), a demyelinating motor and sensory neuropathy that starts in infancy, characterized by the presence of foci of abnormally folded myelin sheaths and Schwann cell proliferation in peripheral nerves (Berger *et al.*, 2002). CMT4B causes progressive symmetric distal and proximal weakness starting in the lower extremities. Also, cranial-nerve deficits are observed in most patients. Recently, *MTMR1* was shown to be mis-spliced in DM. Researchers identified 6 alternatively spliced mRNA isoforms of *MTMR1*, one which is muscle specific, induced during myogenesis, and represents the major isoform in adult skeletal muscle. The authors found reduced levels of the muscle-specific isoform and the appearance of an abnormal *MTMR1* transcript in cultured differentiated muscle cells and in skeletal muscle from congenital myotonic dystrophy patients, leading them to postulate that MTMR1 may play a role in muscle formation (Buj-Bello *et al.*, 2002). They also showed that mice deficient for the gene develop a generalized and progressive myopathy (Buj-Bello *et al.*, 2002). Little is known about the function of the remaining myotubularin family members including *MtmR4,* identified in this microarray analysis, but given the effects of *MTMR1* and *MTMR2* mutations*, MTMR4* warrants further investigation. Zhao et al. found expression of *MTMR4* in all human tissues examined except lung, small intestine, stomach, salivary gland, adrenal gland, and uterus (Zhao *et al.*, 2001).

MtmR4 was found to have three significant exon differences. Interestingly expression of either 5 or 250 CUG repeats lead to alternative exon splicing in different exons (Figure 56). It is impossible to compare alternative splicing at the core probeset level with the DM missplicing of *MTMR1* since alternative exons 2.1; 2.2 and 2.3 are not represented, this would need to be investigated molecularly. Buj-Bello *et al.* (2002) also reports, but does not specify, newly identified mis-spliced exons (Buj-Bello *et al.*, 2002), which could relate to the differences documented here.

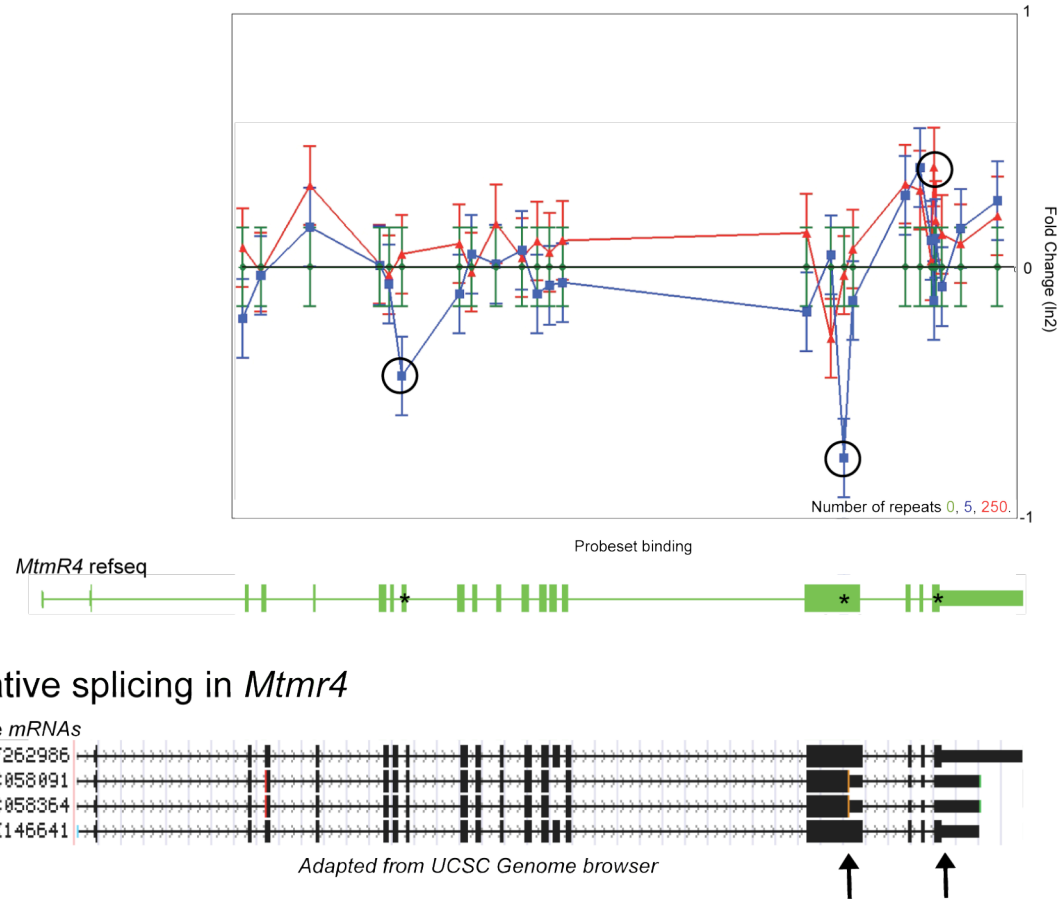


Figure 56 Alternative exon usage in *MtmR4.* Graph plotting probeset value. Data for 5 and 250 repeats were subtracted from 0 repeats (pLoxEGFP). Each exon is represented by one or more probesets that have been spaced proportionately to the genomic sequence for clarity. Probesets corresponding to probable alternative exon usage are circled. Refseq sequences are aligned to show the position of the exon within the gene. Refseq: NM133215 length 24,080 bp . Arrows (and red lines) indicate single nucleotide differences, which terminate the open reading frame. This may affect probe hybridisation at this position, but does not account for the differences between the samples, and therefore probably has no relevance.

5.3.3.2.2 *Kcnk4*

A little mentioned indication in myotonic dystrophy, but problematic when performing surgical procedures, is that of anaesthetic sensitivity. In DM1, patients have a higher sensitivity to sedatives, anaesthetic, and neuromuscular blocking agents which can result in cardiovascular and respiratory complications, thought to originate from muscular fatigue and myotonia (Aldridge, 1985; Klompe *et al.*, 2007). *KCNK4* is a stretch and polyunsaturated fatty acids activated tandem pore-domain potassium (K_{2P}) channel. A voltage insensitive background potassium channel, which restores the membrane resting potential close to equilibrium (Lesage *et al.*, 2000). The channel is highly expressed in brain (Fink *et al.*, 1998; Lesage *et al.*, 2000) and the role of K_{2P} channels in general anesthesia and neuroprotection have been proposed recently. Harinath *et al.* 2004, identified TREK-1 and TRAAK channels (also known as Kcnk2 and Kcnk4) as molecular

targets for trichloroethanol (an active metabolite of the general anaesthetic chloral hydrate) and suggested that activation of these channels might contribute to the CNS depressant effects of chloral hydrate (Harinath *et al.*, 2004). The channel is also activated by the glutamate release inhibitor Riluzole, a potent neuroprotective agent with anticonvulsant, anaesthetic (Mantz *et al.*, 1992) and anti-ischaemic properties in brain (Heurteaux *et al.*, 2006), retina (Ettaiche *et al.*, 1999; Izumi *et al.*, 2003) and spinal cord (Lang-Lazdunski *et al.*, 1999). This drug also has shown protective effects in animal models of Parkinson's disease (Boireau *et al.*, 1994; Benazzouz *et al.*, 1995); in other models of acute neurodegenerative diseases (Kennel *et al.*, 2000; Fumagalli *et al.*, 2006) and in human neurodegenerative diseases (Howard *et al.*, 2002; Jankovic *et al.*, 2002; Killestein *et al.*, 2005; Wu *et al.*, 2006; Mitsumoto *et al.*, 2007). If activation of this channel by Riluzole confers such beneficial neuroprotective effects, it is feasible then that dysfunction of KCNK4 –possibly brought about by alternative splicing– could have the opposite degenerative effects, such as retinal degeneration and cortical atrophy.

Two *KCNK4* alternative splice variants have been characterised in human with a third detected on northern analysis. Gene expression in mouse is limited to the neuronal system, but in human is expressed more widely, mainly in the heart and brain but also in the liver, skeletal muscle, kidney and pancreas (Fink *et al.*, 1998; Ozaita *et al.*, 2002). Any differences between their function, or developmental expression of the alternative transcripts, has not yet been established.

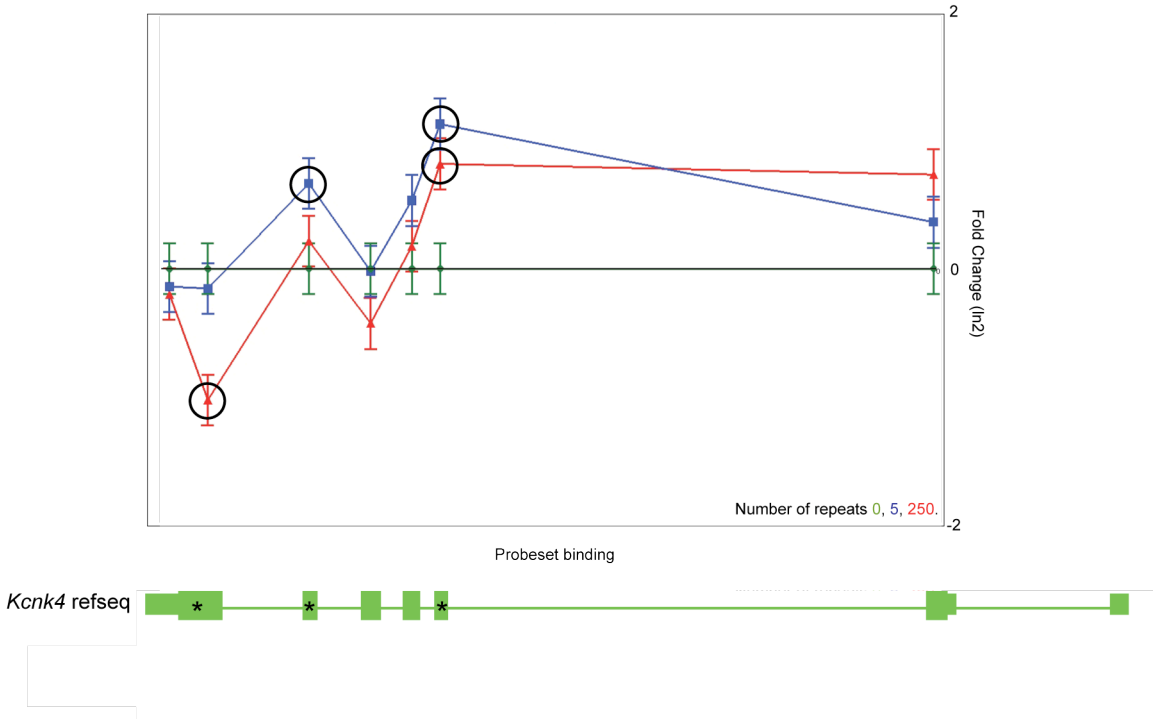


Figure 57 Alternative exon usage in *Kcnk4.* Graph plotting probeset value. Data for 5 and 250 repeats were subtracted from 0 repeats (pLoxEGFP). Each exon is represented by one or more probesets which are spaced proportionately to the genomic sequence. Probesets corresponding to probable alternative exon usage are circled. Refseq sequences are aligned to show the position of the exon within the gene. Refseq: NM008431 length 8,826 bp.

Kcnk4 was found to have three possible exon differences, which differ from those of the human orthologue published by Ozaita *et al.* No information was found in the literature concerning embryonic transcript expression. Again, expression of either 5 or 250 CUG repeats lead to alternative exon splicing (Figure 57).

5.3.3.2.3 Narg1

Localized at excitatory synapses, NMDA receptors are one of the major classes of ionotropic neurotransmitter receptors of mammalian brain (Stephenson, 2006). Sugiura *et al.* showed that highly specific NMDA receptor-dependent regulation of *NARG1, NARG2* and *NARG3* gene expression plays an important role in the transition from proliferation of neuronal precursors to differentiation of neurons. *Narg1* (also known as *murine N-terminal acetyltransferase 1*, *mNAT1*) is an embryonic gene highly expressed in the developing brain in regions of neuronal proliferation and migration and subsequently down-regulated during early postnatal development, in part, by the onset of N-methyl-d-aspartate (NMDA) receptor function (Sugiura *et al.*, 2001). Narg1 and its co-subunit, Ard1, assemble to form a functional acetyltransferase. Throughout brain development, *Narg1* and *Ard1* are highly expressed in areas of cell division and migration, and are down-regulated as neurons differentiate (Sugiura *et al.*, 2003). Suguira *et al.*, also found that *Narg1* and *Ard1* are expressed in proliferating mouse P19 embryonic carcinoma cells; treatment of these cells with retinoic acid initiated exit from the cell cycle, neuronal differentiation, and down-regulation of *Narg1* and *Ard1* as the *NMDA receptor 1* gene was induced. The results provided direct evidence that Narg1 and Ard1 form an N-terminal acetyltransferase in mice and suggested that expression and down-regulation of this enzyme complex played an important role in the generation and differentiation of neurons. (Sugiura *et al.*, 2003), so it follows that inappropriate regulation could affect neuronal generation in the newborn.

NARG1 could have a role in other symptoms of DM, such as retinal degeneration, cataracts and testicular atrophy: In the eye, NARG1 is also required to control retinal neovascularisation in adult ocular endothelial cells (Paradis *et al.*, 2002), and involved in

corneal morphogenesis. *Narg1* (also known as *Tubedown1*) is highly expressed in mouse embryonic cornea (E18.5) and also in adult corneal endothelium (Saika *et al.*, 2001).

There are two known isoforms of the protein produced by alternative splicing in the mouse (swissprot Q9BXJ9-1 and Q9BXJ9-4). Interestingly, NARG1 may also be important in spermatogenesis since expression levels varied throughout spermatogenesis in the rat (He *et al.*, 2002).

In our analysis, *Narg1* was found to have three possible exon differences. Again, expression of either 5 or 250 CUG repeats lead to alternative exon splicing (Figure 58).

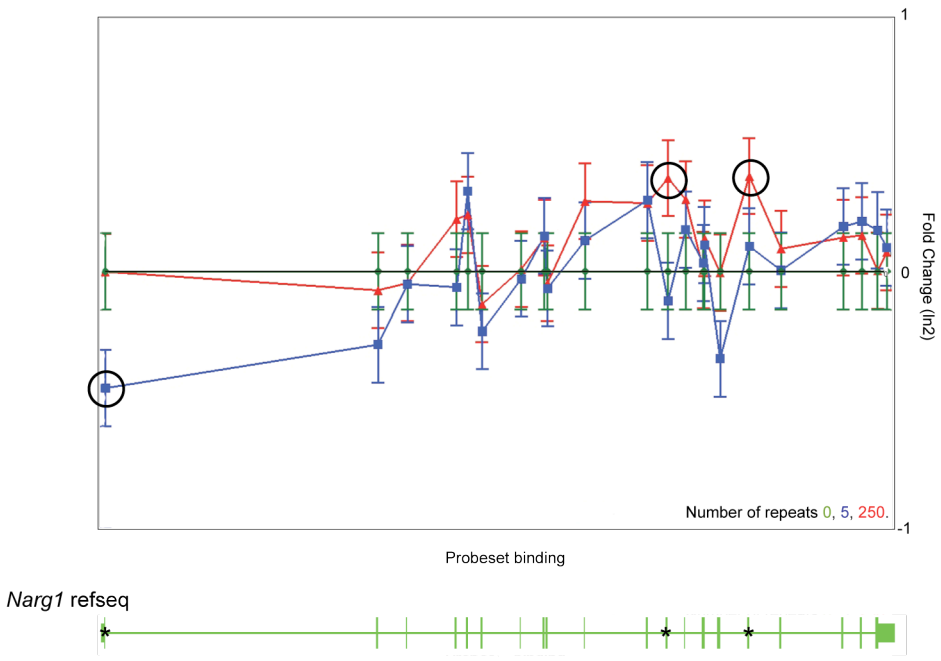


Figure 58 Alternative exon usage in *Narg1*. Graph plotting probeset value. Data for 5 and 250 repeats were subtracted from 0 repeats (pLoxEGFP). Each exon is represented by one or more probesets which are spaced proportionately to the genomic sequence. Probesets corresponding to probable alternative exon usage are circled. Refseq sequences are aligned to show the position of the exon within the gene. Refseq: NM053089 length 57,883 bp.

5.3.3.2.4 Ttyh1

Myotonic dystrophy is a multisystemic disease, which affects the brain to varying degrees. Cerebral atrophy and white matter lesions are often observed in DM1 (Damian *et al.*, 1993; Antonini *et al.*, 2004). Neurofibriliary tangles (NFTs) –which are present in Alzheimer's and most other dementias (Robert *et al.*, 2007)– are also seen in DM1 brains (Kiuchi *et al.*, 1991). These are formed from insoluble aggregates of abnormally phosphorylated neuronal

microtubule associated protein tau. It would seem that the tau mis-splicing of exon 2 is responsible for neuronal loss since changes in tau isoform ratios leads to NFT formation, but it is not known if tangles are pathogenic; a benign marker of pathogenesis, or protective. There is a link between Tau mutations and learning disabilities, but the association is not absolutely clear since in mouse models, the formation of NFTs can be disassociated from tau induced neuronal loss (Andorfer *et al.*, 2005; Santacruz *et al.*, 2005). Alzheimer's disease is classed as a tauopathy, but lacks any genetic tau mutation yet NFT are formed, and there is neuronal loss, so the mechanism of pathogenesis is in part unclear. The brain lesions observed in DM1 can be explained by tau induced neuronal loss, nevertheless this need not exclude involvement of other genes.

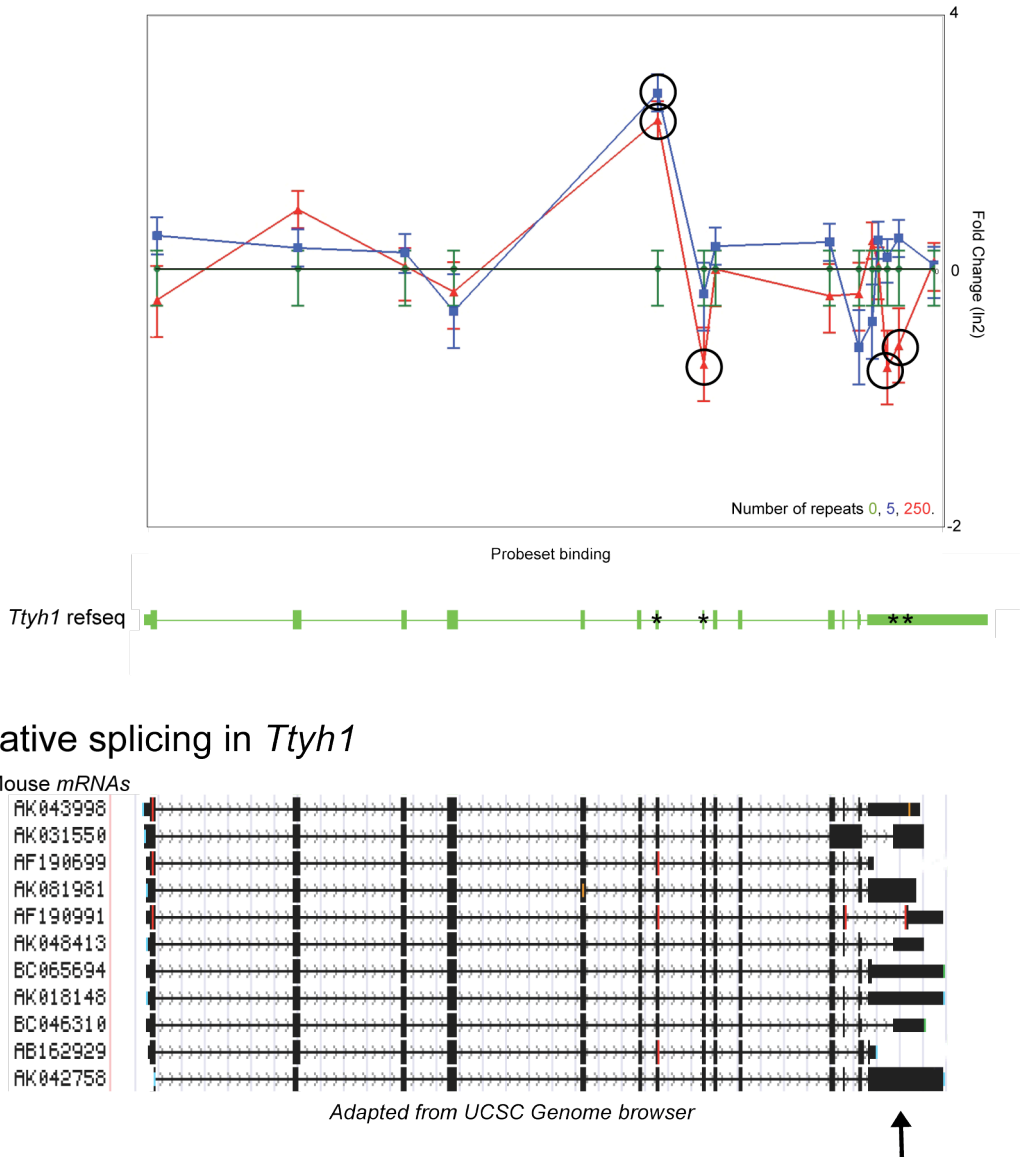


Figure 59 Alternative exon usage in *Ttyh1*. Graph plotting probeset values. Data for 5 and 250 repeats were subtracted from 0 repeats (pLoxEGFP). Each exon is represented by one or more probesets which are spaced proportionately to the genomic sequence. Probesets corresponding to probable alternative exon usage are circled. Genbank mRNA sequences are aligned. Red markings indicate base differences. The arrow marks known alternative exon usage. Refseq

sequences are aligned to show the position of the exon within the gene. Refseq: NM021324 length 16,709 bp.

Human *TTYH1* is related to the drosophila gene family *tweety* (flightless behavioural, Campbell *et al.*, 2000), expression is restricted to neuronal tissue and testis (Matthews *et al.*, 2007). There are two known splice variants, the least common one of which is a swell-activated maxi chloride channel, probably controlling cell volume (Suzuki, 2006). Mathews *et al.*, expressed mouse *Ttyh1* in non-neuronal human kidney (HEK) cells which caused axon-like protrusions to occur. They observed the highest concentration of TTYH1 at the tips of the protrusions, where adhesion and guidance molecules are also expressed. The protein was also found to be associated with integrin alpha 5. Integrins play a key role in cell migration relative to extracellular matrix, in cell adhesion to basement membranes and resulting cell polarization. Their data suggested a role for TTYH1 in process formation, cell adhesion and possibly as a transmembrane receptor (Matthews *et al.*, 2007), functions important during brain development.

Ttyh1 was found to have four possible exon differences, and yet again, expression of either 5 or 250 CUG repeats lead to alternative exon splicing. The latter two differences were positioned within the 3' UTR, where alternative splice variants have been sequenced previously and deposited in genbank (Figure 59).

5.3.3.2.5 Bptf

BPTF is involved in brain development and chromatin remodelling. Also Known as *Fac1, Nurf301* and *falz*, *Bptf* was identified by the reactivity of its encoded protein to a monoclonal antibody prepared against brain homogenates from patients with Alzheimer's disease. The gene's discoverers determined that it is developmentally regulated, since they found abundant expression in fetal brain, where it is present throughout the gray and white matter of the developing spinal cord at 18-22 gestational weeks. In the adult, expression was found to be at low levels in brain and spinal cord, except in neurodegenerative diseases where the protein is up-regulated (Bowser *et al.*, 1995). Functionally, the subunit composition suggests that it represents the human ortholog of the *nurf301* component of the *Drosophila* nucleosome remodelling factor (NURF) complex. Wysoka *et al.*, discovered that the gene product preferentially bound to trimethylated histone H3 lysine 4 tails (H3K4me3), which mark the transcription start sites of virtually all transcriptionally active genes (Yan *et al.*, 2006), allowing activation by chromatin remodelling by the

NURF complex (Wysocka *et al.*, 2006). Human NURF is enriched in brain, Barak *et al.* demonstrated that it regulates human ENGRAILED, a homeodomain protein that regulates neuronal development in the mid-hindbrain. They also showed that hNURF potentiates neurite outgrowth in cell culture, and suggest a role for the transcription factor complex in neuronal growth (Barak *et al.*, 2003).

Human *BPTF* is alternatively spliced, two different isoforms have been described completely (Entrez nucleotide sequence database ID: NM182641 -isoform 1 and NM004459 -isoform 2) and four named isoforms formed by alternative splicing are listed in the Swissprot database (Swissprot ID Q12830-1, Q12830-2, Q12830-3, Q12830-4). From the analysis, expression of either 5 or 250 CUG repeats lead to six possible exon differences between the two samples compared to EGFP. The exon difference noted in pLoxEGFP5 corresponds to a known alternatively spliced exon in the human gene (Figure 60), which is excluded in isoform 2 (NM004459) and included in isoform 1 (NM182641). As mentioned earlier, *BPTF* is thought to be developmentally regulated, but it is not clear whether this is done by isoform switching, or if the isoforms co-exist.

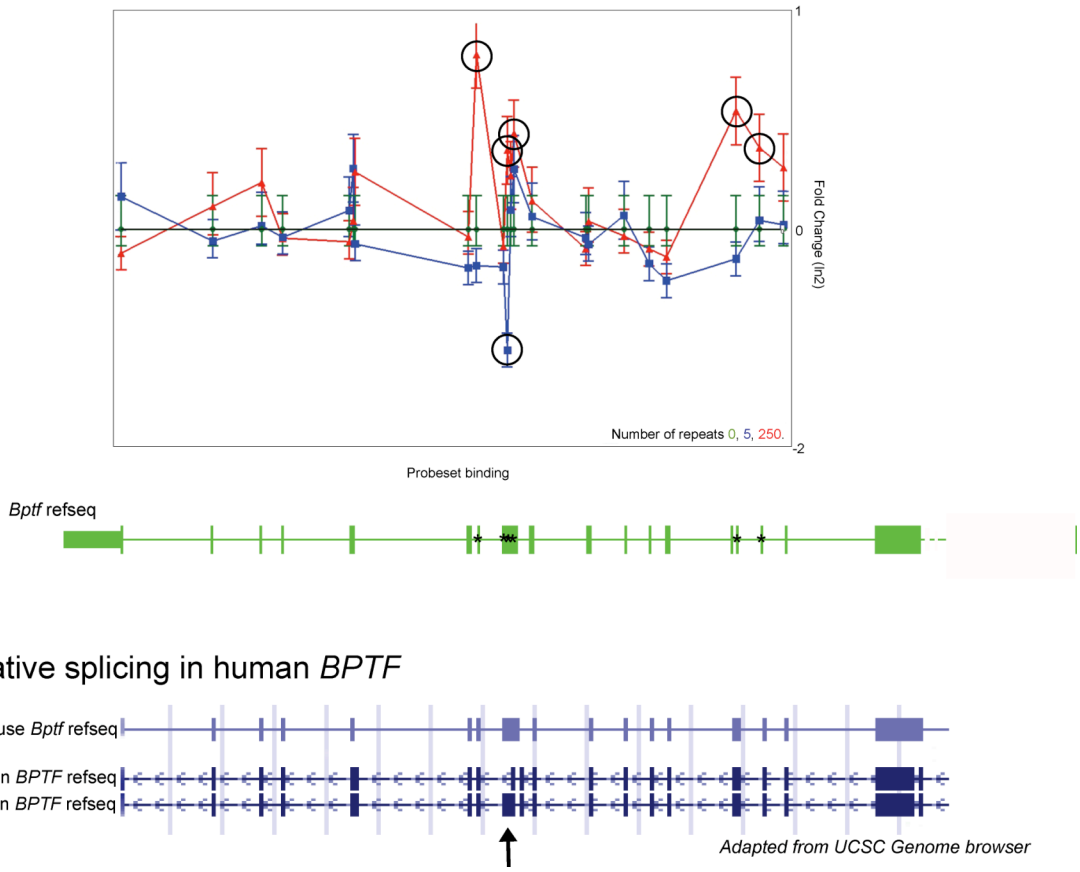


Figure 60 Alternative exon usage in *Bptf.* Graph plotting probeset values. Data for 5 and 250 repeats were subtracted from 0 repeats (pLoxEGFP). Each exon is represented by one or more probesets which are spaced proportionately to the genomic sequence. Probesets corresponding to probable alternative exon usage are circled. Refseq sequences are aligned to show the position of

the exon within the gene. The arrow identifies the exon excluded after expression of the pLoxEGFP5 transgene also excluded in human isoform 2. Refseq: mouse, NM176850 length 98,842 bp; human, NM004459.6 and NM182641.3.

5.3.3.2.6 Cacna1c

CACNA1C is the pore-forming subunit of the L-type, voltage dependent calcium channel responsible for excitation-contraction coupling in the heart (see Striessnig, 1999 for review); heart development (Rottbauer *et al.*, 2001), and is expressed in developing skeletal muscle (Chaudhari *et al.*, 1993). Cardiac conduction defects have been well documented in myotonic dystrophy, both in patients and mouse models (Berul *et al.*, 2000; Harper, 2001; Pall *et al.*, 2003; Mahadevan *et al.*, 2006).

Excitation-contraction coupling in the heart is a carefully ordered process. In order for blood to be pumped directionally through the heart, atrial blood must first be squeezed into the ventricles and then out into the aorta. The atrial contraction triggered by the sino-atrial node, and the delay whilst atrial action potentials are restored and the ventricles fill with blood, is referred to as the PR interval. The QRS complex (the peak on an ECG, Figure 61) relates to the spread of electrical activity triggered by the atrio-ventricular node. This spreads down the centre of the heart and upwards through the ventricles resulting in ventricular contraction, which after returning to resting potential is termed the QT interval.

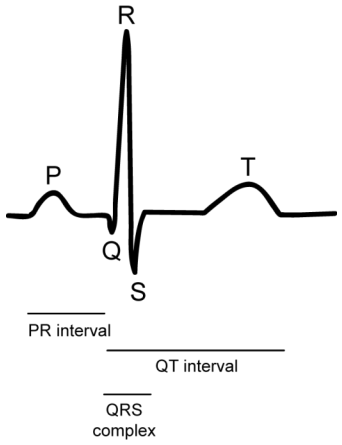


Figure 61 The electrocardiogram (ECG). Schematic diagram depicting the nomenculature of electrical peaks and troughs during a heartbeat. P, P wave; PR, PR interval; QRS, QRS complex; QT, QT interval; T, T wave. Adapted from Merck Manuals online medical library (http://www.merck.com/mmpe/index.html).

Molecularly excitation-contraction coupling is controlled by ion transport. Myocardiocytes possess a negative membrane potential when at rest. Electrical signals cause voltage gated calcium channels to depolarise, releasing calcium ions into the cell. The influx of calcium results in further release of free calcium into the cell from the sarcoplasmic reticulum,

causing the cell to contract. After a delay potassium channels open, allowing potassium to flow out of the cell for repolarisation. $Ca_v1.2$ is the L-type or long lasting, voltage dependent calcium channel responsible for excitation-contraction coupling in the heart. It mediates calcium influx in response to membrane depolarisation, and controls sustained contraction of the myocardium via calcium-induced Ca^{2+} release from the sarcoplasmic reticulum. The *Cacna1c* gene product is the chief pore-forming subunit of $Ca_v1.2$, responsible for the calcium current, and is associated with auxiliary subunits $\alpha^2\delta$, β and γ, which modulate membrane expression and current properties (Tuluc *et al.*, 2007). Although the $Ca_v1.2$ controls the main calcium current in the heart, it also has other cellular functions; excitation-coupling in smooth muscle, hormone release (in mice, selective ablation in beta cells resulted in impaired insulin secretion and systemic glucose intolerance (Schulla *et al.*, 2003)), regulation of transcription, and synaptic integration (Catterall *et al.*, 2005 for review).

Mutations in *CACNA1C* lead to conductance defects in the heart –a major complication in myotonic dystrophy. Particular amino acid substitutions in the cardiac $Ca_v1.2$ channel have been shown to cause elevated ST levels and short QT intervals (Figure 61) on electrocardiograms and result in sudden cardiac death (Antzelevitch *et al.*, 2007). In DM1 however ECGs generally reveal different variations –a long PR interval and wide QRS complex (Figure 61) (Bu'Lock *et al.*, 1999), which could result from different voltage sensitivities in the same gene caused by aberrant splicing (Soldatov *et al.*, 1995; Tang *et al.*, 2004). Indeed, mutations in exon8/8a of this gene lead to Timothy syndrome, an autosomal dominant disorder characterised by long QT intervals (Figure 61) and ventricular cardiac arrhythmias (Splawski *et al.*, 2004; Splawski *et al.*, 2005; Faber *et al.*, 2007).

Aberrant splicing of $Ca_v1.2$ could also play a role in anaesthetic sensitivity. Volatile anaesthetics depress cardiac contractility by inhibition of cardiac L-type calcium channels (Gingrich *et al.*, 2005), so further depression of an already sub-functional channel could worsen complications and result in heart and respiratory failure. It seems likely then that dysregulation of *CACNA1C* alternative splicing could play a major role in the defective heart conductance observed in DM, and also contribute to anaesthetic sensitivity.

CACNA1C is subject to extensive alternative splicing with 19 of 55 exons alternatively spliced (Tang *et al.*, 2004; Murakami *et al.*, 2006). The N-terminus is associated with trafficking and the C-terminus Ca2+-dependent and slow voltage-dependent inactivation

(Kobrinsky *et al.*, 2005). Both embryonic and adult specific isoforms are produced from combinatorial splicing of exons 31-33 resulting in shifts in the voltage dependence of activation (Diebold *et al.*, 1992; Tang *et al.*, 2004).

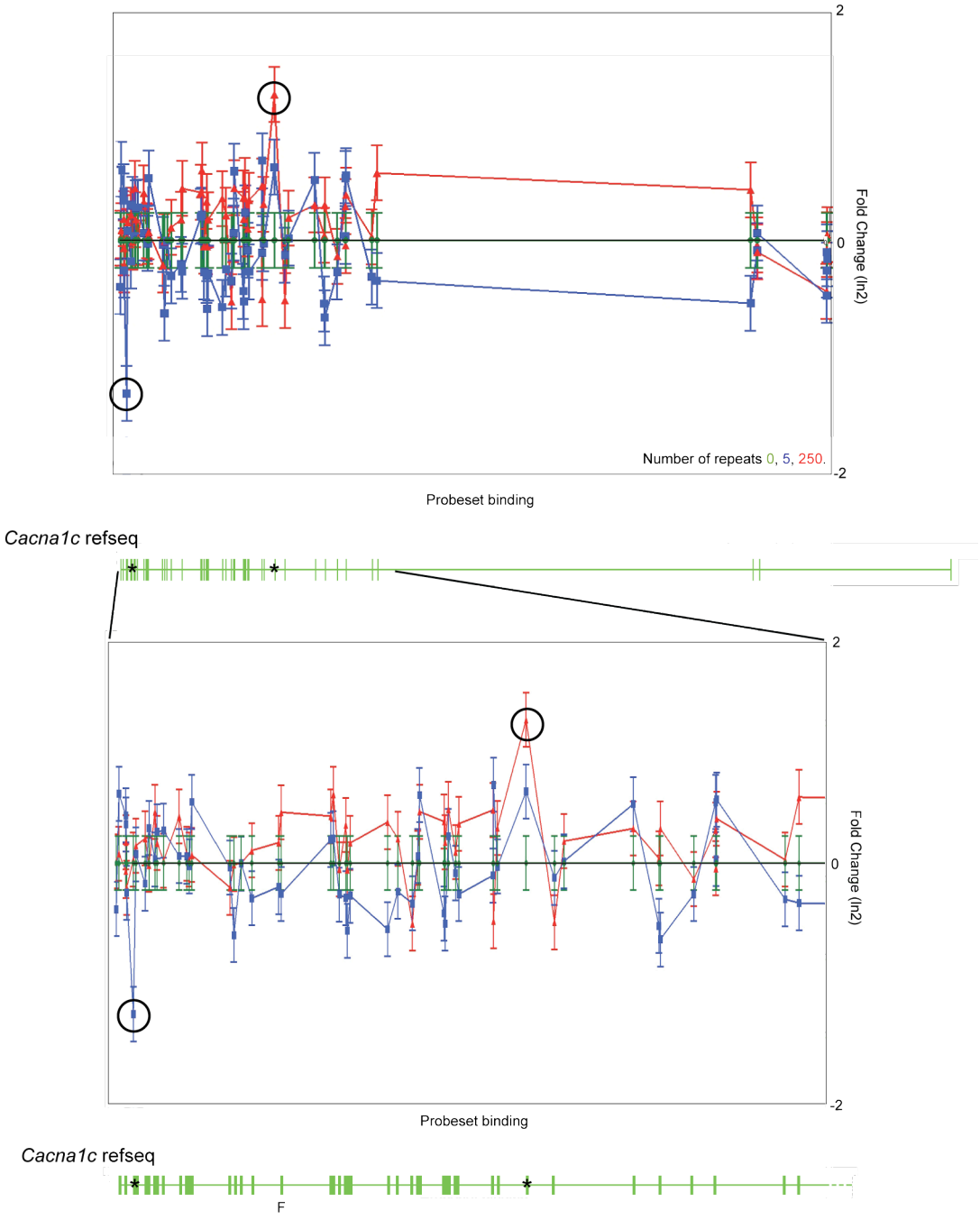


Figure 62 Alternative exon usage in *Cacna1c*. Graph plotting probeset values. Data for 5 and 250 repeats were subtracted from 0 repeats (pLoxEGFP). Each exon is represented by one or more probesets which have been spaced equally rather than proportionately to the genomic sequence for clarity. Probesets corresponding to probable alternative exon usage are circled. Refseq sequences are aligned. Refseq: NM009781 length 602,340 bp. F denotes area of embryonic alternative splicing.

Cacna1c was found to have at least two possible exon differences. Expression of either 5 or 250 CUG repeats lead to alternative exon splicing (Figure 62).

Preliminary analysis of probeset data at the full metaprobeset level (data not shown), which has a much lower *p*-value (1.48E-05 *vs.* 2.87E-02 at the core level) indicates differential exon usage around the foetal region. This is consistent with the inappropriate foetal isoform switching seen in DM patients and warrants further investigation.

5.4 Expression of known genes

In anticipation of the microarray results we wanted to look at the expression of genes already known to be mis-spliced in myotonic dystrophy. We chose the *chloride channel 1* (*Clcn1*) and *cardiac troponin T* (*Tnnt2*) genes for study since the splicing patterns had already been well characterised in mouse models. For *Tnnt2*, exon 5 inclusion is promoted by the expression of expanded CUG repeats and in *Clcn1*, exon 7a inclusion which leads to the formation of a premature termination codon and probably nonsense mediated decay (Mankodi *et al.*, 2002; Kanadia *et al.*, 2003; Kanadia *et al.*, 2006). Mouse 3T3 cells were transfected with reagent alone; pLoxEGFP; pLoxEGFP5 or pLoxEGFP250 and harvested after 24 hours incubation. Total RNA was extracted, and used as a template for first strand cDNA synthesis primed with random hexamers. Primers were designed as described in Kanadia *et al.* 2003 to amplify *Tnnt2* cDNA between exons 2 and 6 and *Clcn1* between exons 5 and 8. The PCR annealing temperature and magnesium chloride concentrations were optimised in order to detect reaction products using ethidium bromide staining and agarose gel electrophoresis (Figure 63).

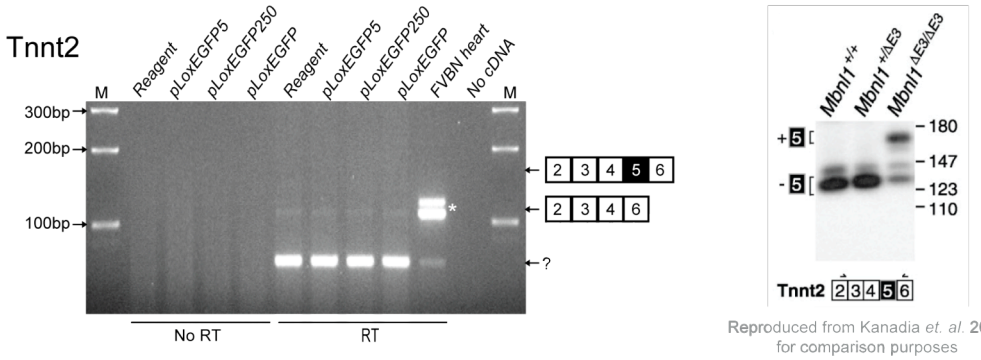


Figure 63 Detection of Tnnt2 transcripts in 3T3 cells. Agarose gel separated RT-PCR. 3T3 cells were transfected with constitutively expressing pLoxEGFP constructs containing 0; 5 or 250 CUG

repeats and analysed by RT-PCR . Primers were positioned between exons 2 and 6. Isoform sizes were estimated from Kanadia *et al.* 2003, the positions and exon content indicated on the RHS. The expression of expanded CUG repeats has been reported to increase the inclusion of exon 5, but in this experiment levels were too low to detect. No difference in the expression pattern was seen between treatments. *See Figure 64 legend.

The *Tnnt2* banding pattern obtained for the adult FVBN heart cDNA corresponded in size to that previously published (Figure 63) (Kanadia *et al.*, 2003) where exon 5 exclusion is the predominant isoform. Inclusion of exon 5, as seen in DM1 patients, should result in a product of approximately 170 bp, but a fragment of this size was not detected in any of the transfected 3T3 cells or in the control heart cDNA. A small product of approximately 75bp, not referred to in the literature, was generated using these primers. It was evident in all 3T3 cDNA samples and also in FVBN heart control cDNA (Figure 63, top, question mark). The size of the band could have corresponded to a product of exon 2 spliced directly to exon 6 (calculated using genescan intron-exon prediction on mouse chromosome 1 genomic sequence AC108813.8) yet a product of identical size was also obtained using tail genomic DNA (data not shown), indicating the sequence template lacked introns. In the published literature, figures were tightly cropped (Kanadia *et al.*, 2003; Ho *et al.*, 2005), so it was impossible to ascertain whether this fragment was an artefact in our hands alone. During optimisation of the PCR, where a range of $MgCl_2$ concentrations were tested, the presence and intensity of the small reaction product did not change in proportion to the two larger bands, and always appeared to be the most significant product (Figure 64), indicating good sequence identity with the primers. Increasing the annealing temperature to the calculated melting temperature of the probe (72ºC) had no effect (data not shown), also indicating good sequence identity. Ideally, the fragment could have been excised, cloned and sequenced. Since small PCR products preferentially amplify in a reaction, to try to increase the detection sensitivity of the upper bands, the original gel was Southern blotted and hybridised to a probe generated from the doublet PCR product from FVBN heart (Figure 63*). No Tnnt2 exon 5 inclusion– corresponding to a 170bp fragment –was detected in any of the samples (Figure 64). There is a smudge around this area in the pLoxEGFP250-transfected samples, but the result is not clear and it is difficult to conclude anything from the poor separation of bands, but there clearly are slight differences. Ideally, the PCR products should be separated using polyacrylamide gel electrophoresis, to improve resolution.

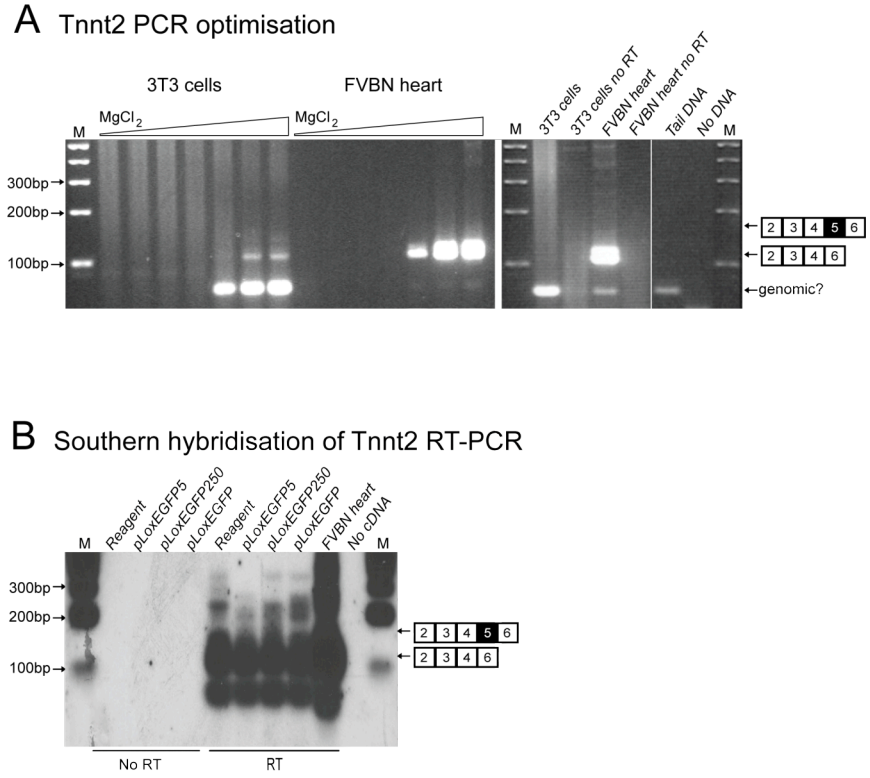


Figure 64 cTnnt2 RT-PCR A A decrease in the magnesium chloride concentration does not reduce the predominance of the 70 bp band, which originates from RNA since no product is present in the no RT control, but contains only genomic sequence and so is not an alternatively spliced product of *cTnnt2*. **B** Southern blot of the *Tnnt2* gel (Figure 63) hybridised to FVBN doublet band (Figure 63*) The 180bp band corresponding to exon 5 inclusion is not present in control samples but may be present in pLoxEGFP250 transfected 3T3 cells as a smudge.

The prepared cDNA samples were also used to ascertain the presence of the *Clcn1* transcript in 3T3 cells, and the effect of CTG repeat expression on *Clcn1* splicing patterns. Briefly, random-primed cDNA prepared from 3T3 cells transfected with pLoxEGFP; pLoxEGFP5, pLoxEGFP250 or reagent alone and FVBN skeletal muscle, was used as a template for the exon 5-8 *Clcn1* primer set published by Kanadia *et al.* 2003. The PCR products were separated by agarose gel electrophoresis. From the 3T3 samples, four bands were obtained (Figure 65). The predicted sizes for isoforms including and excluding exon 7a were calculated using cDNA accession numbers AY046403, AY046404, and the sizes correspond to the smallest PCR products. The larger two products could be further products of alternative exon missplicing implicated in DM1, extended 6-7a and 8a (Mankodi *et al.*, 2002), but identification of the exon content would require sequencing, since information relating to the sequence and the PCR product sizes has not been published. The proportion of alternatively-spliced transcripts in 3T3 cells is different to the skeletal muscle control, which has a much higher incidence of the smallest product corresponding to the adult form of Clcn1, excluding exons 7a and 8a. No variation is seen

between the test samples, therefore expression of expanded CUG repeats has surprisingly no effect on the pattern of *Clcn1* alternative splicing in 3T3 cells.

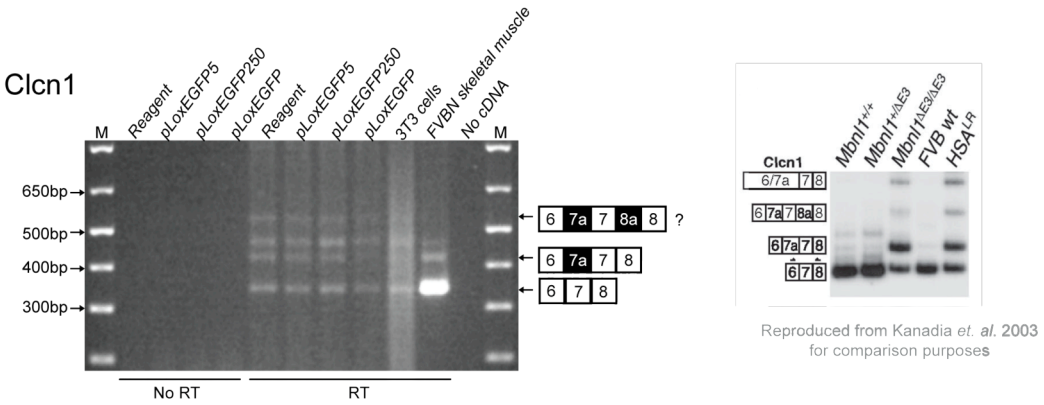


Figure 65 Detection of Clcn1 transcripts in 3T3 cells. Agarose gel separated RT-PCR. 3T3 cells were transfected with constitutively expressing pLoxEGFP constructs containing 0; 5 or 250 CUG repeats and analysed by RT-PCR. Primers were positioned within exons 5 and 8. The predicted sizes for isoforms indicated on the RHS are: 6-7-8, 340 bp; 6-7a-7-8, 419 bp (cDNA accession numbers AY046403, AY046404). The proportion of alternatively-spliced transcripts in 3T3 cells is different to the skeletal muscle control, which has a much higher incidence of the smallest product corresponding to the adult form of *Clcn1* with no inclusion of exons 7a and 8a. No variation is seen between the test samples. Therefore expression of expanded CUG repeats has no effect on the pattern of *Clcn1* alternative splicing in 3T3 cells.

5.5 Discussion

Microarray analyses were carried out in order to identify further novel genes aberrantly spliced after the expression of expanded CUG repeats in a cell culture system. Mouse fibroblast 3T3 cells transfected with 5 or 250 repeats from the pLoxEGFP construct were separated based on EGFP fluorescence. RNA was isolated and used for whole transcript microarray analysis. In the most reliable comparison between EGFP positive fractions pLoxEGFP250+ *vs.* pLoxEGFP5+, 138 genes were up-regulated and 6 down-regulated. No genes were selected for further analysis at this point due to the launch of the more ideally suited exon microarray chips. For this, RNA was isolated from mouse fibroblast 3T3 cells expressing 0, 5 or 250 repeats from the pLoxEGFP construct without FACS separation. Between the test samples: pLoxEGFP0; pLoxEGFP5 and pLoxEGFP250 repeats, 29; 54 and 58 genes showed alternative exon usage within the core (Table 18), extended (Table 19) and full (Table 20) probesets respectively. Six genes were selected from those identified, based on previously known alternative exon usage, and the possible correlation of gene function and symptoms in myotonic dystrophy explored in the literature and discussed.

Since transient transfection does not result in all cells expressing the EGFP construct, in the whole transcript analysis, cells were separated using FACS to enrich the number of cells expressing the reporter with or without repeats. Even with total separation, this would include some cells expressing EGFP with no apparent foci, which has been explained by the heterologous population of repeat lengths in the starting plasmid, where a small percentage of deletions have been identified (Chapter 3: Growing repeats and Chapter 4 EGFP and RNA foci). Since repeat lengths as small as 100 formed foci the small amount of EGFP positive; non-foci containing cells collected during FACS was not expected to skew the results. However, after the analysis, EGFP RT-PCR on the leftover samples revealed EGFP template within the "negative EGFP" cell fractions, indicating that either a significant proportion of "EGFP negative foci positive" cells were present, or that the separation was not pure. Ideally, Cy3-$(CAG)_{10}$ *in situ* hybridisation analysis should have been attempted on "EGFP negative" separated samples before RNA isolation, but it seems that separation is most likely the fault since high numbers of "EGFP negative foci positive" cells were not observed during pLoxEGFP250 transfections in general. If sample separation was very poor then the numbers of differentially expressed genes would be expected to be near zero. In test 4 (Table 2: pLoxEGFP5+ *vs.* pLoxEGFP5-) the differences in the numbers of genes unregulated (130; listed in Table 16) and down-regulated (90; listed in Table 17) were notable however, this was also the case with test 2 (Table 2: pLoxEGFP250+ *vs.* pLoxEGFP5-) the differences in the numbers of genes unregulated (312; listed in Table 14) and down-regulated (79; listed in Table 15), suggesting that some form of enrichment between positive and negative EGFP fractions had occurred. Indeed, the *Cacna1c* gene was identified in both pLoxEGFP5+ *vs.* pLoxEGFP5- (Table 17) and pLoxEGFP250+ *vs.* pLoxEGFP5- (Table 15) down-regulated gene comparisons. It would be prudent then to reanalyse these lists to possibly identify further mis-spliced genes.

The experiment at the time it was designed was well controlled for the possibility of poor separation. PLoxEGFP5 samples were considered to be equivalent to the normal or control situation, and results could then be reanalysed by using the difference between pLoxEGFP5 positive and pLoxEGFP250 positive chips only. Doubt was cast on the negativity of the pLoxEGFP5 samples when Mahadevan *et al.* (2006) reported that over-expression of 5 CUG repeats in the mouse resulted in myotonia, cardiac conduction abnormalities, histopathology and also RNA splicing defects, in the absence of detectable nuclear inclusions. Since our model also utilised the pCMV promoter, it was reasonable to expect that the pLoxEGFP5 could also be pathogenic. For the exon array analysis then,

FACS separation was not carried out, and pLoxEGFP (0 repeats) was added as the negative control.

Myotonic dystrophy affects so many systems, that it is difficult to select a gene whose deficiency does not replicate at least some of the symptoms of DM. Consequently little bias has been used in the selection of genes for further analysis. It is interesting to note that the *Cacna1c* gene, the candidate gene for conduction defects in DM, was also identified in the whole transcript analysis in the down-regulated pLoxEGFP5+ *vs.* pLoxEGFP 5- comparison, and in the down-regulated pLoxEGFP250+ *vs.* pLoxEGFP5- comparison. Here the levels of the *Cacna1c* transcript were reduced 1.75 fold in the 5+ fraction and 1.64 fold in the 250+ fraction compared to out best negative pLoxEGFP5- (Table 17 and Table 15). This was the only gene of those selected from the exon array analysis that was also identified in the whole transcript arrays. This is not surprising since differences in transcript levels using these arrays rely on 3' end alternative splice events or nonsense mediated decay due to the use of cryptic splice sites outside the terminal exon, which is not expected to be a frequent event. In DM1 missplicing identified to date, this happens only in the *chloride channel 1* gene.

It is not clear from the literature whether the splicing differences identified here match existing developmental patterns. As was already mentioned, the patterns in *Mtmr4* are impossible to match at the core probeset level since these intron-derived exons are not represented; this is also true of *Cacna1c.* For the other selected genes, no developmental patterns have been published.

It was noted that in the genes selected that often, alternative exon usage was apparent in both samples compared to the EGFP baseline, but not necessarily in the same exon. Is this an artefact of the selection procedure, or is it statistically significant? It could be an artefact of the analysis. Using Partek Genomic suite, a higher alternative exon usage score may be given to those genes differing between both samples. Of the 25 highest ranking genes from the full data set 21 genes showed alternative exon usage in both samples compared to EGFP, and in different exons, so an artefact does seem likely. In light of this, it would be prudent to repeat the analyses, and separately compare EGFP *vs.* pLoxEGFP5 and EGFP *vs.* pLoxEGFP250, to determine whether the same genes are selected. If the differences are real it may be that we are looking at the result of two different effects: That of increased levels of CUG-BP1 (pLoxEGFP 5), as seen in the recent mouse model, and increased CUG-BP1 combined with depletion of MBNL (pLoxEGFP250).

The results of the exon array analyses were searched for known genes *Clcn1*; *Insr*; *Ryr1*; *Tnnt2*; *Serca* (also known as *atp2a1*) and *Amphiphysin* (*Amph* -also known as *Bin1*). The *p*-values were high (Table 22), indicating that the genes were unlikely to be mis-spliced. Why these genes were not mis-spliced in 3T3 cells is uncertain. It may be that since the origin of the cells was embryonic, the levels of MBNL1 and CUG-BP1 are already 'altered' and that further depletion of MBNL by expanded repeats would make no change to the pattern of splicing. Of those genes known to be misspliced in DM1, there is no information in the literature concerning alternative splicing in 3T3 cells.

Table 22 Alternative splicing *p*-values of genes known to be mis-spliced in DM1

	Core	Extended	Full
Clcn1	0.586979	0.609543	0.535062
Insr	0.395370	0.106255	0.059114
Ryr1	0.396188	0.319517	0.130963
Tnnt2	0.965664	1	0.775952
Serca	0.172591	0.176744	0.675686
Amph	0.492576	0.0217275	0.357725

In the analysis presented in this chapter, core probeset data has been used. Hence before primer design, the alternative exon usage should also be assessed at the extended and the full metaprobeset levels, since changes in these further probesets would affect primer position. Indeed in most of the genes selected, *p*-values are highest in these sets (Table 21). Ideally, the selected genes would next be characterised by RT-PCR using primers surrounding the exons in question, firstly using samples generated with the culture system described, then patient cells and also mouse DM models in order to assess possible isoform changes during development. Candidate cDNAs would then be cloned and sequenced to verify the exon-intron boundaries.

6 Discussion

6.1 DM1 as an RNA processing disorder

Up until fairly recently our understanding of the pathogenesis of myotonic dystrophy was making slow progress. The DM1 locus is a gene dense region, and the effects of the expanded CTG repeat are complex. Expansion of the CTG repeat results in the spread of heterochromatin centered on the CTG repeat region (Otten *et al.*, 1995), resulting in reduced expression of surrounding genes such as *SIX5* (Klesert *et al.*, 1997; Thornton *et al.*, 1997). Since only repeat-containing transcripts are retained in the nucleus, the proportional expression pattern of functionally distinct isoforms of the *DMPK* gene itself is also altered (Groenen *et al.*, 2000; Tiscornia *et al.*, 2000; Wansink *et al.*, 2003; van Herpen *et al.*, 2005), leading to reduced levels of functionally distinct isoforms (Groenen *et al.*, 2000). Despite extensive research into these effects, haploinsufficiency at the DM1 locus alone does not appear to account for the extensive multisystemic effects seen in patients. The main pathogenic effects arise from a toxic gain-of-function of the expanded repeat RNA (Philips *et al.*, 1998; Kuyumcu-Martinez *et al.*, 2006). Expression of the mutant RNA results in nuclear retention of the transcript and sequestration of the splicing regulator MBNL1 to that transcript within foci (Miller *et al.*, 2000; Fardaei *et al.*, 2002). At the same time mutant RNA expression increases the MBNL1 antagonist CUG-BP1, by some unknown mechanism that may involve phosphorylation whilst bound to short single stranded CUG repeats within the cytoplasm (possibly generated by dicer), followed by nuclear relocalisation (Timchenko, personal communication). The alteration in the antagonistic balance between CUG-BP1 and MBNL1 splicing regulators results in a failure of embryonic to adult alternative splicing events in a number of genes, of which there are now many identified in DM1 (Ho *et al.*, 2004; Lin *et al.*, 2006). Some are clearly directly responsible for specific aspects of the pathology, for instance the *chloride channel* to myotonia and the *insulin receptor* to insulin resistance, for others the relationship is less clear (Savkur *et al.*, 2001; Mankodi *et al.*, 2002).

In order to identify further genes misspliced in DM1, we generated a murine cell culture model of myotonic dystrophy type 1 which mimicked the nuclear expanded RNA foci formation and MBNL1 co-localisation seen in patient cells. The formation of foci was observed with as few as 100 CUG repeats in HeLa cells. MBNL1 co-localisation to the foci was apparent using a GFP/MBNL1 fusion protein construct co-transfected with

pLoxEGFP250. Notably, the foci formed in these cells appeared to be larger, and fewer in number, than cells transfected with pLoxEGFP250 alone, suggesting the extent of foci formation was limited by the availability of MBNL1. Using this system and Affymetrix exon array analysis to compare effects of 0; 5 and 250 CUG repeat expression on the transcriptome, we identified 29; 54 and 58 genes that showed alternative exon usage within the core, extended and full probesets respectively. From this list, bioinformatics analysis revealed several nominee genes possibly mis-spliced in DM1 –*MtmR4*, which has possible neuromuscular involvement; *Kcnk4*, *Narg1*, *Ttyh1* and *Bptf*, potentially related to brain development; and *Cacna1c*, a promising candidate for heart conductance defects and sudden death. The discovery of mis-splicing in the pore-forming unit of *Cacna1c,* if confirmed, could be of major consequence to patients undergoing surgery. Identification of the gene as a target in DM allows the anaesthetist to avoid the use of sedatives with effects on this channel and select alternative agents, possibly avoiding the complications of surgery in DM patients altogether.

However DM1 is not just a spliceopathy, *DMPK* insufficiency and DM1 locus chromatin disruption also play a probable role, but there are further questions. CUG-BP1 is a developmental splicing regulator instrumental in the embryonic-adult alternative splicing switch, but CUG-BP1 has another function in addition to this. The protein has deadenylase activity, controlling polyA tail length and mRNA stability (Paillard *et al.*, 2003; Moraes *et al.*, 2006). What are the effects of elevated levels in relation to this role, and are there other functions? Are there as yet unidentified additional functions of MBNL1, MBNL2 or MBNL3, and what would be the consequence of depleted reserves on them? What about RNAi? Small 21nt RNAs have been reported by both Cho *et al.* and Krol *et al.*. Cho *et al.* showed that the DM1 3'UTR is transcribed bi-directionally generating 21nt RNAs in both the normal and expanded allele, and suggested a role for the RNAi pathway in DMPK regulation (Cho *et al.*, 2005; Krol *et al.*, 2007).

Here, other RNA processing disorders are contrasted and compared, including DM2, looking for similarities and dissimilarities to try to gain insight into as yet unanswered questions, and also to solicit new ones.

6.2 Other RNA processing disorders

The pathogenesis of myotonic dystrophy can probably be mostly explained by the combination of deregulated splicing of CUG-BP1 and MBNL1 target genes during

development, and haploinsufficiency of *DMPK* and neighbouring genes. Yet one big question remains: How does the congenital form arise? Missplicing in myotonic dystrophy is in some instances borne from a failure of isoform switching between the embryonic and adult forms of target genes, therefore it is unlikely to be caused by MBNL1 depleted mis-regulated splicing since embryonic forms are entirely appropriate in the neonate. More importantly, there have been no reports of a congenital form in DM2, and therefore in the newborn, DM1 pathogenesis must most likely arise from the differences between myotonic dystrophy type 1 and type 2.

Historically, patients with presentations clinically similar to DM1, but without the *DMPK* expansion were diagnosed with proximal myotonic myopathy (PROMM) (Ricker, 1999). These patients had symptoms similar to DM1, but muscle weakness was mainly proximal, and wasting slight. Over time, with the accumulation of clinical data, the differences between PROMM and DM1 became less significant, and the disease was re-classed as a second type of myotonic dystrophy, DM2 (Ranum *et al.*, 1998; Ricker *et al.*, 1999; Liquori *et al.*, 2001). Type 2 in general is a milder form of the disorder, closely resembling that of adult onset DM1. Whereas DM1 patients initially present because of daytime sleepiness and apathy, or distal weakness and stiffness, or in juvenile cases –developmental delay, DM2 patients first complain of muscle pain, stiffness or fatigue, or lower extremity weakness (Harper, 2001). Distal weakness such as facial muscle involvement is less pronounced as is bulbar involvement, so speech and swallowing are less often affected. Muscle atrophy is less pronounced. Although the muscle defects between the two types differ slightly, other symptoms, such as cataracts; hypogonadism and insulin sensitivity are equivalent (see Finsterer, 2002 and Day *et al.*, 2003 for review).

Both mutations consist of an untranslated expanded repeat, which becomes trapped in the nucleus: A CTG expansion in the 3' UTR of the *DMPK* gene in DM1 and in DM2, a CCTG expansion within intron 1 of the *CNBP* gene. MBNL proteins 1, 2 and 3 are recruited to foci within their respective aggregate mutant transcripts and sequestered (Fardaei *et al.*, 2002). This is believed to result in deregulation of alternative splicing events dynamically controlled by MBNL1 and CUG-BP1 during development. Indeed, in a mouse MBNL1$^{\Delta E3}$ model missplicing of *Clcn1, Tnnt2* and *Tnnt3* was observed in the homozygote, also cataracts; centralised nuclei and split fibres within skeletal muscle, but no atrophy was apparent at 11 weeks. There was also no evidence of neonatal muscle weakness associated with the congenital form (Kanadia *et al.*, 2003). MBNL1 sequestration therefore, can account for the splicing defects, and some muscle histology

relevant to both types 1 and 2, but appears an unlikely explanation for the increase in severity of muscle wasting or CNS involvement in DM1.

It is the differences then that become interesting. In DM1, 50 to ~3000 repeat tracts result in the disease (Harper, 2001). In DM2, between 75 and 11,000 CCTG repeats are required (Liquori *et al.*, 2001). Why are more repeats necessary in DM2 to elicit a pathological response? What of the genes themselves; the flanking regions, expression patterns and expression levels?

The genes at the two DM loci have no apparent similarities. Research indicates that flanking genes around a locus can be responsible for disease pathogenesis, such as in fascioscapularhumeral muscular dystrophy, where deletion of a repressor binding site results in increased transcription of the neighbouring *FRG1* gene (Gabellini *et al.*, 2002). At the DM1 locus however, none of the surrounding genes *RSHL1*, *DMWD or SIX5* have been satisfactorily implicated in the disease. *Six5* insufficiency in mice leads to cataracts as in DM1, but not of the same type. It would be reasonable to use existing mouse models to attempt to create the complete pathogenic spectrum, by perhaps combining existing models MBNL1$^{\Delta E3}$ and the *Six5* or *DMPK* knockout mice.

Phosphorylation of CUG-BP1 by DMPK results in a decrease of nuclear CUG-BP1 *in vitro* (Roberts *et al.*, 1997), so would haploinsufficiency of the kinase result in a reduced requirement for MBNL1 sequestration in order to elicit a response from a smaller number of repeats? In DM2, longer repeats would then be required to deplete MBNL1 to shift the dynamic CUG-BP1/MBNL1 balance as far.

In DM2, only the repeat array is retained within the nucleus, not the *CNBP* transcript – levels of CNBP protein are not affected (Margolis *et al.*, 2006). In DM1 however, the mutant transcript is retained resulting in an alteration of the proportion of functionally distinct DMPK isoforms (Groenen *et al.*, 2000). Most isoforms in normal individuals are expressed in many tissues including heart, skeletal muscle, liver and brain, except for isoform 2 which is only found in the heart and skeletal muscle (Fu *et al.*, 1993; swissprot Q09013), and isoform 14 which is only found in the brain, with high levels in the striatum (cognitive function), cerebellar cortex (motor function) and pons (respiration regulation), all functions which are affected in DM1 (Gennarelli *et al.*, 1995). Alteration of the isoform proportions could lead to functional insufficiency specifically in these tissue types.

FXTAS shares interesting links with DM pathology. Although from the symptoms there is no obvious similarity (gait ataxia and intention tremor), except that the disease is late-onset (in Wells *et al.*, 2006, chapter 10). In FXTAS, miRNA-mediated post-transcriptional control is probably disrupted by over-expression of the miRNA guide protein FMR1 (Plante *et al.*, 2006; Plante *et al.*, 2006). Of DM1, the CTG array is known to form a single hairpin of a type capable of invoking RNAi (Langlois *et al.*, 2005). Could this result in short CUG repeats entering the cytoplasm and lead to an increase of CUG-BP1 protein? Small 21nt CUG RNAs have been detected in DM1 patient cells, and a recent mouse model has shown that over-expression of short 5-repeat CTG arrays is sufficient to mimic DM pathogenesis in mice (Mahadevan *et al.*, 2006). Presumably the 5-repeat transcripts generated in this model play a role in the alteration of cellular CUG-BP1 levels perhaps by phosphorylation of bound protein increasing localisation to the nucleus. These mice suffer from myotonia and splicing defects, but there is no mention of a congenital defect such as brain involvement. This may be due to the design of the transgene whereby repeat transcription is induced at any one point in time, rather than being expressed during development, which incidentally, would be difficult to orchestrate in this model since animals die 3-4 weeks after induction. However, it may also be because production of 5 CUG repeats may not mimic the microRNA function of dicer-generated short 21nt RNAs, possibly generated from CUG hairpins. In addition to increasing nuclear CUG-BP1 levels by binding and phosphorylation, this could lead to protein insufficiency through RNA mediated gene silencing. If *DMPK* is regulated by dicer as indicated by the presence of 21nt CUG RNAs, then expanded repeats may be detrimental to the cellular RNAi process affecting unrelated post transcriptionally-regulated genes in *DMPK* positive tissues.

It has not yet been established whether the DM2 mutant array is capable of forming similar hairpins, if so would short 21nt CCTG repeats be less toxic than CTG repeats? RNAi generation of short cytoplasmic CTG repeats leading to increased CUG-BP1 levels is consistent with the observation that MBNL1 sequestration is separable from misregulated splicing (Ho *et al.*, 2005), since CAG repeats may not lead to increased levels of CUG-BP1. Further research is required to establish whether RNAi gene silencing plays a role in the pathogenesis in either or both DM1 and DM2 and whether this could be implicated in the congenital form, perhaps by the silencing of specific *DMPK* isoforms in the brain and other tissues.

It was mentioned earlier that CUG-BP1 levels increase in concentration and activity within the nucleus. We have seen the effects of over-expression of CUG-BP1 delineated in a

mouse model (Ho *et al.*, 2005). Here the splicing disruption and abnormal muscle histology was reproduced, but effects in other DM systems such as the brain are not apparent since expression was limited to heart and skeletal muscle. What are the effects of increased CUG-BP1 activity in other tissues during development and on other roles of CUG-BP1? Deadenylation by CUG-BP1 was reported in 2003 by Paillard *et al*. Possible over activity by CUG-BP1 could lead to shortened target polyA tails of other genes within the cell, affecting their stability. This is probably the cause of occulopharangeal muscular dystrophy. The PABP2B protein –defective in OPMD, binds with high affinity to the polyA tail of messenger RNA, controlling adenylate addition to approximately 250 nucleotides. In OPMD patients, the protein becomes aggregated with bound RNA and other proteins. It is not clear whether the disease arises from disrupted adenylation, or sequestration of the 'other' unidentified proteins, but symptoms are late onset, and include ptosis, dysphagia and generalised muscle wasting –as in DM. Analysis of DM2 skeletal muscle showed no increase in CUG-BP1 (Lin *et al.*, 2006). If levels are not elevated in DM2, it may be that DM2 symptoms arise from MBNL1 depletion rather than over-expression of CUG-BP1. This would still alter the antagonistic balance between the splicing regulators, but given the additional role of CUG-BP1 in deadenylation, could result in a milder phenotype. It would be interesting to delineate the effects of CUG-BP1 over-expression with respect to deadenylation from splicing regulation. This could be possible by over-expression of MBNL1 in the existing CUG-BP1 over-expressor mouse model.

A protein known to associate with PABPB2 is FRG1 implicated in fascioscapulohumeral muscular dystrophy (FSHD), which is caused by a deletion, rather than an expansion, of *D4Z4* repeats believed to function as a binding site for a transcriptional repressor complex (van Deutekom *et al.*, 1993; Gabellini *et al.*, 2002). The repressor limits transcription of flanking genes *FRG1; FRG2* and *ANT1*, but in mouse models, only over-expression of *FRG1* develops a muscular dystrophy with features characteristic of human FSHD (Gabellini *et al.*, 2006). In this model, missplicing was identified in muscle-related genes *Mtmr1* and *Tnnt3* transcripts in the mice and in FSHD patient muscle cell cultures, genes also mis-spliced in DM (Buj-Bello *et al.*, 2002; Kanadia *et al.*, 2003). In contrast, Osborne *et al.* found that *FRG1* levels in FSHD patient skeletal muscle biopsies were not raised, and splicing patterns in *MTMR1* and *TNNT3* were normal (Osborne *et al.*, 2007). Both sets of data appear robust so it is not clear where the discrepancies arise. The presence of *FRG1* pseudogenes should be noted however, which would require careful consideration for the choice of PCR amplimer design to accurately measure transcript levels.

Huntington disease-like 2 (HDL2) also shares missplicing with DM1. Abnormal splicing is observed in *amyloid precursor protein* (*APP*) and *microtubule associated protein tau* (*MAPT*) (Rudnicki *et al.*, 2007), as in DM1 (Jiang *et al.*, 2004; Leroy *et al.*, 2006). HDL2 is almost indistinguishable from the polyglutamine associated classical Huntington disease, yet the CTG expansion at the 3' end of the *junctophilin-3* (*JPH3*) gene leads variably to polyleucine; polyalanine; or is untranslated depending on the splice variant (Margolis *et al.*, 2001). How the expansion in HDL2 leads to an almost identical phenotype to HD is unclear, but an RNA gain-of function mechanism seems likely. The HDL2 repeat is transcribed in the CTG orientation, which produces RNA inclusions that co-localise with MBNL1. Lengthwise, the expansion in HDL2 is shorter ~60 repeats, which in DM1 terms is only mildly pathogenic. This makes MBNL1 sequestration less likely, yet if the expression levels of the affected gene were higher than *DMPK*, it would corroborate over-expression of 5 repeats being toxic (Mahadevan *et al.*, 2006). Expression patterns differ between the two genes *JPH3* and *DMPK*, which could account for differences in pathology to DM. *DMPK* has low levels of expression in brain, and is high in heart, skeletal muscle and testis (Sarkar *et al.*, 2004). Whereas *JPH*3 is present at high levels in brain, modest in testis and minimal elsewhere (Takeshima *et al.*, 2000).

Expression of the *ATXN8OS* gene with 116 CTG repeats in mice successfully demonstrates the pathogenicity of the non-coding expansion in SCA8 (Moseley *et al.*, 2006). This length of repeat is in the pathogenic range for DM1, but symptoms are not related. In SCA8 movement is affected, –gait, speech hand and eye coordination (in Wells *et al.*, 2006, chapter 28), caused by degeneration of the cerebellum, in DM1 pathogenesis is primarily muscle-related. The differences again –as in HDL2, probably derive from the expression patterns of the two genes. In the mouse model, repeat positive *ATXN8OS* transcripts were only found in the cerebellum; basal ganglia; frontal lobe and parietal lobe (Moseley *et al.*, 2006).

Of the RNA processing disorders discussed here, it seems likely that SCA 8 and HDL2 share a similar RNA gain of function mechanism with DM1 -notably MBNL1 associated RNA foci are apparent in HDL2. The differences in the pathology probably arise from the pattern of mutant RNA expression. Less is known about these pathogenic mechanisms than DM1, so perhaps DM1 pathogenesis may shed light on the disease process in these instances. Two interesting areas have been highlighted since they take further research of DM1 pathogenesis in two directions little studied so far–the involvement of RNAi, and the effects of over-deadenylation by CUG-BP1. The consequences of shortened polyA tails in

OPMD share remarkable similarity to characteristic symptoms of DM1 –ptosis and dysphagia, perhaps hinting at the possibility of over-active deadenylation by CUG-BP1 in DM1. In FXTAS, miRNA-mediated post-transcriptional control is probably disrupted. If *DMPK* is regulated by dicer as indicated by the presence of 21nt CUG RNAs in DM1 patient cells, then expanded repeats may be detrimental to this universal process but limited to *DMPK* positive tissues, which would explain the phenotypic differences between the two diseases. These two pathways are applicable to both myotonic dystrophy type 1 and type 2, but with perhaps variable effect since the expression patterns of *DMPK* and *CNBP* are not the same, which may therefore account for some of the differences between the two types. Our model could be used to address some of these questions. Since the model is based on the expression of expanded CUG repeats within the immediate *DMPK* 3' UTR effects would be limited to this and would not include the consequences of haploinsufficiency of *DMPK* or the surrounding genes. Chromatin disruption would not be an issue since here the expansion is extra-genomic. The model would be ideal for investigation into the role of the RNAi silencing mechanism, both on the transgene itself and RNA levels generally within the transcriptome. Also, with a simple modification, DM2 CCTG repeats could be expressed allowing the assessment of the relative increases of CUG-BP1 after CTG or CCTG expression to help determine whether DM2 is primarily MBNL1-depletion mediated, lacking additional CUG-BP1 effects. This would indicate that the differences between the two types arise from the other role(s) of CUG-BP1, such as increased deadenylation, which could also be investigated using this system. The length of PolyA tails could be determined and the affected genes identified. In this model the cell-type is not restricted. Results could be compared between cell-lines derived from different tissues, and at different developmental stages, perhaps gaining further insight into the elusive pathogenesis of the congenital form.

The progression of our understanding of DM1 pathogenesis as an RNA processing disorder has enabled access to a plethora of information from other research areas within this expanding group of diseases. Insights gained from the similarities and differences has opened up further avenues of research into myotonic dystrophy pathogenesis, and what is already known about DM will facilitate understanding of other diseases within the field. Understanding has not only progressed in areas of disease, but also in the mechanisms of normal human biology such as the role of CUG-BP1 as a splicing regulator. New directions of research into the pathogenesis of myotonic dystrophy should in turn reveal new therapeutic targets, and subsequently, yield new therapies.

References

Abremski, K., R. Hoess and N. Sternberg (1983). "Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination." *Cell* **32**(4): 1301-11.

Aldridge, L. M. (1985). "Anaesthetic problems in myotonic dystrophy. A case report and review of the Aberdeen experience comprising 48 general anaesthetics in a further 16 patients." *Br J Anaesth* **57**(11): 1119-30.

Amack, J. D., A. P. Paguio and M. S. Mahadevan (1999). "Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model [published erratum appears in Hum Mol Genet 1999 Dec;8(13):2573]." *Hum Mol Genet* **8**(11): 1975-84.

Andorfer, C., C. M. Acker, Y. Kress, P. R. Hof, K. Duff and P. Davies (2005). "Cell-cycle reentry and cell death in transgenic mice expressing nonmutant human tau isoforms." *J Neurosci* **25**(22): 5446-54.

Antonini, G., C. Mainero, A. Romano, F. Giubilei, V. Ceschin, F. Gragnani, S. Morino, M. Fiorelli, F. Soscia, A. Di Pasquale*, et al.* (2004). "Cerebral atrophy in myotonic dystrophy: a voxel based morphometric study." *J Neurol Neurosurg Psychiatry* **75**(11): 1611-3.

Antzelevitch, C., G. D. Pollevick, J. M. Cordeiro, O. Casis, M. C. Sanguinetti, Y. Aizawa, A. Guerchicoff, R. Pfeiffer, A. Oliva, B. Wollnik*, et al.* (2007). "Loss-of-function mutations in the cardiac calcium channel underlie a new clinical entity characterized by ST-segment elevation, short QT intervals, and sudden cardiac death." *Circulation* **115**(4): 442-9.

Aslanidis, C., G. Jansen, C. Amemiya, G. Shutler, M. Mahadevan, C. Tsilfidis, C. Chen, J. Alleman, N. G. Wormskamp, M. Vooijs*, et al.* (1992). "Cloning of the essential myotonic dystrophy region and mapping of the putative defect." *Nature* **355**(6360): 548-51.

Bae, S., Y. Bessho, M. Hojo and R. Kageyama (2000). "The bHLH gene Hes6, an inhibitor of Hes1, promotes neuronal differentiation." *Development* **127**(13): 2933-43.

Bao, Y. P., L. J. Cook, D. O'Donovan, E. Uyama and D. C. Rubinsztein (2002). "Mammalian, yeast, bacterial, and chemical chaperones reduce aggregate formation and death in a cell model of oculopharyngeal muscular dystrophy." *J Biol Chem* **277**(14): 12263-9.

Bao, Y. P., S. Sarkar, E. Uyama and D. C. Rubinsztein (2004). "Congo red, doxycycline, and HSP70 over-expression reduce aggregate formation and cell death in cell models of oculopharyngeal muscular dystrophy." *J Med Genet* **41**(1): 47-51.

Barak, O., M. A. Lazzaro, W. S. Lane, D. W. Speicher, D. J. Picketts and R. Shiekhattar (2003). "Isolation of human NURF: a regulator of Engrailed gene expression." *Embo J* **22**(22): 6089-100.

Baskar, J. F., P. P. Smith, G. S. Ciment, S. Hoffmann, C. Tucker, D. J. Tenney, A. M. Colberg-Poley, J. A. Nelson and P. Ghazal (1996). "Developmental analysis of the cytomegalovirus enhancer in transgenic animals." *J Virol* **70**(5): 3215-26.

Beffy, P., C. Barsanti, R. Del Carratore, S. Simi, P. A. Benedetti, L. Benzi, A. Prelle, P. Ciscato and M. Simili (2005). "Expression and localization of myotonic dystrophy protein kinase in human skeletal muscle cells determined with a novel antibody: possible role of the protein in cytoskeleton rearrangements during differentiation." *Cell Biol Int* **29**(9): 742-53.

Begemann, G., N. Paricio, R. Artero, I. Kiss, M. Perez-Alonso and M. Mlodzik (1997). "muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins." *Development* **124**(21): 4321-31.

Benazzouz, A., T. Boraud, P. Dubedat, A. Boireau, J. M. Stutzmann and C. Gross (1995). "Riluzole prevents MPTP-induced parkinsonism in the rhesus monkey: a pilot study." *Eur J Pharmacol* **284**(3): 299-307.

Benders, A. A., R. A. Wevers and J. H. Veerkamp (1996). "Ion transport in human skeletal muscle cells: disturbances in myotonic dystrophy and Brody's disease." *Acta Physiol Scand* **156**(3): 355-67.

Berger, P., S. Bonneick, S. Willi, M. Wymann and U. Suter (2002). "Loss of phosphatase activity in myotubularin-related protein 2 is associated with Charcot-Marie-Tooth disease type 4B1." *Hum Mol Genet* **11**(13): 1569-79.

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

Berul, C. I., C. T. Maguire, J. Gehrmann and S. Reddy (2000). "Progressive Atrioventricular Conduction Block in a Mouse Myotonic Dystrophy Model." *J Interv Card Electrophysiol* **4**(2): 351-58.

Boireau, A., P. Dubedat, F. Bordier, C. Peny, J. M. Miquet, G. Durand, M. Meunier and A. Doble (1994). "Riluzole and experimental parkinsonism: antagonism of MPTP-induced decrease in central dopamine levels in mice." *Neuroreport* **5**(18): 2657-60.

Bonifazi, E., F. Gullotta, L. Vallo, R. Iraci, A. M. Nardone, E. Brunetti, A. Botta and G. Novelli (2006). "Use of RNA fluorescence in situ hybridization in the prenatal molecular diagnosis of myotonic dystrophy type I." *Clin Chem* **52**(2): 319-22.

Borg, J., L. Edstrom, G. S. Butler-Browne and L. E. Thornell (1987). "Muscle fibre type composition, motoneuron firing properties, axonal conduction velocity and refractory period for foot extensor motor units in dystrophia myotonica." *J Neurol Neurosurg Psychiatry* **50**(8): 1036-44.

Bothe, G. W., J. A. Haspel, C. L. Smith, H. H. Wiener and S. J. Burden (2000). "Selective expression of Cre recombinase in skeletal muscle fibers." *Genesis* **26**(2): 165-6.

Botta, A., S. Caldarola, L. Vallo, E. Bonifazi, D. Fruci, F. Gullotta, R. Massa, G. Novelli and F. Loreni (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., L. Vallo, F. Rinaldi, E. Bonifazi, F. Amati, M. Biancolella, S. Gambardella, E. Mancinelli, C. Angelini, G. Meola*, et al.* (2007). "Gene expression analysis in myotonic dystrophy: indications for a common molecular pathogenic pathway in DM1 and DM2." *Gene Expr* **13**(6): 339-51.

Boucher, C. A., S. K. King, N. Carey, R. Krahe, C. L. Winchester, S. Rahman, T. Creavin, P. Meghji, M. E. Bailey, F. L. Chartier*, et al.* (1995). "A novel homeodomain-encoding gene is associated with a large CpG island interrupted by the myotonic dystrophy unstable (CTG)n repeat." *Hum Mol Genet* **4**(10): 1919-25.

Bowater, R. P., A. Jaworski, J. E. Larson, P. Parniewski and R. D. Wells (1997). "Transcription increases the deletion frequency of long CTG.CAG triplet repeats from plasmids in Escherichia coli." *Nucleic Acids Res* **25**(14): 2861-8.

Bowen, T., C. A. Guy, A. G. Cardno, J. B. Vincent, J. L. Kennedy, L. A. Jones, M. Gray, R. D. Sanders, G. McCarthy, K. C. Murphy*, et al.* (2000). "Repeat sizes at CAG/CTG loci CTG18.1, ERDA1 and TGC13-7a in schizophrenia." *Psychiatr Genet* **10**(1): 33-7.

Bowser, R., A. Giambrone and P. Davies (1995). "FAC1, a novel gene identified with the monoclonal antibody Alz50, is developmentally regulated in human brain." *Dev Neurosci* **17**(1): 20-37.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." *Anal Biochem* **72**: 248-54.

Bradley, J. L., J. C. Blake, S. Chamberlain, P. K. Thomas, J. M. Cooper and A. H. Schapira (2000). "Clinical, biochemical and molecular genetic correlations in Friedreich's ataxia." *Hum Mol Genet* **9**(2): 275-82.

Breitling, R., A. Amtmann and P. Herzyk (2004). "Iterative Group Analysis (iGA): a simple tool to enhance sensitivity and facilitate interpretation of microarray experiments." *BMC Bioinformatics* **5**(34): 1-8.

Breitling, R., P. Armengaud, A. Amtmann and P. Herzyk (2004). "Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments." *FEBS Lett* **573**(1-3): 83-92.

Breschel, T. S., M. G. McInnis, R. L. Margolis, G. Sirugo, B. Corneliussen, S. G. Simpson, F. J. McMahon, D. F. MacKinnon, J. F. Xu, N. Pleasant*, et al.* (1997). "A novel, heritable, expanding CTG repeat in an intron of the SEF2-1 gene on chromosome 18q21.1." *Hum Mol Genet* **6**(11): 1855-63.

Bretag, A. H. (1987). "Muscle chloride channels." *Physiol Rev* **67**(2): 618-724.

Brinster, R. L., H. Y. Chen, M. Trumbauer, A. W. Senear, R. Warren and R. D. Palmiter (1981). "Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs." *Cell* **27**(1 Pt 2): 223-31.

Brinster, R. L., H. Y. Chen, M. E. Trumbauer, M. K. Yagle and R. D. Palmiter (1985). "Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs." *Proc Natl Acad Sci U S A* **82**(13): 4438-42.

Brook, J. D., M. E. McCurrach, H. G. Harley, A. J. Buckler, D. Church, H. Aburatani, K. Hunter, V. P. Stanton, J. P. Thirion, T. Hudson*, 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* **68**(4): 799-808.

Bu'Lock, F. A., M. Sood, J. V. De Giovanni and S. H. Green (1999). "Left ventricular diastolic function in congenital myotonic dystrophy." *Arch Dis Child* **80**(3): 267-70.

Buj-Bello, A., V. Biancalana, C. Moutou, J. Laporte and J. L. Mandel (1999). "Identification of novel mutations in the MTM1 gene causing severe and mild forms of X-linked myotubular myopathy." *Hum Mutat* **14**(4): 320-5.

Buj-Bello, A., D. Furling, H. Tronchere, J. Laporte, T. Lerouge, G. S. Butler-Browne and J. L. Mandel (2002). "Muscle-specific alternative splicing of myotubularin-related 1 gene is impaired in DM1 muscle cells." *Hum Mol Genet* **11**(19): 2297-307.

Buj-Bello, A., V. Laugel, N. Messaddeq, H. Zahreddine, J. Laporte, J. F. Pellissier and J. L. Mandel (2002). "The lipid phosphatase myotubularin is essential for skeletal muscle maintenance but not for myogenesis in mice." *Proc Natl Acad Sci U S A* **99**(23): 15060-5.

Bunting, M., K. E. Bernstein, J. M. Greer, M. R. Capecchi and K. R. Thomas (1999). "Targeting genes for self-excision in the germ line." *Genes Dev* **13**(12): 1524-8.

Buxton, J., P. Shelbourne, J. Davies, C. Jones, T. Van Tongeren, C. Aslanidis, P. de Jong, G. Jansen, M. Anvret, B. Riley*, et al.* (1992). "Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy." *Nature* **355**(6360): 547-8.

Campbell, H. D., M. Kamei, C. Claudianos, E. Woollatt, G. R. Sutherland, Y. Suzuki, M. Hida, S. Sugano and I. G. Young (2000). "Human and mouse homologues of the Drosophila melanogaster tweety (tty) gene: a novel gene family encoding predicted transmembrane proteins." *Genomics* **68**(1): 89-92.

Campuzano, V., L. Montermini, Y. Lutz, L. Cova, C. Hindelang, S. Jiralerspong, Y. Trottier, S. J. Kish, B. Faucheux, P. Trouillas*, et al.* (1997). "Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes." *Hum Mol Genet* **6**(11): 1771-80.

Campuzano, V., L. Montermini, M. D. Molto, L. Pianese, M. Cossee, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, A. Monticelli*, et al.* (1996). "Friedreich's ataxia:

autosomal recessive disease caused by an intronic GAA triplet repeat expansion." *Science* **271**(5254): 1423-7.

Catterall, W. A., E. Perez-Reyes, T. P. Snutch and J. Striessnig (2005). "International Union of Pharmacology. XLVIII. Nomenclature and structure-function relationships of voltage-gated calcium channels." *Pharmacol Rev* **57**(4): 411-25.

Cavadini, P., H. A. O'Neill, O. Benada and G. Isaya (2002). "Assembly and iron-binding properties of human frataxin, the protein deficient in Friedreich ataxia." *Hum Mol Genet* **11**(3): 217-27.

Chakrabarti, L., S. J. Knight, A. V. Flannery and K. E. Davies (1996). "A candidate gene for mild mental handicap at the FRAXE fragile site." *Hum Mol Genet* **5**(2): 275-82.

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

Chaudhari, N. and K. G. Beam (1993). "mRNA for cardiac calcium channel is expressed during development of skeletal muscle." *Dev Biol* **155**(2): 507-15.

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

Cheyette, B. N., P. J. Green, K. Martin, H. Garren, V. Hartenstein and S. L. Zipursky (1994). "The Drosophila sine oculis locus encodes a homeodomain-containing protein required for the development of the entire visual system." *Neuron* **12**(5): 977-96.

Cho, D. H., C. P. Thienes, S. E. Mahoney, E. Analau, G. N. Filippova and S. J. Tapscott (2005). "Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF." *Mol Cell* **20**(3): 483-9.

Chung, S., T. Andersson, K. C. Sonntag, L. Bjorklund, O. Isacson and K. S. Kim (2002). "Analysis of different promoter systems for efficient transgene expression in mouse embryonic stem cell lines." *Stem Cells* **20**(2): 139-45.

Cooper, T. A., E. A. Grice, A. V. Philips and R. S. Savkur (2001). Aberrant splicing of the Clc1 chloride channel pre-mRNA in DM1 skeletal muscle: a possible explanation for myotonia. *The Third International Conference on Unstable Microsatellites and Human Disease. April 21st-April 25th*, Noordwijkerhout,The Netherlands.

Cougot, N., E. van Dijk, S. Babajko and B. Seraphin (2004). "'Cap-tabolism'." *Trends Biochem Sci* **29**(8): 436-44.

Damian, M. S., G. Bachmann, D. Herrmann and W. Dorndorf (1993). "Magnetic resonance imaging of muscle and brain in myotonic dystrophy." *J Neurol* **240**(1): 8-12.

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

Davis, B. M., M. E. McCurrach, K. L. Taneja, R. H. Singer and D. E. Housman (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., K. Ricker, J. F. Jacobsen, L. J. Rasmussen, K. A. Dick, W. Kress, C. Schneider, M. C. Koch, G. J. Beilman, A. R. Harrison*, et al.* (2003). "Myotonic dystrophy type 2: molecular, diagnostic and clinical spectrum." *Neurology* **60**(4): 657-64.

de Haro, M., I. Al-Ramahi, B. De Gouyon, L. Ukani, A. Rosa, N. A. Faustino, T. Ashizawa, T. A. Cooper and J. Botas (2006). "MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1." *Hum Mol Genet* **15**(13): 2138-45.

De Sandre-Giovannoli, A., M. Chaouch, S. Kozlov, J. M. Vallat, M. Tazir, N. Kassouri, P. Szepetowski, T. Hammadouche, A. Vandenberghe, C. L. Stewart*, et al.* (2002).

"Homozygous defects in LMNA, encoding lamin A/C nuclear-envelope proteins, cause autosomal recessive axonal neuropathy in human (Charcot-Marie-Tooth disorder type 2) and mouse." *Am J Hum Genet* **70**(3): 726-36.

Dekel, I., Y. Magal, S. Pearson-White, C. P. Emerson and M. Shani (1992). "Conditional conversion of ES cells to skeletal muscle by an exogenous MyoD1 gene." *New Biol* **4**(3): 217-24.

Diebold, R. J., W. J. Koch, P. T. Ellinor, J. J. Wang, M. Muthuchamy, D. F. Wieczorek and A. Schwartz (1992). "Mutually exclusive exon splicing of the cardiac calcium channel alpha 1 subunit gene generates developmentally regulated isoforms in the rat heart." *Proc Natl Acad Sci U S A* **89**(4): 1497-501.

Eriksson, M., T. Ansved, M. Anvret and N. Carey (2001). "A Mammalian Radial Spokehead-Like Gene, RSHL1, at the Myotonic Dystrophy-1 Locus." *Biochem Biophys Res Commun* **281**(4): 835-41.

Eriksson, M., T. Ansved, L. Edstrom, M. Anvret and N. Carey (1999). "Simultaneous analysis of expression of the three myotonic dystrophy locus genes in adult skeletal muscle samples: the CTG expansion correlates inversely with DMPK and 59 expression levels, but not DMAHP levels." *Hum Mol Genet* **8**(6): 1053-60.

Ettaiche, M., K. Fillacier, C. Widmann, C. Heurteaux and M. Lazdunski (1999). "Riluzole improves functional recovery after ischemia in the rat retina." *Invest Ophthalmol Vis Sci* **40**(3): 729-36.

Faber, G. M., J. Silva, L. Livshitz and Y. Rudy (2007). "Kinetic properties of the cardiac L-type Ca2+ channel and its role in myocyte electrophysiology: a theoretical investigation." *Biophys J* **92**(5): 1522-43.

Fan, X., P. Dion, J. Laganiere, B. Brais and G. A. Rouleau (2001). "Oligomerization of polyalanine expanded PABPN1 facilitates nuclear protein aggregation that is associated with cell death." *Hum Mol Genet* **10**(21): 2341-51.

Fan, X., C. Messaed, P. Dion, J. Laganiere, B. Brais, G. Karpati and G. A. Rouleau (2003). "HnRNP A1 and A/B interaction with PABPN1 in oculopharyngeal muscular dystrophy." *Can J Neurol Sci* **30**(3): 244-51.

Fardaei, M., K. Larkin, J. D. Brook and M. G. Hamshere (2001). "In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts." *Nucleic Acids Res* **29**(13): 2766-71.

Fardaei, M., M. T. Rogers, H. M. Thorpe, K. Larkin, M. G. Hamshere, P. S. Harper and J. D. Brook (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.

Filippova, G. N., C. P. Thienes, B. H. Penn, D. H. Cho, Y. J. Hu, J. M. Moore, T. R. Klesert, V. V. Lobanenkov and S. J. Tapscott (2001). "CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus." *Nat Genet* **28**(4): 335-43.

Fink, M., F. Lesage, F. Duprat, C. Heurteaux, R. Reyes, M. Fosset and M. Lazdunski (1998). "A neuronal two P domain K+ channel stimulated by arachidonic acid and polyunsaturated fatty acids." *Embo J* **17**(12): 3297-308.

Finsterer, J. (2002). "Myotonic dystrophy type 2." *Eur J Neurol* **9**(5): 441-7.

Foiry, L., L. Dong, C. Savouret, L. Hubert, H. te Riele, C. Junien and G. Gourdon (2006). "Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice." *Hum Genet* **119**(5): 520-6.

Fortune, M. T., C. Vassilopoulos, M. I. Coolbaugh, M. J. Siciliano and D. G. Monckton (2000). "Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability." *Hum Mol Genet* **9**(3): 439-45.

Fu, Y. H., D. L. Friedman, S. Richards, J. A. Pearlman, R. A. Gibbs, A. Pizzuti, T. Ashizawa, M. B. Perryman, G. Scarlato, R. G. Fenwick, Jr*., 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., D. P. Kuhl, A. Pizzuti, M. Pieretti, J. S. Sutcliffe, S. Richards, A. J. Verkerk, J. J. Holden, R. G. Fenwick, Jr., S. T. Warren*, et al.* (1991). "Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox." *Cell* **67**(6): 1047-58.

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

Fumagalli, E., P. Bigini, S. Barbera, M. De Paola and T. Mennini (2006). "Riluzole, unlike the AMPA antagonist RPR119990, reduces motor impairment and partially prevents motoneuron death in the wobbler mouse, a model of neurodegenerative disease." *Exp Neurol* **198**(1): 114-28.

Gabellini, D., G. D'Antona, M. Moggio, A. Prelle, C. Zecca, R. Adami, B. Angeletti, P. Ciscato, M. A. Pellegrino, R. Bottinelli*, et al.* (2006). "Facioscapulohumeral muscular dystrophy in mice over-expressing FRG1." *Nature* **439**(7079): 973-7.

Gabellini, D., M. R. Green and R. Tupler (2002). "Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle." *Cell* **110**(3): 339-48.

Gatchel, J. R. and H. Y. Zoghbi (2005). "Diseases of unstable repeat expansion: mechanisms and common principles." *Nat Rev Genet* **6**(10): 743-55.

Gavrilov, D. K., X. Shi, K. Das, T. C. Gilliam and C. H. Wang (1998). "Differential SMN2 expression associated with SMA severity." *Nat Genet* **20**(3): 230-1.

Gecz, J., B. A. Oostra, A. Hockey, P. Carbonell, G. Turner, E. A. Haan, G. R. Sutherland and J. C. Mulley (1997). "FMR2 expression in families with FRAXE mental retardation." *Hum Mol Genet* **6**(3): 435-41.

Gennarelli, M., M. Lucarelli, G. Zelano, A. Pizzuti, G. Novelli and B. Dallapiccola (1995). "Different expression of the myotonin protein kinase gene in discrete areas of human brain." *Biochem Biophys Res Commun* **216**(2): 489-94.

Geoffroy, G., A. Barbeau, G. Breton, B. Lemieux, M. Aube, C. Leger and J. P. Bouchard (1976). "Clinical description and roentgenologic evaluation of patients with Friedreich's ataxia." *Can J Neurol Sci* **3**(4): 279-86.

George, A. L., Jr., M. A. Crackower, J. A. Abdalla, A. J. Hudson and G. C. Ebers (1993). "Molecular basis of Thomsen's disease (autosomal dominant myotonia congenita)." *Nat Genet* **3**(4): 305-10.

Gerbasi, V. R. and A. J. Link (2007). "The myotonic dystrophy type 2 protein ZNF9 is part of an ITAF complex that promotes cap-independent translation." *Mol Cell Proteomics* **6**(6): 1049-58.

Gingrich, K. J., S. Tran, I. M. Nikonorov and T. J. Blanck (2005). "Halothane inhibition of recombinant cardiac L-type Ca2+ channels expressed in HEK-293 cells." *Anesthesiology* **103**(6): 1156-66.

Gomes-Pereira, M., M. T. Fortune, L. Ingram, J. P. McAbney and D. G. Monckton (2004). "Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion." *Hum Mol Genet* **13**(16): 1815-25.

Gomes-Pereira, M., M. T. Fortune and D. G. Monckton (2001). "Mouse tissue culture models of unstable triplet repeats: in vitro selection for larger alleles, mutational expansion bias and tissue specificity, but no association with cell division rates." *Hum Mol Genet* **10**(8): 845-54.

Gomes-Pereira, M. and D. G. Monckton (2006). "Chemical modifiers of unstable expanded simple sequence repeats: what goes up, could come down." *Mutat Res* **598**(1-2): 15-34.

Gourdon, G., F. Radvanyi, A. S. Lia, C. Duros, M. Blanche, M. Abitbol, C. Junien and H. Hofmann-Radvanyi (1997). "Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice." *Nat Genet* **15**(2): 190-2.

Groenen, P. J., D. G. Wansink, M. Coerwinkel, W. van den Broek, G. Jansen and B. Wieringa (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.

Gronemeier, M., A. Condie, J. Prosser, K. Steinmeyer, T. J. Jentsch and H. Jockusch (1994). "Nonsense and missense mutations in the muscular chloride channel gene Clc-1 of myotonic mice." *J Biol Chem* **269**(8): 5963-7.

Group, T. H. s. D. C. R. (1993). "A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group." *Cell* **72**(6): 971-83.

Gu, Y., Y. Shen, R. A. Gibbs and D. L. Nelson (1996). "Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island." *Nat Genet* **13**(1): 109-13.

Guiraud-Dogan, C., A. Huguet, M. Gomes-Pereira, E. Brisson, G. Bassez, C. Junien and G. Gourdon (2007). "DM1 CTG expansions affect insulin receptor isoforms expression in various tissues of transgenic mice." *Biochim Biophys Acta* **1772**(11-12): 1183-91.

Gurnett, C. A., S. D. Kahl, R. D. Anderson and K. P. Campbell (1995). "Absence of the skeletal muscle sarcolemma chloride channel ClC-1 in myotonic mice." *J Biol Chem* **270**(16): 9035-8.

Gurwin, E. B., R. B. Fitzsimons, K. S. Sehmi and A. C. Bird (1985). "Retinal telangiectasis in facioscapulohumeral muscular dystrophy with deafness." *Arch Ophthalmol* **103**(11): 1695-700.

Hallauer, P. L. and K. E. Hastings (2000). "Human cytomegalovirus IE1 promoter/enhancer drives variable gene expression in all fiber types in transgenic mouse skeletal muscle." *BMC Genet* **1**: 1.

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

Harada, Y., R. Sutomo, A. H. Sadewa, T. Akutsu, Y. Takeshima, H. Wada, M. Matsuo and H. Nishio (2002). "Correlation between SMN2 copy number and clinical phenotype of spinal muscular atrophy: three SMN2 copies fail to rescue some patients from the disease severity." *J Neurol* **249**(9): 1211-9.

Harinath, S. and S. K. Sikdar (2004). "Trichloroethanol enhances the activity of recombinant human TREK-1 and TRAAK channels." *Neuropharmacology* **46**(5): 750-60.

Harley, H. G., S. A. Rundle, J. C. MacMillan, J. Myring, J. D. Brook, S. Crow, W. Reardon, I. Fenton, D. J. Shaw and P. S. Harper (1993). "Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy." *Am J Hum Genet* **52**(6): 1164-74.

Harley, H. G., S. A. Rundle, W. Reardon, J. Myring, S. Crow, J. D. Brook, P. S. Harper and D. J. Shaw (1992). "Unstable DNA sequence in myotonic dystrophy." *Lancet* **339**(8802): 1125-8.

Harper, P., S. (1989). *Myotonic dystrophy*. London ; Philadelphia, W.B. Saunders.

Harper, P., S. (1998). *Myotonic Dystrophy as a trinucletide repeat disorder - a clinical perspective. In "genetic Instabilities and Hereditary Neurological diseases"*. San Diego, Academic Press.

Harper, P. S. (2001). *Myotonic dystrophy*. London, W. B. Saunders.

Hashem, V. I., E. A. Klysik, W. A. Rosche and R. R. Sinden (2002). "Instability of repeated DNAs during transformation in Escherichia coli." *Mutat Res* **502**(1-2): 39-46.

He, Y. G., Y. F. Xie, Y. Chen, W. Qian, J. H. Lai and D. Y. Tan (2002). "[Cloning and analysis of a novel gene encoding N-terminal acetyltransferase subunit]." *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* **34**(3): 353-7.

Heurteaux, C., C. Laigle, N. Blondeau, G. Jarretou and M. Lazdunski (2006). "Alpha-linolenic acid and riluzole treatment confer cerebral protection and improve survival after focal brain ischemia." *Neuroscience* **137**(1): 241-51.

Hill, M. E., G. A. Creed, T. F. McMullan, A. G. Tyers, D. Hilton-Jones, D. O. Robinson and S. R. Hammans (2001). "Oculopharyngeal muscular dystrophy: phenotypic and genotypic studies in a UK population." *Brain* **124**(Pt 3): 522-6.

Hino, H., K. Araki, E. Uyama, M. Takeya, M. Araki, K. Yoshinobu, K. Miike, Y. Kawazoe, Y. Maeda, M. Uchino*, et al.* (2004). "Myopathy phenotype in transgenic mice expressing mutated PABPN1 as a model of oculopharyngeal muscular dystrophy." *Hum Mol Genet* **13**(2): 181-90.

Ho, T. H., D. Bundman, D. L. Armstrong and T. A. Cooper (2005). "Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy." *Hum Mol Genet* **14**(11): 1539-47.

Ho, T. H., B. N. Charlet, M. G. Poulos, G. Singh, M. S. Swanson and T. A. Cooper (2004). "Muscleblind proteins regulate alternative splicing." *Embo J* **23**(15): 3103-12.

Ho, T. H., R. S. Savkur, M. G. Poulos, M. A. Mancini, M. S. Swanson and T. A. Cooper (2005). "Co-localization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy." *J Cell Sci* **118**(Pt 13): 2923-33.

Hoess, R. H. and K. Abremski (1984). "Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP." *Proc Natl Acad Sci U S A* **81**(4): 1026-9.

Hofmann-Radvanyi, H. and C. Junien (1993). "Myotonic dystrophy: over-expression or/and under-expression? A critical review on a controversial point." *Neuromuscul Disord* **3**(5-6): 497-501.

Hofmann-Radvanyi, H., C. Lavedan, J. P. Rabes, D. Savoy, C. Duros, K. Johnson and C. Junien (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.

Hogan, B. (1994). *Manipulating the mouse embryo : a laboratory manual*. Cold Spring Harbor, N.Y, Cold Spring Harbor Laboratory.

Holmes, S. E., E. O'Hearn, A. Rosenblatt, C. Callahan, H. S. Hwang, R. G. Ingersoll-Ashworth, A. Fleisher, G. Stevanin, A. Brice, N. T. Potter*, et al.* (2001). "A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2." *Nat Genet* **29**(4): 377-8.

Holmes, S. E., E. E. O'Hearn, M. G. McInnis, D. A. Gorelick-Feldman, J. J. Kleiderlein, C. Callahan, N. G. Kwak, R. G. Ingersoll-Ashworth, M. Sherr, A. J. Sumner*, et al.* (1999). "Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12." *Nat Genet* **23**(4): 391-2.

Holmes, S. E., O'Hearn, E. (2000). SCA12: Additional evidence for a causative role of the CAG repeat expansion in PPP2R2B. *American Society for Human Genetics*.

Holmes, S. E., O'Hearn, E. (2003). *In "genetics of Movement Disorder"*, Academic Press, San Diego.

Houseley, J. M., Z. Wang, G. J. Brock, J. Soloway, R. Artero, M. Perez-Alonso, K. M. O'Dell and D. G. Monckton (2005). "Myotonic dystrophy associated expanded CUG repeat muscleblind positive ribonuclear foci are not toxic to Drosophila." *Hum Mol Genet* **14**(6): 873-83.

Howard, R. S. and R. W. Orrell (2002). "Management of motor neurone disease." *Postgrad Med J* **78**(926): 736-41.

Ikeuchi, T., R. Koide, O. Onodera, H. Tanaka, M. Oyake, H. Takano and S. Tsuji (1995). "Dentatorubral-pallidoluysian atrophy (DRPLA). Molecular basis for wide clinical features of DRPLA." *Clin Neurosci* **3**(1): 23-7.

Inukai, A., M. Doyu, T. Kato, Y. Liang, S. Kuru, M. Yamamoto, Y. Kobayashi and G. Sobue (2000). "Reduced expression of DMAHP/SIX5 gene in myotonic dystrophy muscle." *Muscle Nerve* **23**(9): 1421-6.

Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf and T. P. Speed (2003). "Exploration, normalization, and summaries of high density oligonucleotide array probe level data." *Biostatistics* **4**(2): 249-64.

Izumi, Y., S. B. Hammerman, C. O. Kirby, A. M. Benz, J. W. Olney and C. F. Zorumski (2003). "Involvement of glutamate in ischemic neurodegeneration in isolated retina." *Vis Neurosci* **20**(2): 97-107.

Jankovic, J. and C. Hunter (2002). "A double-blind, placebo-controlled and longitudinal study of riluzole in early Parkinson's disease." *Parkinsonism Relat Disord* **8**(4): 271-6.

Jansen, G., P. J. Groenen, D. Bachner, P. H. Jap, M. Coerwinkel, F. Oerlemans, W. van den Broek, B. Gohlsch, D. Pette, J. J. Plomp*, et al.* (1996). "Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice." *Nat Genet* **13**(3): 316-24.

Jansen, G., M. Mahadevan, C. Amemiya, N. Wormskamp, B. Segers, W. Hendriks, K. O'Hoy, S. Baird, L. Sabourin, G. Lennon*, et al.* (1992). "Characterization of the myotonic dystrophy region predicts multiple protein isoform-encoding mRNAs." *Nat Genet* **1**(4): 261-6.

Jansen, G., P. Willems, M. Coerwinkel, W. Nillesen, H. Smeets, L. Vits, C. Howeler, H. Brunner and B. Wieringa (1994). "Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in (CTG)n repeat variation and selection against extreme expansion in sperm." *Am J Hum Genet* **54**(4): 575-85.

Jeffreys, A. J., K. Tamaki, A. MacLeod, D. G. Monckton, D. L. Neil and J. A. Armour (1994). "Complex gene conversion events in germline mutation at human minisatellites." *Nat Genet* **6**(2): 136-45.

Jiang, H., A. Mankodi, M. S. Swanson, R. T. Moxley and C. A. Thornton (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.

Jin, S., M. Shimizu, A. Balasubramanyam and H. F. Epstein (2000). "Myotonic dystrophy protein kinase (DMPK) induces actin cytoskeletal reorganization and apoptotic-like blebbing in lens cells." *Cell Motil Cytoskeleton* **45**(2): 133-48.

Kageyama, R., T. Ohtsuka and K. Tomita (2000). "The bHLH gene Hes1 regulates differentiation of multiple cell types." *Mol Cells* **10**(1): 1-7.

Kanadia, R. N., K. A. Johnstone, A. Mankodi, C. Lungu, C. A. Thornton, D. Esson, A. M. Timmers, W. W. Hauswirth and M. S. Swanson (2003). "A muscleblind knockout model for myotonic dystrophy." *Science* **302**(5652): 1978-80.

Kanadia, R. N., J. Shin, Y. Yuan, S. G. Beattie, T. M. Wheeler, C. A. Thornton and M. S. Swanson (2006). "Reversal of RNA missplicing and myotonia after muscleblind over-expression in a mouse poly(CUG) model for myotonic dystrophy." *Proc Natl Acad Sci U S A* **103**(31): 11748-53.

Kanadia, R. N., C. R. Urbinati, V. J. Crusselle, D. Luo, Y. J. Lee, J. K. Harrison, S. P. Oh and M. S. Swanson (2003). "Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3." *Gene Expr Patterns* **3**(4): 459-62.

Kennedy, L. and P. F. Shelbourne (2000). "Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease?" *Hum Mol Genet* **9**(17): 2539-44.

Kennel, P., F. Revah, G. A. Bohme, R. Bejuit, P. Gallix, J. M. Stutzmann, A. Imperato and J. Pratt (2000). "Riluzole prolongs survival and delays muscle strength deterioration in mice with progressive motor neuronopathy (pmn)." *J Neurol Sci* **180**(1-2): 55-61.

Killestein, J., N. F. Kalkers and C. H. Polman (2005). "Glutamate inhibition in MS: the neuroprotective properties of riluzole." *J Neurol Sci* **233**(1-2): 113-5.

Kimura, T., M. Nakamori, J. D. Lueck, P. Pouliquin, F. Aoike, H. Fujimura, R. T. Dirksen, M. P. Takahashi, A. F. Dulhunty and S. Sakoda (2005). "Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca2+-ATPase in myotonic dystrophy type 1." *Hum Mol Genet* **14**(15): 2189-200.

Kinoshita, M., R. Takahashi, T. Hasegawa, T. Komori, R. Nagasawa, K. Hirose and H. Tanabe (1996). "(CTG)n expansions in various tissues from a myotonic dystrophy patient." *Muscle Nerve* **19**(2): 240-2.

Kiuchi, A., N. Otsuka, Y. Namba, I. Nakano and M. Tomonaga (1991). "Presenile appearance of abundant Alzheimer's neurofibrillary tangles without senile plaques in the brain in myotonic dystrophy." *Acta Neuropathol* **82**(1): 1-5.

Klesert, T. R., D. H. Cho, J. I. Clark, J. Maylie, J. Adelman, L. Snider, E. C. Yuen, P. Soriano and S. J. Tapscott (2000). "Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy." *Nat Genet* **25**(1): 105-9.

Klesert, T. R., A. D. Otten, T. D. Bird and S. J. Tapscott (1997). "Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP." *Nat Genet* **16**(4): 402-6.

Klompe, L., M. Lance, D. van der Woerd, T. Scohy and A. J. Bogers (2007). "Anaesthesiological and ventilatory precautions during cardiac surgery in Steinert's disease." *J Card Surg* **22**(1): 74-5.

Kobrinsky, E., S. Tiwari, V. A. Maltsev, J. B. Harry, E. Lakatta, D. R. Abernethy and N. M. Soldatov (2005). "Differential role of the alpha1C subunit tails in regulation of the Cav1.2 channel by membrane potential, beta subunits, and Ca2+ ions." *J Biol Chem* **280**(13): 12474-85.

Koga, R., Y. Nakao, Y. Kurano, T. Tsukahara, A. Nakamura, S. Ishiura, I. Nonaka and K. Arahata (1994). "Decreased myotonin-protein kinase in the skeletal and cardiac muscles in myotonic dystrophy." *Biochem Biophys Res Commun* **202**(1): 577-85.

Kovtun, I. V. and C. T. McMurray (2001). "Trinucleotide expansion in haploid germ cells by gap repair." *Nat Genet* **27**(4): 407-11.

Krahe, R., T. Ashizawa, C. Abbruzzese, E. Roeder, P. Carango, M. Giacanelli, V. L. Funanage and M. J. Siciliano (1995). "Effect of myotonic dystrophy trinucleotide repeat expansion on DMPK transcription and processing." *Genomics* **28**(1): 1-14.

Krol, J., A. Fiszer, A. Mykowska, K. Sobczak, M. de Mezer and W. J. Krzyzosiak (2007). "Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets." *Mol Cell* **25**(4): 575-86.

Kuhn, R., F. Schwenk, M. Aguet and K. Rajewsky (1995). "Inducible gene targeting in mice." *Science* **269**(5229): 1427-9.

Kuhn, U. and E. Wahle (2004). "Structure and function of poly(A) binding proteins." *Biochim Biophys Acta* **1678**(2-3): 67-84.

Kuyumcu-Martinez, N. M. and T. A. Cooper (2006). "Misregulation of alternative splicing causes pathogenesis in myotonic dystrophy." *Prog Mol Subcell Biol* **44**: 133-59.

La Spada, A. R., E. M. Wilson, D. B. Lubahn, A. E. Harding and K. H. Fischbeck (1991). "Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy." *Nature* **352**(6330): 77-9.

Ladd, A. N., N. Charlet and T. A. Cooper (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.

Ladd, A. N., M. G. Stenberg, M. S. Swanson and T. A. Cooper (2005). "Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development." *Dev Dyn* **233**(3): 783-93.

Ladd, P. D., L. E. Smith, N. A. Rabaia, J. M. Moore, S. A. Georges, R. S. Hansen, R. J. Hagerman, F. Tassone, S. J. Tapscott and G. N. Filippova (2007). "An Antisense Transcript Spanning the CGG Repeat Region of FMR1 is Upregulated in Premutation Carriers but Silenced in Full Mutation Individuals." *Hum Mol Genet* **16**(24):3174-87.

Lam, L. T., Y. C. Pham, T. M. Nguyen and G. E. Morris (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.

Lang-Lazdunski, L., C. Heurteaux, N. Vaillant, C. Widmann and M. Lazdunski (1999). "Riluzole prevents ischemic spinal cord injury caused by aortic crossclamping." *J Thorac Cardiovasc Surg* **117**(5): 881-9.

Langlois, M. A., C. Boniface, G. Wang, J. Alluin, P. M. Salvaterra, J. Puymirat, J. J. Rossi and N. S. Lee (2005). "Cytoplasmic and nuclear retained DMPK mRNAs are targets for RNAi in myotonic dystrophy cells." *J Biol Chem* **280**(17):16949-54.

Langlois, M. A., N. S. Lee, J. J. Rossi and J. Puymirat (2003). "Hammerhead ribozyme-mediated destruction of nuclear foci in myotonic dystrophy myoblasts." *Mol Ther* **7**(5 Pt 1): 670-80.

Laporte, J., L. J. Hu, C. Kretz, J. L. Mandel, P. Kioschis, J. F. Coy, S. M. Klauck, A. Poustka and N. Dahl (1996). "A gene mutated in X-linked myotubular myopathy defines a new putative tyrosine phosphatase family conserved in yeast." *Nat Genet* **13**(2): 175-82.

Lavedan, C., H. Hofmann-Radvanyi, P. Shelbourne, J. P. Rabes, C. Duros, D. Savoy, I. Dehaupas, S. Luce, K. Johnson and C. Junien (1993). "Myotonic dystrophy: size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism." *Am J Hum Genet* **52**(5): 875-83.

Lee, H. C., M. K. Patel, D. J. Mistry, Q. Wang, S. Reddy, J. R. Moorman and J. P. Mounsey (2003). "Abnormal Na channel gating in murine cardiac myocytes deficient in myotonic dystrophy protein kinase." *Physiol Genomics* **12**(2): 147-57.

Lefebvre, S., L. Burglen, S. Reboullet, O. Clermont, P. Burlet, L. Viollet, B. Benichou, C. Cruaud, P. Millasseau, M. Zeviani*, et al.* (1995). "Identification and characterization of a spinal muscular atrophy-determining gene." *Cell* **80**(1): 155-65.

Leroy, O., J. Wang, C. A. Maurage, M. Parent, T. Cooper, L. Buee, N. Sergeant, A. Andreadis and M. L. Caillet-Boudin (2006). "Brain-specific change in alternative splicing of Tau exon 6 in myotonic dystrophy type 1." *Biochim Biophys Acta* **1762**(4): 460-7.

Lesage, F., F. Maingret and M. Lazdunski (2000). "Cloning and expression of human TRAAK, a polyunsaturated fatty acids-activated and mechano-sensitive K(+) channel." *FEBS Lett* **471**(2-3): 137-40.

Leung, T., X. Q. Chen, I. Tan, E. Manser and L. Lim (1998). "Myotonic dystrophy kinase-related Cdc42-binding kinase acts as a Cdc42 effector in promoting cytoskeletal reorganization." *Mol Cell Biol* **18**(1): 130-40.

Lewis, J. D. and E. Izaurralde (1997). "The role of the cap structure in RNA processing and nuclear export." *Eur J Biochem* **247**(2): 461-9.

Lia, A. S., H. Seznec, H. Hofmann-Radvanyi, F. Radvanyi, C. Duros, C. Saquet, M. Blanche, C. Junien and G. Gourdon (1998). "Somatic instability of the CTG repeat

in mice transgenic for the myotonic dystrophy region is age dependent but not correlated to the relative intertissue transcription levels and proliferative capacities." *Hum Mol Genet* **7**(8): 1285-91.

Lim, J. H., A. B. Booker and J. R. Fallon (2005). "Regulating fragile X gene transcription in the brain and beyond." *J Cell Physiol* **205**(2): 170-5.

Lin, X., J. W. Miller, A. Mankodi, R. N. Kanadia, Y. Yuan, R. T. Moxley, M. S. Swanson and C. A. Thornton (2006). "Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy." *Hum Mol Genet* **15**(13): 2087-97.

Liquori, C. L., K. Ricker, M. L. Moseley, J. F. Jacobsen, W. Kress, S. L. Naylor, J. W. Day and L. P. Ranum (2001). "Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9." *Science* **293**(5531): 864-7.

Liu, J. X. and J. F. Gui (2005). "Expression pattern and developmental behaviour of cellular nucleic acid-binding protein (CNBP) during folliculogenesis and oogenesis in fish." *Gene* **356**: 181-92.

Lodi, R., J. M. Cooper, J. L. Bradley, D. Manners, P. Styles, D. J. Taylor and A. H. Schapira (1999). "Deficit of in vivo mitochondrial ATP production in patients with Friedreich ataxia." *Proc Natl Acad Sci U S A* **96**(20): 11492-5.

Lorson, C. L., E. Hahnen, E. J. Androphy and B. Wirth (1999). "A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy." *Proc Natl Acad Sci U S A* **96**(11): 6307-11.

Lorson, C. L., J. Strasswimmer, J. M. Yao, J. D. Baleja, E. Hahnen, B. Wirth, T. Le, A. H. Burghes and E. J. Androphy (1998). "SMN oligomerization defect correlates with spinal muscular atrophy severity." *Nat Genet* **19**(1): 63-6.

Lunt, P. W., P. E. Jardine, M. C. Koch, J. Maynard, M. Osborn, M. Williams, P. S. Harper and M. Upadhyaya (1995). "Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD)." *Hum Mol Genet* **4**(5): 951-8.

Maeda, M., C. S. Taft, E. W. Bush, E. Holder, W. M. Bailey, H. Neville, M. B. Perryman and R. D. Bies (1995). "Identification, tissue-specific expression, and subcellular localization of the 80- and 71-kDa forms of myotonic dystrophy kinase protein." *J Biol Chem* **270**(35): 20246-9.

Mahadevan, M., C. Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barcelo, K. O'Hoy*, 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., C. Amemiya, G. Jansen, L. Sabourin, S. Baird, C. E. Neville, N. Wormskamp, B. Segers, M. Batzer, J. Lamerdin*, et al.* (1993). "Structure and genomic sequence of the myotonic dystrophy (DM kinase) gene." *Hum Mol Genet* **2**(3): 299-304.

Mahadevan, M. S., R. S. Yadava, Q. Yu, S. Balijepalli, C. D. Frenzel-McCardell, T. D. Bourne and L. H. Phillips (2006). "Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy." *Nat Genet* **38**(9): 1066-70.

Mankodi, A., E. Logigian, L. Callahan, C. McClain, R. White, D. Henderson, M. Krym and C. A. Thornton (2000). "Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat." *Science* **289**(5485): 1769-73.

Mankodi, A., M. P. Takahashi, H. Jiang, C. L. Beck, W. J. Bowers, R. T. Moxley, S. C. Cannon and C. A. Thornton (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., C. R. Urbinati, Q. P. Yuan, R. T. Moxley, V. Sansone, M. Krym, D. Henderson, M. Schalling, M. S. Swanson and C. A. Thornton (2001). "Muscleblind

localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2." *Hum Mol Genet* **10**(19): 2165-70.

Manley, K., T. L. Shirley, L. Flaherty and A. Messer (1999). "Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice." *Nat Genet* **23**(4): 471-3.

Mantz, J., A. Cheramy, A. M. Thierry, J. Glowinski and J. M. Desmonts (1992). "Anesthetic properties of riluzole (54274 RP), a new inhibitor of glutamate neurotransmission." *Anesthesiology* **76**(5): 844-8.

Margolis, J. M., B. G. Schoser, M. L. Moseley, J. W. Day and L. P. Ranum (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.

Margolis, R. L., E. O'Hearn, A. Rosenblatt, V. Willour, S. E. Holmes, M. L. Franz, C. Callahan, H. S. Hwang, J. C. Troncoso and C. A. Ross (2001). "A disorder similar to Huntington's disease is associated with a novel CAG repeat expansion." *Ann Neurol* **50**(3): 373-80.

Martorell, L., I. Illa, J. Rosell, J. Benitez, M. J. Sedano and M. Baiget (1996). "Homozygous myotonic dystrophy: clinical and molecular studies of three unrelated cases." *J Med Genet* **33**(9): 783-5.

Martorell, L., D. G. Monckton, J. Gamez and M. Baiget (2000). "Complex patterns of male germline instability and somatic mosaicism in myotonic dystrophy type 1 [In Process Citation]." *Eur J Hum Genet* **8**(6): 423-30.

Martorell, L., D. G. Monckton, J. Gamez, K. J. Johnson, I. Gich, A. L. de Munain and M. Baiget (1998). "Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients." *Hum Mol Genet* **7**(2): 307-12.

Martorell, L., D. G. Monckton, A. Sanchez, A. Lopez De Munain and M. Baiget (2001). "Frequency and stability of the myotonic dystrophy type 1 premutation." *Neurology* **56**(3): 328-35.

Mastroyiannopoulos, N. P., M. L. Feldman, J. B. Uney, M. S. Mahadevan and L. A. Phylactou (2005). "Woodchuck post-transcriptional element induces nuclear export of myotonic dystrophy 3' untranslated region transcripts." *EMBO Rep* **6**(5): 458-63.

Matsuura, T., P. Fang, C. E. Pearson, P. Jayakar, T. Ashizawa, B. B. Roa and D. L. Nelson (2006). "Interruptions in the expanded ATTCT repeat of spinocerebellar ataxia type 10: repeat purity as a disease modifier?" *Am J Hum Genet* **78**(1): 125-9.

Matthews, C. A., J. E. Shaw, J. A. Hooper, I. G. Young, M. F. Crouch and H. D. Campbell (2007). "Expression and evolution of the mammalian brain gene Ttyh1." *J Neurochem* **100**(3): 693-707.

Maurage, C. A., B. Udd, M. M. Ruchoux, P. Vermersch, H. Kalimo, R. Krahe, A. Delacourte and N. Sergeant (2005). "Similar brain tau pathology in DM2/PROMM and DM1/Steinert disease." *Neurology* **65**(10): 1636-8.

Mayer, R. E., P. Hendrix, P. Cron, R. Matthies, S. R. Stone, J. Goris, W. Merlevede, J. Hofsteenge and B. A. Hemmings (1991). "Structure of the 55-kDa regulatory subunit of protein phosphatase 2A: evidence for a neuronal-specific isoform." *Biochemistry* **30**(15): 3589-97.

McGrath, C. F., J. S. Buckman, T. D. Gagliardi, W. J. Bosche, L. V. Coren and R. J. Gorelick (2003). "Human cellular nucleic acid-binding protein Zn2+ fingers support replication of human immunodeficiency virus type 1 when they are substituted in the nucleocapsid protein." *J Virol* **77**(15): 8524-31.

Meira-Lima, I. V., J. Zhao, P. Sham, A. C. Pereira, J. E. Krieger and H. Vallada (2001). "Association and linkage studies between bipolar affective disorder and the polymorphic CAG/CTG repeat loci ERDA1, SEF2-1B, MAB21L and KCNN3." *Mol Psychiatry* **6**(5): 565-9.

Mendlewicz, J., D. Souery, J. Del-Favero, I. Massat, K. Lindblad, C. Engstrom, D. Van den Bossche, R. Adolfsson, M. Schalling and C. Van Broeckhoven (2004). "Expanded RED products and loci containing CAG/CTG repeats on chromosome 17 (ERDA1) and chromosome 18 (CTG18.1) in trans-generational pairs with bipolar affective disorder." *Am J Med Genet B Neuropsychiatr Genet* **128**(1): 71-5.

Messaed, C., P. A. Dion, A. Abu-Baker, D. Rochefort, J. Laganiere, B. Brais and G. A. Rouleau (2007). "Soluble expanded PABPN1 promotes cell death in oculopharyngeal muscular dystrophy." *Neurobiol Dis* **26**(3): 546-57.

Michalowski, S., J. W. Miller, C. R. Urbinati, M. Paliouras, M. S. Swanson and J. Griffith (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., C. R. Urbinati, P. Teng-Umnuay, M. G. Stenberg, B. J. Byrne, C. A. Thornton and M. S. Swanson (2000). "Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy." *Embo J* **19**(17): 4439-48.

Miniou, P., D. Tiziano, T. Frugier, N. Roblot, M. Le Meur and J. Melki (1999). "Gene targeting restricted to mouse striated muscle lineage." *Nucleic Acids Res* **27**(19): e27.

Mitsumoto, H. and J. G. Rabkin (2007). "Palliative care for patients with amyotrophic lateral sclerosis: "prepare for the worst and hope for the best"." *Jama* **298**(2): 207-16.

Miwa, T., T. Koyama and M. Shirai (2000). "Muscle specific expression of Cre recombinase under two actin promoters in transgenic mice." *Genesis* **26**(2): 136-8.

Monani, U. R., C. L. Lorson, D. W. Parsons, T. W. Prior, E. J. Androphy, A. H. Burghes and J. D. McPherson (1999). "A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2." *Hum Mol Genet* **8**(7): 1177-83.

Monckton, D. G., Ashizawa, T. and Siciliano, M.J. (1998). *Murine models for myotonic dystrophy.* San Diego, Academic Press, Inc.

Monckton, D. G., M. I. Coolbaugh, K. T. Ashizawa, M. J. Siciliano and C. T. Caskey (1997). "Hypermutable myotonic dystrophy CTG repeats in transgenic mice." *Nat Genet* **15**(2): 193-6.

Monckton, D. G., R. Neumann, T. Guram, N. Fretwell, K. Tamaki, A. MacLeod and A. J. Jeffreys (1994). "Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism." *Nat Genet* **8**(2): 162-70.

Monckton, D. G., L. J. Wong, T. Ashizawa and C. T. Caskey (1995). "Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses." *Hum Mol Genet* **4**(1): 1-8.

Moraes, K. C., C. J. Wilusz and J. Wilusz (2006). "CUG-BP binds to RNA substrates and recruits PARN deadenylase." *Rna* **12**(6): 1084-91.

Moseley, M. L., T. Zu, Y. Ikeda, W. Gao, A. K. Mosemiller, R. S. Daughters, G. Chen, M. R. Weatherspoon, H. B. Clark, T. J. Ebner*, et al.* (2006). "Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8." *Nat Genet* **38**(7): 758-69.

Murakami, M., T. Ohba, Y. Takahashi, H. Watanabe, I. Miyoshi, S. Nakayama, K. Ono, H. Ito and T. Iijima (2006). "Identification of a cardiac isoform of the murine calcium channel alpha1C (Cav1.2-a) subunit and its preferential binding with the beta2 subunit." *J Mol Cell Cardiol* **41**(1): 115-25.

Nagamitsu, S., T. Matsuura, M. Khajavi, R. Armstrong, C. Gooch, Y. Harati and T. Ashizawa (2000). "A "dystrophic" variant of autosomal recessive myotonia congenita caused by novel mutations in the CLCN1 gene." *Neurology* **55**(11): 1697-703.

Nagy, A., J. Rossant, R. Nagy, W. Abramow-Newerly and J. C. Roder (1993). "Derivation of completely cell culture-derived mice from early-passage embryonic stem cells." *Proc Natl Acad Sci U S A* **90**(18): 8424-8.

Nagy, A. and K. Vintersten (2006). "Murine embryonic stem cells." *Methods Enzymol* **418**: 3-21.

Nakamoto, M., H. Takebayashi, Y. Kawaguchi, S. Narumiya, M. Taniwaki, Y. Nakamura, Y. Ishikawa, I. Akiguchi, J. Kimura and A. Kakizuka (1997). "A CAG/CTG expansion in the normal population." *Nat Genet* **17**(4): 385-6.

Napierala, M. and W. J. Krzyzosiak (1997). "CUG repeats present in myotonin kinase RNA form metastable 'slippery' hairpins." *J Biol Chem* **272**(49): 31079-85.

Narang, M. A., J. D. Waring, L. A. Sabourin and R. G. Korneluk (2000). "Myotonic dystrophy (DM) protein kinase levels in congenital and adult DM patients." *Eur J Hum Genet* **8**(7): 507-12.

Nichols, J., E. P. Evans and A. G. Smith (1990). "Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity." *Development* **110**(4): 1341-8.

O'Cochlain, D. F., C. Perez-Terzic, S. Reyes, G. C. Kane, A. Behfar, D. M. Hodgson, J. A. Strommen, X. K. Liu, W. van den Broek, D. G. Wansink*, et al.* (2004). "Transgenic over-expression of human DMPK accumulates into hypertrophic cardiomyopathy, myotonic myopathy and hypotension traits of myotonic dystrophy." *Hum Mol Genet* **13**(20): 2505-18.

O'Rourke J, R., S. A. Georges, H. R. Seay, S. J. Tapscott, M. T. McManus, D. J. Goldhamer, M. S. Swanson and B. D. Harfe (2007). "Essential role for Dicer during skeletal muscle development." *Dev Biol* **311**(2):359-68.

O'Shea, K. S. (1999). "Embryonic stem cell models of development." *Anat Rec* **257**(1): 32-41.

Osborne, R. J., S. Welle, S. L. Venance, C. A. Thornton and R. Tawil (2007). "Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy." *Neurology* **68**(8): 569-77.

Otten, A. D. and S. J. Tapscott (1995). "Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure." *Proc Natl Acad Sci U S A* **92**(12): 5465-9.

Ouellet, D. L., M. P. Perron, L. A. Gobeil, P. Plante and P. Provost (2006). "MicroRNAs in Gene Regulation: When the Smallest Governs It All." *J Biomed Biotechnol* **2006**(4): 69616.

Ozaita, A. and E. Vega-Saenz de Miera (2002). "Cloning of two transcripts, HKT4.1a and HKT4.1b, from the human two-pore K+ channel gene KCNK4. Chromosomal localization, tissue distribution and functional expression." *Brain Res Mol Brain Res* **102**(1-2): 18-27.

Paillard, L., V. Legagneux and H. Beverley Osborne (2003). "A functional deadenylation assay identifies human CUG-BP as a deadenylation factor." *Biol Cell* **95**(2): 107-13.

Pall, G. S., K. J. Johnson and G. L. Smith (2003). "Abnormal contractile activity and calcium cycling in cardiac myocytes isolated from DMPK knockout mice." *Physiol Genomics* **13**(2): 139-46.

Paradis, H., C. Y. Liu, S. Saika, M. Azhar, T. Doetschman, W. V. Good, R. Nayak, N. Laver, C. W. Kao, W. W. Kao*, et al.* (2002). "Tubedown-1 in remodeling of the developing vitreal vasculature in vivo and regulation of capillary outgrowth in vitro." *Dev Biol* **249**(1): 140-55.

Paul, S., W. Dansithong, D. Kim, J. Rossi, N. J. Webster, L. Comai and S. Reddy (2006). "Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing." *Embo J* **25**(18): 4271-83.

Pellizzoni, L., J. Yong and G. Dreyfuss (2002). "Essential role for the SMN complex in the specificity of snRNP assembly." *Science* **298**(5599): 1775-9.

Pham, Y. C., N. Man, L. T. Lam and G. E. Morris (1998). "Localization of myotonic dystrophy protein kinase in human and rabbit tissues using a new panel of monoclonal antibodies." *Hum Mol Genet* **7**(12): 1957-65.

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

Pinkert, C. A. (1994). *Transgenic animal technology : a laboratory handbook*. San Diego ; London, Academic.

Plante, I., L. Davidovic, D. L. Ouellet, L. A. Gobeil, S. Tremblay, E. W. Khandjian and P. Provost (2006). "Dicer-Derived MicroRNAs Are Utilized by the Fragile X Mental Retardation Protein for Assembly on Target RNAs." *J Biomed Biotechnol* **2006**(4): 64347.

Plante, I. and P. Provost (2006). "Hypothesis: A Role for Fragile X Mental Retardation Protein in Mediating and Relieving MicroRNA-Guided Translational Repression?" *J Biomed Biotechnol* **2006**(4): 16806.

Puccio, H. and M. Koenig (2000). "Recent advances in the molecular pathogenesis of Friedreich ataxia." *Hum Mol Genet* **9**(6): 887-92.

Rajavashisth, T. B., A. K. Taylor, A. Andalibi, K. L. Svenson and A. J. Lusis (1989). "Identification of a zinc finger protein that binds to the sterol regulatory element." *Science* **245**(4918): 640-3.

Raker, V. A., K. Hartmuth, B. Kastner and R. Luhrmann (1999). "Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner." *Mol Cell Biol* **19**(10): 6554-65.

Raker, V. A., G. Plessel and R. Luhrmann (1996). "The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro." *Embo J* **15**(9): 2256-69.

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

Ranum, L. P., P. F. Rasmussen, K. A. Benzow, M. D. Koob and J. W. Day (1998). "Genetic mapping of a second myotonic dystrophy locus." *Nat Genet* **19**(2): 196-8.

Reddy, S., D. B. Smith, M. M. Rich, J. M. Leferovich, P. Reilly, B. M. Davis, K. Tran, H. Rayburn, R. Bronson, D. Cros*, et al.* (1996). "Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy." *Nat Genet* **13**(3): 325-35.

Ricker, K. (1999). "Myotonic dystrophy and proximal myotonic myophathy." *J Neurol* **246**(5): 334-8.

Ricker, K., T. Grimm, M. C. Koch, C. Schneider, W. Kress, C. D. Reimers, W. Schulte-Mattler, B. Mueller-Myhsok, K. V. Toyka and C. R. Mueller (1999). "Linkage of proximal myotonic myopathy to chromosome 3q." *Neurology* **52**(1): 170-1.

Rijli, F. M., P. Dolle, V. Fraulob, M. LeMeur and P. Chambon (1994). "Insertion of a targeting construct in a Hoxd-10 allele can influence the control of Hoxd-9 expression." *Dev Dyn* **201**(4): 366-77.

Robert, M. and P. S. Mathuranath (2007). "Tau and tauopathies." *Neurol India* **55**(1): 11-6.

Roberts, R., N. A. Timchenko, J. W. Miller, S. Reddy, C. T. Caskey, M. S. Swanson and L. T. Timchenko (1997). "Altered phosphorylation and intracellular distribution of a (CUG)n triplet repeat RNA-binding protein in patients with myotonic dystrophy and in myotonin protein kinase knockout mice." *Proc Natl Acad Sci U S A* **94**(24): 13221-6.

Rottbauer, W., K. Baker, Z. G. Wo, M. A. Mohideen, H. F. Cantiello and M. C. Fishman (2001). "Growth and function of the embryonic heart depend upon the cardiac-specific L-type calcium channel alpha1 subunit." *Dev Cell* **1**(2): 265-75.

Rudnicki, D. D., S. E. Holmes, M. W. Lin, C. A. Thornton, C. A. Ross and R. L. Margolis (2007). "Huntington's disease--like 2 is associated with CUG repeat-containing RNA foci." *Ann Neurol* **61**(3): 272-82.

Sabouri, L. A., M. S. Mahadevan, M. Narang, D. S. Lee, L. C. Surh and R. G. Korneluk (1993). "Effect of the myotonic dystrophy (DM) mutation on mRNA levels of the DM gene." *Nat Genet* **4**(3): 233-8.

Sachinidis, A., B. K. Fleischmann, E. Kolossov, M. Wartenberg, H. Sauer and J. Hescheler (2003). "Cardiac specific differentiation of mouse embryonic stem cells." *Cardiovasc Res* **58**(2): 278-91.

Saika, S., S. Saika, C. Y. Liu, M. Azhar, L. P. Sanford, T. Doetschman, R. L. Gendron, C. W. Kao and W. W. Kao (2001). "TGFbeta2 in corneal morphogenesis during mouse embryonic development." *Dev Biol* **240**(2): 419-32.

Sakamoto, N., P. D. Chastain, P. Parniewski, K. Ohshima, M. Pandolfo, J. D. Griffith and R. D. Wells (1999). "Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia." *Mol Cell* **3**(4): 465-75.

Salvatori, S., D. Biral, S. Furlan and O. Marin (1997). "Evidence for localization of the myotonic dystrophy protein kinase to the terminal cisternae of the sarcoplasmic reticulum." *J Muscle Res Cell Motil* **18**(4): 429-40.

Salvatori, S., M. Fanin, C. P. Trevisan, S. Furlan, S. Reddy, J. I. Nagy and C. Angelini (2005). "Decreased expression of DMPK: correlation with CTG repeat expansion and fibre type composition in myotonic dystrophy type 1." *Neurol Sci* **26**(4): 235-42.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). *Molecular cloning : a laboratory manual*. Cold Spring Harbor, Cold Sprng Harbor Laboratory Press.

Santacruz, K., J. Lewis, T. Spires, J. Paulson, L. Kotilinek, M. Ingelsson, A. Guimaraes, M. DeTure, M. Ramsden, E. McGowan*, et al.* (2005). "Tau suppression in a neurodegenerative mouse model improves memory function." *Science* **309**(5733): 476-81.

Sarkar, P. S., B. Appukuttan, J. Han, Y. Ito, C. Ai, W. Tsai, Y. Chai, J. T. Stout and S. Reddy (2000). "Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts." *Nat Genet* **25**(1): 110-4.

Sarkar, P. S., J. Han and S. Reddy (2004). "In situ hybridization analysis of Dmpk mRNA in adult mouse tissues." *Neuromuscul Disord* **14**(8-9): 497-506.

Savkur, R. S., A. V. Philips and T. A. Cooper (2001). "Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy." *Nat Genet* **29**(1): 40-7.

Savkur, R. S., A. V. Philips and T. A. Cooper (2001). Aberrant splicing of the insulin receptor pre-mRNA is associated with insulin resistance in the trinucleotide repeat disorder of myotonic dystrophy. *The Third International Conference on Unstable Microsatellites and Human Disease. April 21st-April 25th.*, Noordwijkerhout,The Netherlands.

Savkur, R. S., A. V. Philips, T. A. Cooper, J. C. Dalton, M. L. Moseley, L. P. Ranum and J. W. Day (2004). "Insulin receptor splicing alteration in myotonic dystrophy type 2." *Am J Hum Genet* **74**(6): 1309-13.

Savouret, C., C. Garcia-Cordier, J. Megret, H. te Riele, C. Junien and G. Gourdon (2004). "MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice." *Mol Cell Biol* **24**(2): 629-37.

Schalling, M., T. J. Hudson, K. H. Buetow and D. E. Housman (1993). "Direct detection of novel expanded trinucleotide repeats in the human genome." *Nat Genet* **4**(2): 135-9.

Schraen-Maschke, S., S. Brique, M. C. Chartier-Harlin, E. Brique, A. Destee and B. Sablonniere (1999). "Analysis of ERDA1, CTG18.1, and uncloned CAG/CTG repeat sequences in familial Parkinson's disease with anticipation." *Am J Med Genet* **88**(6): 738-41.

Schulla, V., E. Renstrom, R. Feil, S. Feil, I. Franklin, A. Gjinovci, X. J. Jing, D. Laux, I. Lundquist, M. A. Magnuson*, et al.* (2003). "Impaired insulin secretion and glucose

tolerance in beta cell-selective Ca(v)1.2 Ca2+ channel null mice." *Embo J* **22**(15): 3844-54.

Seidl, K. J., A. Bottaro, A. Vo, J. Zhang, L. Davidson and F. W. Alt (1998). "An expressed neo(r) cassette provides required functions of the 1gamma2b exon for class switching." *Int Immunol* **10**(11): 1683-92.

Sergeant, N., B. Sablonniere, S. Schraen-Maschke, A. Ghestem, C. A. Maurage, A. Wattez, P. Vermersch and A. Delacourte (2001). "Dysregulation of human brain microtubule-associated tau mRNA maturation in myotonic dystrophy type 1." *Hum Mol Genet* **10**(19): 2143-55.

Seznec, H., O. Agbulut, N. Sergeant, C. Savouret, A. Ghestem, N. Tabti, J. C. Willer, L. Ourth, C. Duros, E. Brisson*, et al.* (2001). "Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities." *Hum Mol Genet* **10**(23): 2717-26.

Seznec, H., A. S. Lia-Baldini, C. Duros, C. Fouquet, C. Lacroix, H. Hofmann-Radvanyi, C. Junien and G. Gourdon (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.

Shaw, D. J., M. McCurrach, S. A. Rundle, H. G. Harley, S. R. Crow, R. Sohn, J. P. Thirion, M. G. Hamshere, A. J. Buckler, P. S. Harper*, et al.* (1993). "Genomic organization and transcriptional units at the myotonic dystrophy locus." *Genomics* **18**(3): 673-9.

Shaw-Smith, C., A. M. Pittman, L. Willatt, H. Martin, L. Rickman, S. Gribble, R. Curley, S. Cumming, C. Dunn, D. Kalaitzopoulos*, et al.* (2006). "Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability." *Nat Genet* **38**(9): 1032-7.

Shelbourne, P., R. Winqvist, E. Kunert, J. Davies, J. Leisti, H. Thiele, H. Bachmann, J. Buxton, B. Williamson and K. Johnson (1992). "Unstable DNA may be responsible for the incomplete penetrance of the myotonic dystrophy phenotype." *Hum Mol Genet* **1**(7): 467-73.

Shimizu, K., W. Chen, A. M. Ashique, R. Moroi and Y. P. Li (2003). "Molecular cloning, developmental expression, promoter analysis and functional characterization of the mouse CNBP gene." *Gene* **307**: 51-62.

Shimizu, M., W. Wang, E. T. Walch, P. W. Dunne and H. F. Epstein (2000). "Rac-1 and Raf-1 kinases, components of distinct signaling pathways, activate myotonic dystrophy protein kinase." *FEBS Lett* **475**(3): 273-7.

Shimokawa, M., S. Ishiura, N. Kameda, M. Yamamoto, N. Sasagawa, N. Saitoh, H. Sorimachi, H. Ueda, S. Ohno, K. Suzuki*, et al.* (1997). "Novel isoform of myotonin protein kinase: gene product of myotonic dystrophy is localized in the sarcoplasmic reticulum of skeletal muscle." *Am J Pathol* **150**(4): 1285-95.

Sidransky, E., C. Burgess, T. Ikeuchi, K. Lindblad, R. T. Long, R. A. Philibert, J. Rapoport, M. Schalling, S. Tsuji and E. I. Ginns (1998). "A triplet repeat on 17q accounts for most expansions detected by the repeat-expansion-detection technique." *Am J Hum Genet* **62**(6): 1548-51.

Soldatov, N. M., A. Bouron and H. Reuter (1995). "Different voltage-dependent inhibition by dihydropyridines of human Ca2+ channel splice variants." *J Biol Chem* **270**(18): 10540-3.

Splawski, I., K. W. Timothy, N. Decher, P. Kumar, F. B. Sachse, A. H. Beggs, M. C. Sanguinetti and M. T. Keating (2005). "Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations." *Proc Natl Acad Sci U S A* **102**(23): 8089-96; discussion 8086-8.

Splawski, I., K. W. Timothy, L. M. Sharpe, N. Decher, P. Kumar, R. Bloise, C. Napolitano, P. J. Schwartz, R. M. Joseph, K. Condouris*, et al.* (2004). "Ca(V)1.2

calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism." *Cell* **119**(1): 19-31.

Steinmeyer, K., R. Klocke, C. Ortland, M. Gronemeier, H. Jockusch, S. Grunder and T. J. Jentsch (1991). "Inactivation of muscle chloride channel by transposon insertion in myotonic mice." *Nature* **354**(6351): 304-8.

Stephenson, F. A. (2006). "Structure and trafficking of NMDA and GABAA receptors." *Biochem Soc Trans* **34**(Pt 5): 877-81.

Striessnig, J. (1999). "Pharmacology, structure and function of cardiac L-type Ca(2+) channels." *Cell Physiol Biochem* **9**(4-5): 242-69.

Sugiura, N., S. M. Adams and R. A. Corriveau (2003). "An evolutionarily conserved N-terminal acetyltransferase complex associated with neuronal development." *J Biol Chem* **278**(41): 40113-20.

Sugiura, N., R. G. Patel and R. A. Corriveau (2001). "N-methyl-D-aspartate receptors regulate a group of transiently expressed genes in the developing brain." *J Biol Chem* **276**(17): 14257-63.

Sumpter, V., A. Kahrs, U. Fischer, U. Kornstadt and R. Luhrmann (1992). "In vitro reconstitution of U1 and U2 snRNPs from isolated proteins and snRNA." *Mol Biol Rep* **16**(4): 229-40.

Suzuki, M. (2006). "The Drosophila tweety family: molecular candidates for large-conductance Ca2+-activated Cl- channels." *Exp Physiol* **91**(1): 141-7.

Takahashi, M. P., T. Kimura, T. Yanagihara and S. Sakoda (1999). "Calcium increase in mouse skeletal muscles by triparanol: a drug to induce myotonic dystrophy-like clinical manifestations." *Neurosci Lett* **272**(2): 87-90.

Takeshima, H., S. Komazaki, M. Nishi, M. Iino and K. Kangawa (2000). "Junctophilins: a novel family of junctional membrane complex proteins." *Mol Cell* **6**(1): 11-22.

Taneja, K. L., M. McCurrach, M. Schalling, D. Housman and R. H. Singer (1995). "Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues." *J Cell Biol* **128**(6): 995-1002.

Tang, Z. Z., M. C. Liang, S. Lu, D. Yu, C. Y. Yu, D. T. Yue and T. W. Soong (2004). "Transcript scanning reveals novel and extensive splice variations in human l-type voltage-gated calcium channel, Cav1.2 alpha1 subunit." *J Biol Chem* **279**(43): 44335-43.

Tassone, F., C. Iwahashi and P. J. Hagerman (2004). "FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS)." *RNA Biol* **1**(2): 103-5.

Tawil, R., J. Forrester, R. C. Griggs, J. Mendell, J. Kissel, M. McDermott, W. King, B. Weiffenbach and D. Figlewicz (1996). "Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group." *Ann Neurol* **39**(6): 744-8.

Thierfelder, L., H. Watkins, C. MacRae, R. Lamas, W. McKenna, H. P. Vosberg, J. G. Seidman and C. E. Seidman (1994). "Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere." *Cell* **77**(5): 701-12.

Thomas, K. R. and M. R. Capecchi (1987). "Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells." *Cell* **51**(3): 503-12.

Thornton, C. A., K. Johnson and R. T. Moxley, 3rd (1994). "Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes." *Ann Neurol* **35**(1): 104-7.

Thornton, C. A., J. P. Wymer, Z. Simmons, C. McClain and R. T. Moxley, 3rd (1997). "Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene." *Nat Genet* **16**(4): 407-9.

Timchenko, L. T., J. W. Miller, N. A. Timchenko, D. R. DeVore, K. V. Datar, L. Lin, R. Roberts, C. T. Caskey and M. S. Swanson (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., Z. J. Cai, A. L. Welm, S. Reddy, T. Ashizawa and L. T. Timchenko (2001). "RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1." *J Biol Chem* **276**(11):7820-6.

Tiscornia, G. and M. S. Mahadevan (2000). "Myotonic dystrophy: the role of the CUG triplet repeats in splicing of a novel DMPK exon and altered cytoplasmic DMPK mRNA isoform ratios." *Mol Cell* **5**(6): 959-67.

Tohgi, H., K. Utsugisawa, A. Kawamorita, M. Yamagata, K. Saitoh and K. Hashimoto (1997). "Effects of CTG trinucleotide repeat expansion in leukocytes on quantitative muscle histopathology in myotonic dystrophy." *Muscle Nerve* **20**(2): 232-4.

Tuluc, P., G. Kern, G. J. Obermair and B. E. Flucher (2007). "Computer modeling of siRNA knockdown effects indicates an essential role of the Ca2+ channel {alpha}2{delta}-1 subunit in cardiac excitation-contraction coupling." *Proc Natl Acad Sci U S A* **104**(26):11091-6.

Tybulewicz, V. L., C. E. Crawford, P. K. Jackson, R. T. Bronson and R. C. Mulligan (1991). "Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene." *Cell* **65**(7): 1153-63.

Tyler, F. H. and F. E. Stephens (1950). "Studies in disorders of muscle. II Clinical manifestations and inheritance of facioscapulohumeral dystrophy in a large family." *Ann Intern Med* **32**(4): 640-60.

Utomo, A. R., A. Y. Nikitin and W. H. Lee (1999). "Temporal, spatial, and cell type-specific control of Cre-mediated DNA recombination in transgenic mice." *Nat Biotechnol* **17**(11): 1091-6.

van den Pol, A. N. and P. K. Ghosh (1998). "Selective neuronal expression of green fluorescent protein with cytomegalovirus promoter reveals entire neuronal arbor in transgenic mice." *J Neurosci* **18**(24): 10640-51.

van der Ven, P. F., G. Jansen, T. H. van Kuppevelt, M. B. Perryman, M. Lupa, P. W. Dunne, H. J. ter Laak, P. H. Jap, J. H. Veerkamp, H. F. Epstein*, et al.* (1993). "Myotonic dystrophy kinase is a component of neuromuscular junctions." *Hum Mol Genet* **2**(11): 1889-94.

van Deutekom, J. C., C. Wijmenga, E. A. van Tienhoven, A. M. Gruter, J. E. Hewitt, G. W. Padberg, G. J. van Ommen, M. H. Hofker and R. R. Frants (1993). "FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit." *Hum Mol Genet* **2**(12): 2037-42.

van Herpen, R. E., R. J. Oude Ophuis, M. Wijers, M. B. Bennink, F. A. van de Loo, J. Fransen, B. Wieringa and D. G. Wansink (2005). "Divergent mitochondrial and endoplasmic reticulum association of DMPK splice isoforms depends on unique sequence arrangements in tail anchors." *Mol Cell Biol* **25**(4): 1402-14.

van Koningsbruggen, S., R. W. Dirks, A. M. Mommaas, J. J. Onderwater, G. Deidda, G. W. Padberg, R. R. Frants and S. M. van der Maarel (2004). "FRG1P is localised in the nucleolus, Cajal bodies, and speckles." *J Med Genet* **41**(4): e46.

van Koningsbruggen, S., K. R. Straasheijm, E. Sterrenburg, N. de Graaf, H. G. Dauwerse, R. R. Frants and S. M. van der Maarel (2007). "FRG1P-mediated aggregation of proteins involved in pre-mRNA processing." *Chromosoma* **116**(1): 53-64.

Vihola, A., G. Bassez, G. Meola, S. Zhang, H. Haapasalo, A. Paetau, E. Mancinelli, A. Rouche, J. Y. Hogrel, P. Laforet*, et al.* (2003). "Histopathological differences of myotonic dystrophy type 1 (DM1) and PROMM/DM2." *Neurology* **60**(11): 1854-7.

Vitali, T., V. Sossi, F. Tiziano, S. Zappata, A. Giuli, M. Paravatou-Petsotas, G. Neri and C. Brahe (1999). "Detection of the survival motor neuron (SMN) genes by FISH: further evidence for a role for SMN2 in the modulation of disease severity in SMA patients." *Hum Mol Genet* **8**(13): 2525-32.

Wahle, E. (1991). "A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation." *Cell* **66**(4): 759-68.

Wahle, E. (1995). "Poly(A) tail length control is caused by termination of processive synthesis." *J Biol Chem* **270**(6): 2800-8.

Wakamiya, M., T. Matsuura, Y. Liu, G. C. Schuster, R. Gao, W. Xu, P. S. Sarkar, X. Lin and T. Ashizawa (2006). "The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10." *Neurology* **67**(4): 607-13.

Wang, J., E. Pegoraro, E. Menegazzo, M. Gennarelli, R. C. Hoop, C. Angelini and E. P. Hoffman (1995). "Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation." *Hum Mol Genet* **4**(4): 599-606.

Wang, Q. and J. Bag (2006). "Ectopic expression of a polyalanine expansion mutant of poly(A)-binding protein N1 in muscle cells in culture inhibits myogenesis." *Biochem Biophys Res Commun* **340**(3): 815-22.

Wang, Y. H. and J. Griffith (1995). "Expanded CTG triplet blocks from the myotonic dystrophy gene create the strongest known natural nucleosome positioning elements." *Genomics* **25**(2): 570-3.

Wansink, D. G., R. E. van Herpen, M. M. Coerwinkel-Driessen, P. J. Groenen, B. A. Hemmings and B. Wieringa (2003). "Alternative splicing controls myotonic dystrophy protein kinase structure, enzymatic activity, and subcellular localization." *Mol Cell Biol* **23**(16): 5489-501.

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

Weiner, A. M., M. L. Allende, T. S. Becker and N. B. Calcaterra (2007). "CNBP mediates neural crest cell expansion by controlling cell proliferation and cell survival during rostral head development." *J Cell Biochem* **102**(6):1553-70.

Wells, R. D. and T. Ashizawa (2006). *Genetic instabilities and hereditary neurological diseases*. San Diego, Calif. ; London, Academic Press.

Wells, R. D. and T. Warren Stephen (1998). *Genetic instabilities and hereditary neurological diseases*. San Diego, Calif. ; London, Academic Press.

Westerlaken, J. H., C. E. Van der Zee, W. Peters and B. Wieringa (2003). "The DMWD protein from the myotonic dystrophy (DM1) gene region is developmentally regulated and is present most prominently in synapse-dense brain areas." *Brain Res* **971**(1): 116-27.

Wirth, B. (2000). "An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA)." *Hum Mutat* **15**(3): 228-37.

Wojciechowska, M., A. Bacolla, J. E. Larson and R. D. Wells (2005). "The myotonic dystrophy type 1 triplet repeat sequence induces gross deletions and inversions." *J Biol Chem* **280**(2): 941-52.

Wong, L. J., T. Ashizawa, D. G. Monckton, C. T. Caskey and C. S. Richards (1995). "Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent." *Am J Hum Genet* **56**(1): 114-22.

Wu, J., T. Tang and I. Bezprozvanny (2006). "Evaluation of clinically relevant glutamate pathway inhibitors in in vitro model of Huntington's disease." *Neurosci Lett* **407**(3): 219-23.

Wysocka, J., T. Swigut, H. Xiao, T. A. Milne, S. Y. Kwon, J. Landry, M. Kauer, A. J. Tackett, B. T. Chait, P. Badenhorst*, et al.* (2006). "A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling." *Nature* **442**(7098): 86-90.

Yan, C. and D. D. Boyd (2006). "Histone H3 acetylation and H3 K4 methylation define distinct chromatin regions permissive for transgene expression." *Mol Cell Biol* **26**(17): 6357-71.

Yuan, Y., S. A. Compton, K. Sobczak, M. G. Stenberg, C. A. Thornton, J. D. Griffith and M. S. Swanson (2007). "Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs." *Nucleic Acids Res* 35(16):5474-86.

Zhao, R., Y. Qi, J. Chen and Z. J. Zhao (2001). "FYVE-DSP2, a FYVE domain-containing dual specificity protein phosphatase that dephosphorylates phosphotidylinositol 3-phosphate." *Exp Cell Res* **265**(2): 329-38.

Zohn, I. E., Y. Li, E. Y. Skolnik, K. V. Anderson, J. Han and L. Niswander (2006). "p38 and a p38-interacting protein are critical for downregulation of E-cadherin during mouse gastrulation." *Cell* **125**(5): 957-69.

Figure 44. Movie

Please copy to hard drive for optimal viewing.