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

Molecular analysis of myotonic dystrophy type 1 patients with an unusual molecular diagnosis

Claudia Braida Mérola

Thesis submitted for the degree of Doctor of Philosophy

July 2008

Abstract

Myotonic dystrophy type 1 (DM1) is the most common form of muscular dystrophy in adults, characterised by multiple tissue involvement and caused by an expansion of a $(CTG)_n$ repeat within the 3'-UTR of the *DMPK* gene (19q13.3). Normal individuals contain between 5 and 35 CTG repeats, whereas the repeats in DM1 patients expand in the range of 50 to several thousands. Longer alleles are very unstable and generally always increase in size when transmitted from parent to child, explaining the phenomenon of anticipation defined by earlier age of onset and an increase in the severity of the symptoms.

Charcot-Marie-Tooth disease (CMT) is a genetically heterogeneous, hereditary motor and sensory neuropathy of the peripheral nervous system. To date, 30 different loci have been mapped and mutations have been identified in more than 20 different genes.

The DM1+CMT++ family is a very unusual three generation family in which all patients co-segregate both DM1 and CMT (LOD score = 7.03). It was postulated that either a single or two closely linked mutations near the *APOC2* marker must be the cause of DM1 and CMT. Southern blot analysis of restriction digested genomic DNA revealed a fragment equivalent to a small CTG expansion (~200-400) at the *DM1* locus, but an expanded allele could not be amplified by PCR. We postulated that the expanded repeats may have predisposed the repeat tract and the flanking regions to further DNA instability, leading to a secondary deletion, insertion and/or rearrangement. These novel mutations might modify the expression of *DMPK* and/or nearby genes explaining the unusual clinical presentation.

To identify the lesion in the DM1+CMT++ family, a variety of molecular approaches was performed. The molecular lesion identified was an insertion of a GC rich region within the CTG repeats. The allele was comprised of a variable number of CTGs at the 5'-end followed by $(GGC)_3$ G $(CCG)_{20}$ $(CCGCTG)_{14}$ $(CTG)_{35}$. Analysis of single molecule separated alleles revealed

that the interrupted 3'-end of the array was stable, while the CTG repeats at the 5'-end were unstable. Postulated mechanisms to explain the DM1 and CMT symptoms in the family were: a novel RNA gain-of-function, and/or a novel effect on the downstream genes.

Finding an imperfect CTG repeat allele in the DM1+CMT++ family led us to suggest that imperfect CTG repeat alleles may not be unique events and other DM1 patients may also contain similar alleles. To investigate this DNA samples from 14 DM1 patients with an unusual molecular diagnosis were analysed. The majority of these patients presented with an imperfect CTG repeat allele containing CCGCTG hexamers and/or CCG repeats. Five patients contained two or three higher order repeats containing between 18 and 30 bp such as $((CTG)_5\ (CCG)_5)$, $((CTG)_2\,(CCGCTG)_4)$ and $((CTG)_5\,(CCG)_2\,(CCGCTG))$. These findings further suggest that imperfect CTG repeat alleles might not be as rare as was previously believed.

The results of this project point out the importance of performing a more detailed molecular characterisation of the DM1 patients, which could lead to the provision of more accurate prognoses and the development of effective therapies.

Table of contents

List of Tables

List of Figures

Dedicated to my husband with love

Acknowledgement

I would like to thank you, Darren, for all your support, encouragement, feedback and patience during my project and even before. Thanks for taking me as a student. I would also like to thank my assessors, Mark and Sean, for your comments and suggestions. Many thanks to everybody in level 5 including past and present members. All of you have been very friendly, helpful and supportive. I have really enjoyed working in level 5. A special thank to Rhoda for helping me with your undergraduate project, and to David for proofreading my thesis. I would also thank you, Richard and Colin, for your valuable comments in our lab meetings.

I would like to express my appreciation to the patients, who provided samples and consented to participate in a research project. I would also like to thank the Frank Spaans and Jean-Louis Mandel groups, who provided their patient samples.

A big thank to my family and friends for all your support, encouragement and love. Federico, there are no words to express how thankful I am, without you I have probably not started with my PhD.

Last but not least, I would also like to thank FBLS for the studentship, and AFM for funding the research work.

Author's declaration

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

Claudia Braida

July 2008

List of abbreviations

μ	Micro (10^{-6})
°C	Degrees Celsius
[α-^{32}P] dCTP	α-^{32}P- labelled 2'-deoxycitidine-5'-triphosphate
ACM	Acute confusional migraine
AFF2	AF4/FMR2 family, member 2
AR	Androgen receptor
AT	Annealing temperature
ATN1	Atrophin 1
ATP1A2	Alpha-2 subunit of the sodium/potassium pump
ATXN1	Ataxin 1
ATXN2	Ataxin 2
ATXN3	Ataxin 3
ATXN7	Ataxin 7
ATXN8OS	Ataxin 8 opposite strand
ATXN10	Ataxin 10
BSA	Bovine serum albumin
CACNA1A	Calcium channel
CAG	Trinucleotide of cytosine, adenosine and guanine
CADASIL	Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy
CCG	Trinucleotide of cytosine, cytosine and guanine
CELFs	CUGBP1 and ETR3-like factors
CGG	Trinucleotide of cytosine, guanine and guanine
CLCN1	Muscle-specific chloride channel
CMT	Charcot-Marie-Tooth disease
CNBP	CCHC-type zinc finger, nucleic acid binding protein
CpG	Cytosine-phosphorus-guanine
CTG	Trinucleotide of cytosine, thymine and guanine
CUG	Trinucleotide of cytosine, uracil and guanine
CUGBP1	CUG triplet repeat RNA-binding protein 1
DM	Myotonic dystrophy
DM1	Myotonic dystrophy type 1
DM2	Myotonic dystrophy type 2

DMPK	Dystrophia myotonica protein kinase
DMSO	Dimethyl sulfoxide
DMWD	Distrophia myotonica-containing WD repeat motif
DNA	Deoxyribonucleic acid
DNM2	Dynamin 2
DRPLA	Dentatorubral-pallidoluysian atrophy
EDTA	Ethylenediaminetetracetic acid
E. coli	*Escherichia coli*
EtBr	Ethidium bromide
ERDA-1	Expanded expanded repeat domain, CAG/CTG 1
FHM1	Hemiplegic migraine type 1
FHM2	Hemiplegic migraine type 2
FMR1	Fragile X mental retardation 1
FMRP	Fragile X mental retardation protein
FRAXA	Fragile X syndrome
FRAXE	Fragile site, folic acid type, rare, FRA(X)(q28)
FRDA	Friedreich ataxia
FXN	Frataxin
FXTAS	Fragile X associated tremor ataxia syndrome
g	Gram
GAA	Trinucleotide of guanine, adenine and adenine
GARS	Glycil-tRNA synthetase
GC	Dinucleotide of guanine and cytosine
GJB1	Gap junction binding 1
GLM	General linear model
HD	Huntington disease
HDL2	Huntington disease type 2
HMSN	Hereditary motor and sensory neuropathy
HSA	Human skeletal actin
HTT	Huntingtin
IQ	Intellectual ability
IR	Insulin receptor
JBS	Jacobsen syndrome
JPH3	Junctophilin 3
KDa	Kilo Dalton

L	Litre
LB	Luria Bertani
LMNA	Lamin A/C nuclear envelope protein
M	Molar
Mb	Megabase
MBNL	Muscle-blind like
MMR	Mismatch repair
m	Mili (10^{-3})
min	Minutes
mRNA	Messenger ribonucleic acid
MW	Molecular weight
MNCV	Motor conduction velocity
MPZ	Myelin protein zero
MTMR1	Mytoubularin-related 1
n	Nano (10^{-9})
NC	Negative control
NCV	Nerve conduction velocity
ND	Not determined
NR	No response with distal nerve stimulation
PCR	Polymerase chain reaction
PMP22	Peripheral myelin protein 22
PPP2R2B	Protein phosphatase 2 (formerly 2A), regulatory subunit B, beta isoform
PROMM	Proximal myotonic myopathy
RNA	Ribonucleic acid
RP-PCR	Repeat primed PCR
SCA	Spinocerebellar ataxia
SDS	Sodium dodecyl sulphate
sec	Seconds
SIX5	Sine oculis related homeobox 5
SMAX1	Spinal and bulbar muscular atrophy, X-linked 1
SNCV	Sensory nerve conduction velocity
SNP	Single nucleotide polymorphism
SP-PCR	Small pool PCR
TBP	TATA box binding protein
TNNT2	Cardiac troponin T

Tris	Tris(hydroxymethyl)amino methane
UTR	Unstranslated region
UV	Ultraviolet

1 Introduction

1.1 Dynamic mutations and human diseases

Dynamic mutations are defined as a change in the mutation rate of the product versus the progenitor. In contrast to conventional mutations, which are static, these mutations are dynamic and consist of a process of multiple steps rather than a single event (Richards and Sutherland 1992). Dynamic mutations have been identified in simple tandem repeat sequences resulting in variation in the number of repeats. If the repeats are located within a gene, an increase in the number might affect the expression of the gene or neighbouring genes resulting in disease. A number of human disorders have been associated with expanded simple tandem repeats including myotonic dystrophy type 1 (DM1), Huntington disease (HD) and fragile X syndrome (table 1-1). Additionally, two loci *ERDA-1* and *CTG18.1* that are not yet associated with a phenotype have also been associated with expanded repeats (Breschel, *et al*. 1997; Nakamoto, *et al*. 1997; Ikeuchi, *et al*. 1998). In the general population, the number of repeats is usually short (between 5 and 30) and polymorphic. However once the repeats reach a certain threshold the alleles become somatically and intergenerationally unstable. In the majority of the diseases caused by dynamic mutations, longer alleles are associated with an increase in the disease severity and decrease in age of onset in successive generations, a phenomenon known as anticipation (Cummings and Zoghbi 2000a; Richards 2001). The expanded repeats are also somatically unstable, with different number of repeats arising within different cells and tissues of the same individual in a process that is tissue-specific, age-dependant and expansion biased. The fact that longer alleles cause more severe phenotype leads to the postulate that somatic mosaicism might contribute to the progression and severity of the disease (Shelbourne and Monckton 2006).

Table 1-1. Human disorders caused by dynamic mutations

Disorder	Type of repeat	Size of repeats			Gene	Reference
		Normal	Pre-mutation	Full mutation		
Dentatorubral-pallidoluysian atrophy (DRPLA)	$(CAG)_n$	7-23	NN	≥53	*ATN1*	(Koide, *et al*. 1994; Nagafuchi, *et al*. 1994)
Huntington disease (HD)	$(CAG)_n$	10-34	36-39	≥40	*HTT*	(HDCRG, 1993)
Huntington disease type 2 (HDL2)	$(CAG)_n$	5-18	NN	≥50	*JPH3*	(Holmes, *et al*. 2001b; Stevanin, *et al*. 2002)
Spinal and bulbar muscular atrophy, X-linked 1 (SMAX1)	$(CAG)_n$	10-36	-	≥38	*AR*	(La Spada, *et al*. 1991)
Spinocerebellar ataxia 1 (SCA1)	$(CAG)_n$	6-39	-	≥40	*ATXN1*	(Orr, *et al*. 1993)
Spinocerebellar ataxia 2 (SCA2)	$(CAG)_n$	15-24	-	≥35	*ATXN2*	(Imbert, *et al*. 1996; Pulst, *et al*. 1996; Sanpei, *et al*. 1996)
Machado-Joseph disease (SCA3)	$(CAG)_n$	13-44	NN	≥55	*ATXN3*	(Kawaguchi, *et al*. 1994)
Spinocerebellar ataxia 7 (SCA7)	$(CAG)_n$	7-19	28-35	≥37	*ATXN7*	(David, *et al*. 1997)
Spinocerebellar ataxia 17 (SCA17)	$(CAG)_n$	25-42	NN	≥47	*TBP*	(Fujigasaki, *et al*. 2001; Nakamura, *et al*. 2001)
Spinocerebellar ataxia 12 (SCA12)	$(CAG)_n$	7-45	NN	≥55	*PPP2P2B*	(Holmes, *et al*. 1999; Holmes, *et al*. 2001a)
Myotonic dystrophy type 1 (DM1)	$(CTG)_n$	5-35	37-50	≥80	*DMPK*	(Aslanidis, *et al*. 1992; Brook, *et al*. 1992; Buxton, *et al*. 1992; Fu, *et al*. 1992; Harley, *et al*. 1992; Mahadevan, *et al*. 1992)
Spinocerebellar ataxia 8 (SCA8)	$(CTG)_n$	16-92	NN	≥100	*ATXN8OS*	(Koob, *et al*. 1999)
Friedreich ataxia (FRDA)	$(GAA)_n$	6-32	33-60	≥200	*FXN*	(Campuzano, *et al*. 1996)
Myotonic dystrophy type 2 (DM2)	$(CCTG)_n$	up to 26	NN	~5,000	*CNBP*	(Liquori, *et al*. 2001)
Spinocerebellar ataxia 10 (SCA10)	$(ATTCT)_n$	10-22	NN	≥800	*ATXN10*	(Matsuura, *et al*. 2000)
Fragile X syndrome *	$(CGG)_n$	5-44	55-200	>200	*FMR1*	(Fu, *et al*. 1991; Verkerk, *et al*. 1991)
Non-syndromic X-linked mental retardation (FRAXE)*	$(CCG)_n$	4-39	31-61	≥200	*AFF2*	(Flynn, *et al*. 1993; Knight, *et al*. 1993)
Jacobsen syndrome (JBS) *	$(CCG)_n$	11	80	≥100	*JBS*	(Jones, *et al*. 2000)

NN (not known); - (not present); *triplet repeats associated with rare folate-sensitive fragile sites (based on OMIM and Entrez gene databases).

The mechanisms involved in the genetic instability of the repeats are not clear. It was postulated that the expanded repeats form unusual alternative structures that are believed to interfere with the normal process of the cell such as replication, recombination and/or repair (Wells 1996; McMurray 1999; Sinden, *et al*. 2002). The mechanisms of pathogenesis vary from disease to disease depending on the particular repeat sequence and the location of the repeat with respect to the gene. Three main mechanisms have been postulated: protein loss-of-function, protein gain-of-function and RNA gain-of-function (Cummings and Zoghbi 2000b; Galvao, *et al*. 2001a; Richards 2001).

1.1.1 Repeat dynamics

1.1.1.1 Intergenerational instability

The phenomenon of anticipation was under debate for almost a century. The first indication of anticipation in DM1 was reported shortly after the description of the disease in 1909. According to Harper, Fleischer reported that adult onset patients were often connected to older relatives with cataracts, suggesting anticipation in the disease (Fleischer 1918). However, Penrose postulated (in 1948) that the anticipation described might not be a true phenomenon and in contrast suggested a simple ascertainment bias (Penrose 1948). Therefore, anticipation was assumed as an artefact for almost forty years until Howeler and colleagues presented enough evidence against ascertainment bias (Howeler, *et al*. 1989). The study comprised the analysis of 14 DM families and showed a decrease in the fertility of affected men in the second or third decade, the congenital cases were always transmitted from affected mothers, and a decrease in the age of onset in the child in more than 60 parent-child pairs. These findings demonstrated that anticipation in DM was a true phenomenon.

Sherman postulated that anticipation in fragile X syndrome was also observed in families, but not as an increase in the severity and a decrease of age of onset as in DM. In contrast, increased penetrance was observed in later generations. Most notable were the normal transmitting males. These were

men who by their position in the pedigree must have carried the mutant X chromosome but did not display the disease. Typically, an asymptomatic male carrier transmitted the mutation to his daughters (generally asymptomatic) and subsequently the daughters transmitted the mutated allele to affected sons. This phenomenon was known as the Sherman paradox. However, at that time, there was no molecular explanation available to explain anticipation either in DM or in fragile X syndrome (Sherman, *et al*. 1984; Sherman, *et al*. 1985). However, shortly afterwards, first the Sherman paradox was resolved in 1991 with the finding that a CGG repeat expansion was responsible for fragile X syndrome, and then in 1992, it was found that anticipation in DM1 was explained by an expansion of CTG repeats (Fu, *et al*. 1991; Verkerk, *et al*. 1991; Aslanidis, *et al*. 1992; Brook, *et al*. 1992; Buxton, *et al*. 1992; Harley, *et al*. 1992). In both cases, the number of repeats increases through generations and the repeats must reach a certain threshold to become pathogenic. Subsequently, anticipation was considered as one of the characteristics of diseases caused by dynamic mutations.

Intergenerational instability can be modified differently according to the sex of the parent transmitting the mutant allele. In some disorders, the repeats are more unstable and prone to long expansions when transmitted by the mother as in fragile X syndrome (Nolin, *et al*. 1996). However, in DM1, short expansions are generally more unstable when transmitted by males and long expansions associated with congenital cases are almost always transmitted by females (Brunner, *et al*. 1993a; Harley, *et al*. 1993; Lavedan, *et al*. 1993; Ashizawa, *et al*. 1994b; Jansen, *et al*. 1994). The polyglutamine diseases are caused by an expansion of CAG repeats within the coding region. In HD in particular, the repeats are highly unstable when transmitted from father to child, but are generally more stable when transmitted by the mother (Kremer, *et al*. 1995). The mechanism involved in the differential parental origin of transmission is not yet fully understood.

It has been a challenge to model intergenerational instability in DM1 mouse models. Small changes in repeat size were widely reproduced in mouse models, however the large intergenerational repeat expansions observed in

DM1 patients have not been recreated in mouse (Gourdon, *et al*. 1997; Monckton, *et al*. 1997). However, it was reported recently that the transgenic mouse line DM300-328 containing 300 CTG repeats in 45 Kb of human genomic sequence from the DM1 locus showed increases in 250-480 CTG repeats trough transmissions. This model reproduced the dramatic increases observed in DM1 patients, but not the parental effects (Gomes-Pereira, *et al*. 2007). Therefore, the validity of this model should be confirmed before it could represent a tool to further investigate the underlying mechanisms of intergenerational instability.

1.1.1.2 Somatic mosaicism

In DM, as well as in other repeat diseases, the expanded alleles are often visualised as a diffuse smear by Southern blot of restriction digested genomic DNA, suggesting somatic mosaicism (Aslanidis, *et al*. 1992; Buxton, *et al*. 1992; Fu, *et al*. 1992; Harley, *et al*. 1992). Subsequently, it was demonstrated by small pool PCR (SP-PCR) that the smears could be resolved into individual alleles containing different numbers or repeats, revealing the instability of the repeats (Monckton, *et al*. 1995b; Wong, *et al*. 1995). The rate of instability is different amongst the disorders. In DM1, somatic mosaicism has been closely examined and it was determined that the expanded alleles in skeletal muscle are always larger than in lymphocytes (Anvret, *et al*. 1993; Ashizawa, *et al*. 1993; Thornton, *et al*. 1994; Monckton, *et al*. 1995b). Additionally, the expanded repeats increase with the age of the individual. In congenital cases sampled at birth, there is practically no heterogeneity and the expanded alleles are detected as single bands instead of smears. However, as the age at sampling increases the level of heterogeneity increases throughout the life of the individual (Wong, *et al*. 1995; Martorell, *et al*. 1998). Therefore somatic mosaicism has a tendency to be age-dependant, tissue-specific and expansion-biased.

Somatic mosaicism has been reproduced in a number of transgenic mouse models and in cell culture (Scherzinger, *et al*. 1997; Lia, *et al*. 1998; Fortune, *et al*. 2000; Kennedy and Shelbourne 2000; Gomes-Pereira, *et al*. 2001). In

DM1 transgenic mouse models, it was shown that the average length in muscle was always higher than in lymphocytes, consistent with DM1 patients (Fortune, *et al*. 2000). In HD knock-in mouse models, it was shown that mice carrying>72<CAG<80 repeats show high levels of instability in the striatum and a low level in peripheral tissues, whereas mice carrying ~150 repeats not only show a high level of instability in the striatum but also in other brain regions and in peripheral tissues (e.g. liver). These results reveal a striking correlation with the disorder phenotype since the earliest and most severe defects are seen in the neurons of the striatum, suggesting that somatic mosaicism probably contributes towards the progression and severity of the disease (Kennedy and Shelbourne 2000; Kennedy, *et al*. 2003). As a result, it was shown that dynamics of the repeats in mouse reproduce the properties observed in humans including age-dependance, tissue-specificity and expansion-bias.

1.1.2 Modifiers of repeat dynamics

The repeat dynamics of simple tandem repeats is a complex process with several factors involved. These factors have been classified into two groups: factors that are directly associated with the repeats (*cis*-acting modifiers) and factors contributing to the instability through interactions with the repeat (*trans*-acting modifiers)(Richards 2001; Cleary and Pearson 2003; Pearson, *et al*. 2005)

1.1.2.1 *Cis*-acting modifiers of genetic instability

A number of *cis*-acting factors have been identified to modify the instability of triplet repeats: the sequence and the number of repeats, the presence of interrupting repeat units and the content of the flanking sequences. Only certain specific triplet repeats are unstable and the most common motif is $(CNG)_n$. This is probably because these repeats can form stable unusual DNA structures which may interfere with normal processes such as replication and repair leading to genetic instability (Chen, *et al*. 1998; Pearson and Sinden 1998; McMurray 1999; Sinden, *et al*. 2002). Shorter alleles are more stable

than longer ones, and interrupted alleles are more stable than pure repeat tracts (Ashley and Warren 1995). However, the repeats become unstable only when they reach a critical size threshold. A study revealed that expanded repeats at *HD*, *SCA7*, *SCA1* and *DM1* loci are more unstable than repeats at the *MJD*, *SBMA*, *CTG18.1* and *DRPLA* loci. The difference in the stability was correlated with the GC content in the flanking DNA; more unstable loci have a higher GC content in the flanking DNA. It is not clear how the level of GC content in the flanking DNA affects the instability of the repeats. Possible mechanisms involve alterations in the chromatin structure and/or methylation status (Brock, *et al*. 1999). The effect of the flanking DNA on the repeat instability was reproduced in DM1 mice models. Monckton and colleagues generated five transgenic mouse lines incorporating expanded CTG repeats derived from the human *DM1* locus. One of the lines (*Dmt-D*) was significantly more unstable than the others. All lines carry similar sizes of CTG repeats and the differences were mediated by the site of integration in the mouse genome. These results indicated that the genetic instability not only depends on the size of the repeats, but also depends on the effects of the flanking DNA (Monckton, *et al*. 1997).

1.1.2.2 *Trans*-acting modifiers of genetic instability

The dynamics of tandem repeats are affected by: sex of transmitting parent, tissue specificities and age of the individual. Therefore, there must be specific factors responsible for these processes, but these factors have not been identified yet (Gomes-Pereira, *et al*. 2006). Other *trans*-acting modifiers were postulated to be proteins involved in cellular processes such as replication, repair and recombination. Until now, the only proteins identified to act as trans-acting factors are: MSH2, MSH3 and PMS2. These proteins are involved in mismatch repair and it was demonstrated that the absence of these proteins in mouse models reduce the level of instability (Manley, *et al*. 1999; van den Broek, *et al*. 2002; Savouret, *et al*. 2003; Wheeler, *et al*. 2003; Gomes-Pereira, *et al*. 2004b; Gomes-Pereira, *et al*. 2007).

1.1.3 Mechanisms of genetic instability

The mechanisms involved in the genesis of expanded repeats are not yet clear. It has been reported that the triplet repeat regions have sequence motifs and symmetry elements that cause an increase in flexibility and the formation of a number of novel structures (at least *in vitro*). The number and complexity of potential structures becomes pronounced as the repeats increase in number and these unusual structures may alter normal processes including replication, recombination and/or repair leading to genetic instability (Wells 1996; McMurray 1999; Sinden, *et al*. 2002).

Data from humans and mouse models revealed that genetic instability generally occurs at different rates in different tissues, and is an expansion-biased and cell-division-independent process. The instability frequently occurs to a higher degree in cells that do not divide such as neurons and muscle cells, in contrast with cells that divide frequently such as lymphocytes (Anvret, *et al*. 1993; Ashizawa, *et al*. 1993; Thornton, *et al*. 1994; Kennedy, *et al*. 2003). A question arises about the mechanism responsible for the variability of the repeats in cells that do not divide or do not undergo recombination. The accumulated evidence does not support a role of replication slippage and recombination contributing to genetic instability. In contrast, a mismatch repair mechanism might be the major mechanism responsible for genetic instability (Manley, *et al*. 1999; van den Broek, *et al*. 2002; Savouret, *et al*. 2003; Wheeler, *et al*. 2003; Gomes-Pereira, *et al*. 2004b).

1.1.4 Mechanism of pathogenesis

The pathways leading to the disease phenotype vary from disease to disease depending on the particular repeat unit and the location of the repeat with respect to the gene. A gain-of-function mechanism was postulated for disorders caused by repeats located in coding regions. In contrast, if the repeats are in non-coding regions, generally multiple tissue dysfunctions are observed and the mechanism may involve: a protein loss-of-function, RNA

gain-of-function level and/or modification of chromatin structure (Cummings and Zoghbi 2000b; Galvao, *et al*. 2001b; Richards 2001). Recent evidence suggests that more than one mechanism might be involved in diseases such as DM1, SCA3, SCA8, SCA10, SCA12 and HDL2 (Gatchel and Zoghbi 2005; Li, *et al*. 2008).

1.1.4.1 Diseases caused by a protein loss-of-function

In the group of diseases caused by protein loss-of-function, the pathogenesis is caused by transcriptional silencing of the gene resulting in a loss of protein function. Diseases in this group include: fragile X syndrome, fragile XE syndrome and Friedreich ataxia (FRDA). Fragile X syndrome is the most widly studied disease in this group and so will be described as an example.

Fragile X syndrome is the most common cause of inherited mental retardation (Morton, *et al*. 1997). The mutation responsible is an expansion of (CGG)n located in the 5' untranslated region of the fragile X mental retardation 1 gene (*FMR1*), which maps to chromosome Xq27.3 (Fu, *et al*. 1991; Verkerk, *et al*. 1991). The length of the repeat is polymorphic in the general population ranging from 6 to 52 repeats, and generally containing 1 to 3 AGG interruptions, typically at every tenth repeat. However, when the repeats expand to more than 230 repeats (full mutation), the CpG island in the *FMR1* promoter region and the repeats become abnormally hypermethylated and generally there is no production of protein, which gives rise to the pathogenic phenotype (Bell, *et al*. 1991; Pieretti, *et al*. 1991; Eichler, *et al*. 1994). Alleles between 60 and 230 repeats (premutation) are generally unmethylated and unstable when they are transmitted through generations (Feng, *et al*. 1995). Initially, the intermediate alleles were believed to be non-pathogenic, however recently they have been associated with ovarian failure in women and fragile X associated tremor ataxia syndrome (FXTAS) in men (Murray 2000; Hagerman and Hagerman 2004). Intermediate alleles are associated with an RNA gain-of-function mechanism and not a loss-of-function (described in section 1.1.4.3).

1.1.4.2 Diseases caused by a protein gain-of-function

The polyglutamine diseases are a group of dominant inherited neurodegenerative diseases caused by an expansion of CAG repeats located in the coding region of the respective gene. These include: HD, dentatorubralpallidoluysian atrophy (DRPLA), spinal and bulbar muscular atrophy, X-linked (SMAX1) and spino-cerebellar ataxia type 1 (SCA1), 2 (SCA2), 3 (SCA3), 6 (SCA6), 7 (SCA7) and 17 (SCA17) (Zoghbi and Orr 2000; Nakamura, *et al*. 2001). A protein gain-of-function was proposed as the main mechanism inducing neurodegeneration in the polyglutamine diseases (Zoghbi and Orr 2000). Generally, the expansions into the pathogenic range induce a conformational change in the protein involved leading to the formation of aggregates within the cells in the brain. These aggregates are sequestered in inclusions with chaperones and components of the ubiquitin-proteosome system. However, there is still controversy regarding the toxicity of these inclusion bodies, some lines of evidence suggest they represent a cellular protective response (Gatchel and Zoghbi 2005; Ross and Poirier 2005). Recent evidence from a SCA3 *Drosophila* model suggests that the expanded CAG repeats in the RNA may also have a pathogenic role in the polyglatumine diseases, contributing to the neuronal degeneration (Li, *et al*. 2008). Therefore, as in other diseases caused by dynamic mutations, more than one mechanism seems to be involved in the pathogenesis.

1.1.4.3 Diseases caused by an RNA gain-of-function

In this group of diseases, the expanded repeats in the RNA are postulated to cause disease by titration of specific RNA CNG-binding proteins that are normally responsible for regulating the processing of multiple RNAs. The diseases included are: myotonic dystrophy type 1 and type 2 (described in detail in section 1.2.2.3), FXTAS, spinocerebellar ataxia type 8 (SCA8), 10 (SCA10), 12 (SCA12) and Huntington disease like 2 (HDL2) (Ranum and Cooper 2006).

FXTAS is a disorder that affects more than one third of the adult male carriers of premutation alleles (55<CGG<200) in the *FMR1* gene. The clinical features

involve intention tremor, ataxia, parkinsonism and generalised brain atrophy (Hagerman, *et al*. 2001). Increased levels of *FMR1* mRNA were shown in individuals with alleles in the premutation range, suggesting an RNA gain-of-function mechanism to explain the pathogenesis (Tassone, *et al*. 2000a; Tassone, *et al*. 2000b; Hegde, *et al*. 2001). Subsequently, intranuclear ubiquitin-positive inclusions containing *FMR1* mRNA were observed in neurons and astrocytes in autopsies of FXTAS patients (Greco, *et al*. 2002). The protein composition of the intranuclear inclusions was investigated by fluorescence-activated flow sorting and mass spectrometry analysis (Iwahashi, *et al*. 2006). More than 20 different proteins were identified including: hnRNP A2, a number of neurofilaments and lamin A/C (proteins associated with different forms of Charcot-Marie-Tooth disease) and MBNL1 one of the RNA binding proteins interacting with CUG repeats that is recruited in DM1 and DM2 nuclear foci (described in section 1.2.2.3). As a result, it was postulated that intermediate alleles might lead to an RNA-mediated neurodegenerative phenotype.

A "knock-in" mouse model previously generated by replacing the endogenous 8 CGG repeats with 98 CGG repeats was investigated further. It was observed that these mice had increased levels of *FMR1* mRNA and intranuclear ubiquitin-positive inclusions consistent with those observed in FXTAS patients (Willemsen, *et al*. 2003). Similarly, a *Drosophila* transgenic line generated using the 5'-UTR of the human *FMR1* gene containing 90 CGG repeats and ~200 bp of flanking DNA, reproduced the neuronal degeneration and the formation of nuclear inclusions (Jin, *et al*. 2003). These results indicated that intermediate CGG repeats are probably responsible for generating the intranuclear inclusions that lead to a neurodegenerative phenotype.

A similar mechanism seems to be present in patients carrying AFF2 intermediate alleles. *AFF2* is a gene located 600 Kb distal to the *FMR1* gene and is associated with non-syndromic X-linked mental retardation (Gecz, *et al*. 1996; Gu, *et al*. 1996; Gecz 2000; Gu and Nelson 2003). The intermediate alleles of AFF2 were believed to be non-pathogenic. However, a higher than expected proportion of patients with Parkinson's disease were shown to carry

AFF2 premutation alleles (~40 CCG) (Annesi, *et al*. 2004). A transgenic *Drosophila* model carrying ~90 CCG repeats was generated to investigate whether AFF2 intermediate alleles may also lead to an RNA-mediated neurodegenerative phenotype similar to the FXTAS. Indeed a neurodegenerative phenotype with nuclear inclusions with similar characteristics to the ones detected in FXTAS was observed. Additionally, it was demonstrated that the co-expression of both CCG and CGG repeats in *Drosophila* rescues independently their neuronal degeneration by the RNAi mechanism. Further investigations are required to explain how the RNA gain-of-function model could explain the phenotype observed in these diseases, and to evaluate the possibility of developing novel therapies using the RNAi pathway (Sofola, *et al*. 2007).

1.2 Myotonic dystrophy

Myotonic dystrophy is an autosomal dominant, inherited disease characterised by myotonia, muscular dystrophy, cataracts, cardiac conduction defects and abnormalities in a number of other organs. DM is a heterogeneous group of disorders with two different genetic causes identified so far. These are a CTG repeat expansion in the *DMPK* gene causing myotonic dystrophy type 1 (DM1) and a CCTG repeat expansion in the *CNBP* gene causing myotonic dystrophy type 2 (DM2). Both mutations consist of repeat tracts that are transcribed, but not translated, suggesting a similar mechanism of pathogenesis based on a gain-of-function by toxic RNA containing either expanded CUG or CCUG repeats. It could be possible that there are other forms of myotonic dystrophy (Ashizawa and Harper 2006). For instance, Le Beer and colleagues identified a French family with a phenotype consistent with DM and an additional frontotemporal dementia. The patients did not have the DM1 or DM2 mutations and have been mapped to chromosome 15q21-24 (Le Ber, *et al*. 2004).

1.2.1 Myotonic dystrophy type 1

1.2.1.1 Clinical manifestations

DM1 is the most common form of muscular dystrophy in adults and is characterised by multiple tissue involvement including skeletal myotonia, progressive myopathy, cataracts and abnormalities in the heart, brain and endocrine systems. Clinical expression of the disorder is variable and the patients are classified according to the age of onset as: late-onset, adult-onset, childhood-onset or congenital (Harper 2001).

In the late-onset form, the most common finding is a distinctive form of cataract characterised by discrete subcapsular opacities that are iridescent under slit lamp examination (Harper 2001).

In the adult-onset form, the symptoms become evident in mid-life, but sometimes some signs may become apparent in the second decade. Initially in DM1, the distal muscles of the extremities are affected and then the proximal musculature. A general weakness and atrophy of the facial muscles causes the characteristic haggard appearance. The involvement of the extraocular muscles produces ptosis, weakness of eyelid closure. The weakness of facial, jaw and palatal muscles generally result in speech disturbances, which are sometimes confused with a lower IQ. Myotonia, a delayed ability to relax the grip, is commonly observed in the tongue, forearm and hand. Myotonia is the most characteristic symptom of DM and can be detected with an electromyogram. The central nervous system is also involved and several abnormalities can be observed: daytime somnolence (one of the symptoms about which relatives complain most), fatigue, depression and personality changes. Other systemic features include: cardiac conduction defects, insulin resistance and gastrointestinal abnormalities. The gastrointestinal involvement could directly or indirectly be the cause of death by aspiration due to pharyngo-oesophageal dysfunction. General anaesthesia can be life threatening if certain anaesthetics are used (Harper 2001). In males, early balding and testicular atrophy can also be present.

In the childhood-onset form, symptoms such as facial weakness, myotonia and cardiac conduction defects as in the adult-onset form are also evident. In addition, low IQ and other psychological problems are present.

The clinical features of the congenital form are very severe and include generalised hypotonia, respiratory problems, craniofacial abnormalities, feeding difficulties and talipes. Those whose survive early childhood will develop mental retardation, and the symptoms of the classical form when they reach adulthood (Harper 2001).

1.2.1.2 Genetic features

In 1992, the DM1 mutation was identified as an expansion of a $(CTG)_n$ repeat located in the 3'-untranslated region (UTR) of the *DM protein kinase* gene (*DMPK*) (Aslanidis, *et al*. 1992; Brook, *et al*. 1992; Buxton, *et al*. 1992; Harley, *et al*. 1992) and in the promoter of the downstream gene *SIX5* mapped to chromosome 19q13.3 (figure 1-1) (Boucher, *et al*. 1995). In normal individuals, the *DM1* locus is highly polymorphic and contains 5 to 35 copies of the repeat. In DM1 patients, the repeat is expanded in the range of 50 to 100 in those mildly affected, 200 to 1000 in those with the adult classical form and more than 1000 in those with the congenital form (IDMC, 2000). The expanded alleles (with >50 repeats) are genetically highly unstable and nearly always increase in length when transmitted from parent to child, resulting in anticipation with an increase of disease severity and earlier age of onset in successive generations. In the first generation, the affected patients present few or no symptoms. In the second generation, the classical adult form is observed. In the third generation, a congenital case is expected through maternal transmissions (Harper 2001). However, contractions occasionally occur through transmissions (less than 10%) (Ashizawa, *et al*. 1994a), and very rarely a reverse mutation (decrease in size to the normal range) can occur (Shelbourne, *et al*. 1992; Brunner, *et al*. 1993b; Hunter, *et al*. 1993; O'Hoy, *et al*. 1993; Shelbourne, *et al*. 1993). Interestingly, a gene conversion event was suggested in one of the cases of reverse mutation. The father transmitted his DM chromosome to his daughter, but with a number of CTG repeats within the

normal range. Haplotype analysis of six markers revealed that at least 7.2 Kb of the DM1 chromosome including the CTG repeats were exchanged with the normal chromosome (O'Hoy, *et al*. 1993). Strikingly, the reverse mutation is associated with an affected father in all reported cases. Moreover, data from small pool PCR analysis comparing somatic and sperm cells give more evidence suggesting that this process might be specific to the male germline since reversions into the normal range were only seen in sperm samples (Monckton, *et al*. 1995b).To investigate the effects of long CTG repeats in the flanking DNA, plasmids containing 17, 98 and 175 CTG repeats adjacent to the GFP gene were used to transform different strains of *E. coli*. The study revealed that long repeats promote deletions and inversions in the flanking DNA, suggesting that a possible similar mechanism could also be present in DM1 patients (Wojciechowska, *et al*. 2005).

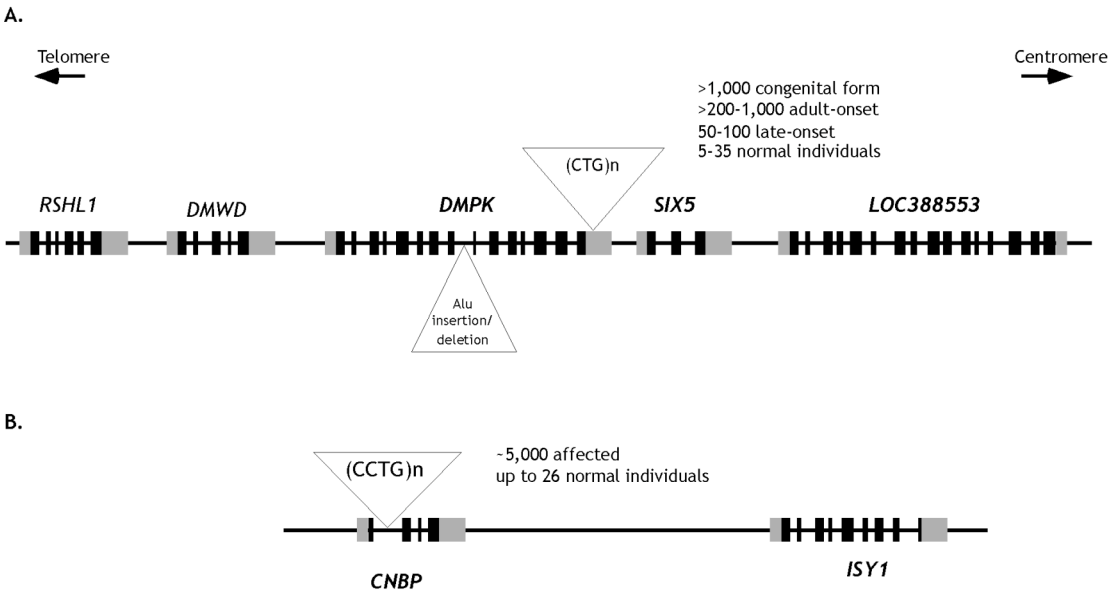


Figure 1-1. Diagram of the *DM1* and *DM2* loci. A. The gene *DMPK* is located in chromosome 19q13.3 in a region of high gene density. The CTG repeats are located in the 3'-UTR of the *DMPK*. The transcription initiation site of *SIX5* is located very closed to the polyadenilation site of *DMPK* (less than 300 bp) and the enhancer of *SIX5* is located in between (Filippova, *et al.* 2001). B. The gene *CNBP* is located in chromosome 3q.21 in not a very gene dense region. The CCTG repeats located in the first intron of *CNBP* gene are included in a complex repeat motif $(TG)_n$ $(TCTG)_n$ $(CCTG)_n$. The coding regions are indicated in black and the untranslated regions in grey (diagram based on Entrez gene database).

A significant correlation exists between the number of repeats and the severity of the clinical symptoms in lymphocytes. A negative correlation exists between the age of onset and the number of repeats. However, these correlations are not precise and it is not possible to offer good prognostic

information on the severity of the symptoms or the age at onset based on the number of repeats (Ashizawa, *et al*. 1992; Harley, *et al*. 1993; Lavedan, *et al*. 1993; Novelli, *et al*. 1993a; Redman, *et al*. 1993). By taking into account the progenitor allele and the age at sampling, it was demonstrated that better estimation of the severity and age of onset of the disease could be provided (Morales 2006).

1.2.1.3 Population genetics

The global incidence of DM1 is estimated to be 1 in 8,000 adults. However, in certain regions such as in the Sanguenay region of the province of Quebec it is as high as 1 in 475 adults, while it is extremely rare among ethnic Sub-Saharan Africans (Mathieu, *et al*. 1990; Ashizawa and Epstein 1991). A founder effect was suspected to be the cause of the high prevalence of DM in the Sanguenay region. To test the hypothesis, a genealogical reconstruction of 88 DM1 families was performed. It was demonstrated that all the patients could be traced to a unique couple, supporting a founder effect (Mathieu, *et al*. 1990). Subsequently, linkage disequilibrium between the DM1 mutation and a number of polymorphisms located in the flanking DNA at both ends of the CTG repeats was observed (Harley, *et al*. 1991). One of the polymorphisms extensively studied is the intragenic Alu repeat insertion/deletion polymorphism located in intron 8 of *DMPK* (figure 1-1). The insertion allele (Alu+) was found to be in complete association with the DM expansion in Caucasian and Japanese populations (Harley, *et al*. 1992; Mahadevan, *et al*. 1992; Yamagata, *et al*. 1992). The analysis of the structure of the Alu repeat and its presence in primates suggested that the ancestral allele is the Alu+ (Rubinsztein, *et al*. 1994). Krahe and colleagues described an exception in a Nigerian family, where the DM1 mutation was associated with the Alu- allele, suggesting that this family probably carries a *de novo* mutation (Krahe, *et al*. 1995). Neville and colleagues extended the analysis and investigated eight more polymorphic sites spanning a region of ~40 Kb. Several haplotypes with different frequencies were defined in Caucasian and Japanese populations, but the DM1 mutation is in linkage disequilibrium with the most common haplotype (A) (Neville, *et al*. 1994; Yamagata, *et al*. 1998). These results support the

hypothesis that the DM1 mutation appears to be the result of one or a few ancestral mutations.

DM1 is characterised by low reproductive fitness, affected males frequently develop testicular atrophy, adult-onset mothers generally transmit long alleles to their children and congenital or child-onset forms occurs. Therefore, expanded alleles are generally lost in just three or four generations. If there is indeed an ancestral mutation in DM1, there must exist a pool of alleles from where new expanded alleles can arise. Imbert and colleagues proposed a model to explain the generation of new CTG repeats expansion. The authors found that the distribution of the normal alleles in Caucasians was trimodal with major frequencies observed in alleles containing $(CTG)_5$, $(CTG)_{11\text{-}13}$ and $(CTG)_{19\text{-}30}$. The $(CTG)_5$ and $(CTG)_{19\text{-}30}$ were found completely associated with the Alu+ allele, whereas $(CTG)_{11\text{-}13}$ was associated with the Alu- allele. It was postulated that the repeat expanded from $(CTG)_5$ to $(CTG)_{19\text{-}30}$ by a small number of events, and that the latter represents the pool from where the DM1 mutation now arises (Imbert, *et al*. 1993). Martorrell and colleagues performed an extensively study in 700 DM1 families and provided evidence to support this model. The authors demonstrated: the trimodal distribution of normal alleles, small increases in sizes through transmissions in alleles containing $(CTG)_{19\text{-}33}$, and dramatic increases in sizes through male transmissions in alleles with $(CTG)_{37\text{-}54}$ (Martorell, *et al*. 1998). Even though there still no direct evidence showing the jump from $(CTG)_5$ to $(CTG)_{19\text{-}30}$, all the data obtained up to now supports the model of this ancestral mutation.

1.2.1.4 The DMPK protein

The *DMPK* gene consists of 15 exons and encodes a serine-threonin protein kinase. In the full length protein, five domains can be distinguished: a leucine-rich repeat amino terminal region, a catalytic domain with the kinase function, VSGGG peptide with an unknown function, a coiled coil domain and a carboxy terminal region (Brook, *et al*. 1992; Fu, *et al*. 1992; Jansen, *et al*. 1992; Mahadevan, *et al*. 1993a; Groenen, *et al*. 2000). By alternative splicing,

six principal mRNA isoforms are generated, differing in the presence or absence of the VSGGG motif and the nature of the carboxy terminus (Fu, *et al*. 1993; Groenen, *et al*. 2000). There is another minor isoform that carries a different carboxy-terminus and lacks the CUG repeats, and consequently is not retained in the nucleus in DM1 patients unlike the other isoforms (Tiscornia and Mahadevan 2000). Analysis of 16 monoclonal antibodies revealed that DMPK is expressed almost exclusively in skeletal, cardiac and smooth muscle (Lam, *et al*. 2000). The function of DMPK is not completely understood. Potential roles include sodium channel regulation in skeletal muscle (Mounsey, *et al*. 2000), membrane trafficking (Dunne, *et al*. 1996; Jin, *et al*. 2000), calcium homeostasis (Benders, *et al*. 1997) and as a regulator of the myogenic pathway (Bush, *et al*. 1996).

1.2.2 Mechanism of DM1 pathogenesis

The basis of DM1 pathogenesis is not yet fully understood. Three models were originally proposed to explain how the repeat expansion in a non-coding region could result in myotonic dystrophy. The models are: haploinsufficiency of DMPK, altered expression of neighbouring genes and a toxic gain-of-function at the RNA level.

1.2.2.1 Haploinsufficiency of DMPK

It was proposed that the expansion of CTG reduces the level of DMPK in *cis*. The levels of mRNA and protein levels were investigated in tissues and cell lines from DM1 adult-onset and congenital patients. However, the data is contradictory, since some studies detected a decrease in the levels of mRNA and protein (Carango, *et al*. 1993; Fu, *et al*. 1993; Hofmann-Radvanyi, *et al*. 1993; Novelli, *et al*. 1993b), whereas others show an increase in the mRNA levels (Sabouri, *et al*. 1993; Laurent, *et al*. 1997). Subsequently, a decrease in the levels of poly(A)+mRNA compared to the total RNA was demonstrated, indicating that hnRNA containing expanded CUG may not be adequately processed into poly(A)+mRNA (Wang, *et al*. 1995). Furthermore, it was shown that mutant transcripts are trapped in the nucleus as discrete foci at the

periphery of nuclear splicing speckles, preventing its translocation to the cytoplasm where it would normally be translated into protein resulting in haploinsufficiency of DMPK protein (Taneja, *et al*. 1995; Davis, *et al*. 1997; Hamshere, *et al*. 1997; Holt, *et al*. 2007). This nuclear retention could explain why some studies observed a decrease in the levels of mRNA and others found an increase. To investigate the biological role and the involvement of DMPK in myotonic dystrophy, transgenic mouse overexpressing *DMPK* and knockout mouse models were developed. Initially, only a partial phenotype with mild progressive skeletal myopathy and abnormalities in cardiac conduction was reproduced in young mice overexpressing DMPK (Jansen, *et al*. 1996; Reddy, *et al*. 1996; Berul, *et al*. 1999). However, it was shown subsequently that older mice (11-15 months) with 25 extracopies of the human *DMPK* gene with 11 CTG repeats developed skeletal and smooth muscle abnormalities together with cardiac defects resembling the symptoms observed in DM1 patients. An overexpression of *DMPK* is not observed in DM1 patients, and therefore the validity of the model could be questioned. However, the authors postulated that the overexpression of *DMPK* with 11 CTG repeats mimics the effects of transcripts containing expanded CTG repeats (O'Cochlain, *et al*. 2004). Despite this evidence, the haploinsuficiency of DMPK is not sufficient to explain the multisystemic phenotype of DM1.The overexpression of *DMPK* reproduced many but not all of the aspects of DM1. There are still no DM1 patients identified with point mutations in the *DMPK* gene and a second form of DM is caused by an expansion of CCTG at another locus.

1.2.2.2 Effects on neighbouring genes

The effects on neighbouring genes was suggested by the fact that the CTG repeat is a strong positioning element for nucleosomes (known to repress transcription) with the efficiency of nucleation increasing with the number of repeats (Wang, *et al*. 1994; Otten and Tapscott 1995; Wang and Griffith 1995). The expansion of repeats may alter the structure of the local chromatin, affecting the expression not only of *DMPK* but also of the contiguous genes. Adjacent to the CTG repeat at the *DM1* locus, there is a nuclease-hypersensitive site containing an enhancer element embedded

within a CpG island that regulates the transcription of the adjacent gene *sine oculis homeobox orthologue 5* (*SIX5*) (Boucher, *et al*. 1995). In addition, two CTCF binding sites flanking the CTG repeat tract were identified to form an insulator (Filippova, *et al*. 2001). In congenital cases, the CpG island is methylated (Steinbach, *et al*. 1998), the chromatin becomes more condensed, the CTCF protein cannot have access to the sites, the activity of the insulator is compromised with the consequence that gene expression is altered. It is possible that the levels of *DMPK* are increased whereas the levels of *SIX5* are decreased, explaining the more severe phenotype of congenital cases (Filippova, *et al*. 2001; Cho, *et al*. 2005).

SIX5 encodes for a protein that belongs to a family of homeobox proteins and is widely expressed in adult tissue including skeletal muscle, heart, brain and eye (Boucher, *et al*. 1995; Winchester, *et al*. 1999). Two studies demonstrated a significant reduction of the levels of *SIX5* mRNA in DM1 patients (Klesert, *et al*. 1997; Thornton, *et al*. 1997), however two other studies failed to detect any difference (Hamshere, *et al*. 1997; Eriksson, *et al*. 1999). Cataracts and cardiac conduction deficits were demonstrated in *SIX5* knockout mice (Klesert, *et al*. 2000; Sarkar, *et al*. 2000; Wakimoto, *et al*. 2002). Kirby and colleagues, from studies with *d-Six4* (the *SIX5* homologue in *Drosophila*), demonstrated that this is required for the normal development of muscle and the mesodermal component of the gonad. Consequently, deficiencies in *SIX5* may contribute toward cataracts, muscle wasting, testicular atrophy and cardiac conduction deficits in DM1 (Kirby, *et al*. 2001). Indeed, Sarkar and colleagues showed in *SIX5* knock-out mice that SIX5 is required for the viability and maturation of sperm cells and suggest that deficiencies in *SIX5* expression could cause the reproductive defects in DM1 males (Sarkar, *et al*. 2004).

The other gene located next to *DMPK* gene is *dystrophia myotonica-containing WD repeat motif* (*DMWD*), which is located upstream of *DMPK* and is mainly expressed in testis and brain (figure 1-1) (Jansen, *et al*. 1992; Jansen, *et al*. 1995). A reduction in the RNA levels of *DMWD* was reported in the cytoplasm of DM1 cell lines as well as in muscle samples from adult-onset patients

(Alwazzan, *et al*. 1999), but other studies showed no differences in the level of RNA (Alwazzan, *et al*. 1999; Eriksson, *et al*. 1999). Further studies are needed to determine if there is a clear reduction in the levels of *DMWD* and subsequently to relate the reduction with the testicular atrophy and/or abnormalities in the central nervous system observed in DM1 patients.

1.2.2.3 RNA-mediated pathogenesis

The RNA-mediated pathogenesis mechanism postulates that the expanded CUG repeats in the RNA cause the disease by titration of specific RNA CUG-binding proteins that are normally responsible for regulating the processing of multiple RNAs. Consistent with this hypothesis, it was shown that the transcripts containing expanded CUG repeats accumulate in the nucleus (Taneja, *et al*. 1995; Davis, *et al*. 1997; Holt, *et al*. 2007) and the CUG repeats fold into stable double-strand RNA hairpin structures *in vitro* (Napieraa and Krzyosiak 1997; Michalowski, *et al*. 1999), recruiting specific RNA CUG-binding proteins and the loss of function of these protein leads to myotonic dystrophy (Miller, *et al*. 2000; Fardaei, *et al*. 2001). To investigate this mechanism, a transgenic mouse model was generated by inserting into the final exon of human skeletal actin (*HSA*) gene either ~250 CTG repeats (HSA^{LR}) or ~5 repeats (HSA^{SR}). The mouse with the expanded repeats (HSA^{LR}) developed myotonia, myopathy and RNA foci characteristic of DM1, whereas those expressing the short repeats (HSA^{SR}) were normal. These results indicated that the expanded CUG repeats are toxic and sufficient to produce most of the clinical features of DM1 in skeletal muscle (Mankodi, *et al*. 2000). Similarly, Seznec and colleagues generated another transgenic mouse model (DM300) by inserting ~300 CTG repeats plus flanking DNA from DM1 patients. The mice exhibit abnormalities in skeletal muscle and abnormal tau expression in the brain consistent with symptoms observed in DM1 patients. These mice differ from the HSA^{LR} in that the expression of repeat-containing RNA is not limited to skeletal muscle and that the repeats are expressed within their natural context. These results provide further evidence for the RNA gain-of-function mechanism not only in muscle, but also in brain (Seznec, *et al*. 2001).

Two families of proteins were identified as playing a critical role in the regulation of alternative splicing. One of these families consists of CUG triplet repeat RNA-binding protein 1 (CUGBP1) and ETR3-like factors (CELFs), and the other families is the muscle-blind like (MBNL) proteins (Miller, *et al*. 2000; Fardaei, *et al*. 2001; Ladd, *et al*. 2001).

The CELF family is widely expressed in adult tissues with the highest levels in skeletal muscle, heart and brain; six genes have been identified in humans (Timchenko, *et al*. 1996b; Lu, *et al*. 1999; Good, *et al*. 2000; Ladd, *et al*. 2001). CUGBP1 was initially characterised by electrophoretic mobility bandshift assay interacting with single stranded RNA containing CUG repeats (Timchenko, *et al*. 1996b). Subsequently, an increase in the levels of CUGBP1 in the nucleus was observed in lymphoblast cells lines of DM1 patients and in homozygous *DMPK* knockout mice (Timchenko, *et al*. 1996a; Roberts, *et al*. 1997). Similarly, an increased expression of ETR-3 in skeletal and cardiac muscle was detected, suggesting that the function could be altered leading to defects in alternative splicing (Savkur, *et al*. 2001; Timchenko, *et al*. 2001). It was shown that the levels of CELF proteins increase, but the proteins do not co-localise with the expanded transcripts.

The other family of RNA-binding proteins involved in DM1 pathogenesis is MBNL. Three proteins were identified in humans, namely MBNL1, MBNL2 and MBNL3, which all interact *in vitro* with CUG repeats and co-localise with the nuclear foci of DM1 and DM2 patients (Miller, *et al*. 2000; Fardaei, *et al*. 2002; Mankodi, *et al*. 2002; Kanadia, *et al*. 2003). The loss of MBNL protein was postulated to play a role in DM1 pathogenesis. To investigate this hypothesis, an *Mbnl1* knockout model was developed. The mice showed defects similar to those seen in DM1 such as myotonia, ocular subcapsular cataracts and misregulation of splicing in a number of pre-mRNAs (Kanadia, *et al*. 2003). Furthermore, it was demonstrated that the overexpression of *Mbnl1* was sufficient to rescue the myotonia and misregulation of splicing in skeletal muscle in HSA^{LR} mice model (Kanadia, *et al*. 2006).

CELF and MBNL proteins might function antagonistically regulating the alternative splicing of specific mRNAs. CELF acts during development promoting the inclusion of fetal exons, whereas MBNL promotes fetal exon skipping and expression of adults isoforms. In DM1 patients, the decrease of MBNL, and the increase of CELF cause a misregulation of the alternative splicing with the result that fetal isoforms are present in adult tissues (Ho, *et al*. 2004). In DM1, an increasing number of pre-mRNA are shown to present defects in alternative splicing including *cardiac troponin T* (*TNNT2*) (Philips, *et al*. 1998), *insulin receptor* (*IR*) (Savkur, *et al*. 2001), *muscle-specific chloride channel* (*CLCN1*) (Charlet, *et al*. 2002; Mankodi, *et al*. 2002), *brain microtubule-associated tau* (Sergeant, *et al*. 2001) and *mytoubularin-related 1* (*MTMR1*) (Buj-Bello, *et al*. 2002). Altered expression of one of the isoforms of *IR* has been observed in skeletal muscles from DM1 patients providing an explanation for the insulin resistance observed among them (Savkur, *et al*. 2001). The alternative splicing of *CLCN1* was investigated in skeletal muscle from transgenic DM mouse models (HSA^{LR}) and in muscle tissue from DM1 patients. A loss of function of CIC-1 due to splicing defects and the development of myotonia was correlated with the level of expanded CUG repeats (Charlet, *et al*. 2002; Mankodi, *et al*. 2002). Alternatively, to investigate if the increase in the levels of CUGBP could reproduce the defects in alternative splicing observed in DM1, a transgenic mouse-overexpressing CUGBP1 in skeletal muscle and heart was generated. The mice developed skeletal muscle pathological features similar to DM1 tissues and defects in the alternative splicing of at least three pre-mRNAs: *CLCN1*, *TNNT2* and *MTMR1* (Ho, *et al*. 2005).

1.2.3 Myotonic dystrophy type 2

1.2.3.1 Clinical manifestations

Myotonic dystrophy type 2 was previously known as proximal myotonic myopathy (PROMM) and is characterised by: proximal muscle weakness, muscle pain and stiffness, cataracts, myotonia, cardiac complications and endocrinological abnormalities. Mild ptosis is present with combined mild

facial weakness in a number of patients. The symptoms are generally milder than in DM1 and there are no congenital cases reported. The clinical expression of the disease is highly variable even within families (Day, *et al*. 2003).

1.2.3.2 Genetic features

In 1998, a DM family that did not carry the DM1 mutation was shown by linkage analysis to map to chromosome 3q.21 (Ranum, *et al*. 1998). It was not until 2001 that the mutation was found to be a CCTG repeat expansion located in the first intron of the *CNBP* gene (Liquori, *et al*. 2001). The CCTG repeat is part of a complex repeat motif $(TG)_n$ $(TCTG)_n$ $(CCTG)_n$ (figure 1-1). Normal individuals contain up to 26 CCTG that are generally interrupted, whereas patients contain uninterrupted expanded alleles with a mean of ~5,000 repeats (Liquori, *et al*. 2001). In contrast to DM1, genetic anticipation is not clearly demonstrated in DM2 families. A decrease in the age of onset through generations is observed, but an increase in the numbers of repeats have not been demonstrated due to the high levels of somatic mosaicism that complicated the analysis (Day, *et al*. 2003).

1.2.3.3 Mechanism of pathogenesis

The parallels in the clinical and molecular aspects of DM1 and DM2 led to the postulation of an RNA gain-of-function model to explain the pathology of DM2 (Ranum and Cooper 2006). Using patient derived myoblast cells lines and skeletal muscle biopsies, it was shown that mutant *CNBP* pre-mRNA is spliced normally, the mRNA is exported from the nucleus and the protein is produced normally in the cytoplasm. Nevertheless, the CCUG repeats are retained in the nucleus similar to the DM1 repeats indicating that the CCUG expansions might cause DM2 (Margolis, *et al*. 2006). The content and location of the DM2 foci are different from DM1 foci. DM2 foci are widely dispersed in the nucleoplasm and contain just the CCUG repeats, whereas DM1 foci accumulate at the periphery of nuclear speckles and contain mutant processed *DMPK* mRNA (Holt, *et al*. 2007). The RNA gain-of-function mechanism seems to explain the similar clinical features of DM1 and DM2,

however there are some differences between both diseases that cannot be explained. DM2 is generally milder and a congenital form is not detected and the weakness is mainly proximal. It was postulated that these differences could be explained by alternative mechanisms such as effects on neighbouring genes and/or haploinsufficiency of DMPK. Further comparisons between DM1 and DM2 will help to understand the complex mechanism of pathogenesis (Day and Ranum 2005).

1.3 Charcot-Marie-Tooth (CMT) disease

In 1886, Charcot and Marie in France simultaneously with Tooth in England described the clinical features of a hereditary motor and sensory neuropathy (HMSN) disorder that was later referred as Charcot-Marie-Tooth disease. CMT is a group of genetically heterogeneous, inherited neuropathies affecting motor and sensory nerves of the peripheral nervous system (Shy 2004). CMT is the most common inherited neuromuscular disease with an estimated prevalence of 1 per 2,500 individuals (Skre 1974).

1.3.1 Clinical manifestations

CMT is characterised by progressive distal muscle wasting and weakness starting in the lower limb and later affecting also the upper extremities, distal sensory loss, decrease or absence of tendon reflexes, and foot deformities such as *pes cavus* (Harding and Thomas 1980). In addition, hearing loss is observed in 5% of patients with CMT (Pareyson, *et al*. 1995). The clinical expression of CMT is highly variable, not only between families but also within families and only occasionally results in severe impairment (Pareyson 1999).

CMT is classified into two main groups according to the nerve conduction velocity (NCV). CMT1 (the demyelinating type) is characterised by motor median nerve conduction velocity less than 38 m/s and CMT2 (the axonal type) with normal or slightly reduced nerve conduction velocities (≥ 45 m/s). Subsequently, an intermediate CMT with a range of 25-45 m/s was also

defined (Davis, *et al*. 1978; Harding and Thomas 1980). Further subdivisions depend on the inheritance pattern, which can be: autosomal dominant, autosomal recessive or X-linked. The classification is not straightforward, because different mutations in the same gene cause different phenotypes (Suter and Scherer 2003).

1.3.2 Genetic features and mechanism of pathogenesis

In 1982, Bird and colleagues showed by linkage analysis that the mutation in some families with CMT1 mapped to chromosome 1 (Bird, *et al*. 1982). Subsequently, studies on DNA polymorphism showed that most patients showed linkage to chromosome 17 (Vance, *et al*. 1989). These results provided the first evidence of locus heterogeneity (mutations in different genes causing a similar phenotype) for CMT. Furthermore, allelic heterogeneity (mutations in the same gene causing a different phenotype) was also demonstrated (Suter and Scherer 2003). To date, mutations causing CMT have been mapped to 30 different loci and more than 20 different genes have been identified. However there are still patients with an unknown aetiology, indicating that there are still a number of mutations remaining to be identified (http://www.molgen.ua.ac.be/CMTMutations/default.cfm).

The majority of CMT patients (40%-50%) carry a 1.5 Mb duplication on chromosome 17p11.2-p12, including the *peripheral myelin protein 22* (*PMP22*) gene (Raeymaekers, *et al*. 1989; Vance, *et al*. 1989; Matsunami, *et al*. 1992; Timmerman, *et al*. 1992; Valentijn, *et al*. 1992). PMP22 is a small membrane protein involved in the correct development of peripheral nerves, determination of myelin thickness, and maintenance of axons (Adlkofer, *et al*. 1995). Other genes mutated in CMT patients include: *myelin protein zero* (*MPZ* or *PO*), *gap junction binding 1* (*GJB1*), *glycil-tRNA synthetase* (*GARS*), *lamin A/C nuclear envelope protein* (*LMNA*) and *dynamin2* (*DNM2*) (Suter and Scherer 2003; Pareyson 2004; Barisic, *et al*. 2008). MPZ is the major peripheral myelin protein and mutations may affect Schwann cell-neuron interactions causing defective myelination of peripheral nerves (Martini and Schachner 1997). *GJB1* encodes for the connexin 32 protein, a gap junction

protein localized at the nodes of ranvier and involved in transportation within the cell. GARS is a housekeeping gene encoding an aminoacyl tRNA synthetase that plays a role in translation. Mutations in GARS might affect protein synthesis necessary for motor neuron maintenance and integrity (Antonellis, *et al*. 2003). Lamins are structural proteins of the nuclear lamina and mutations can cause defective nuclear mechanics and impaired mechanically activated gene transcription (Lammerding, *et al*. 2004). DNM2 is implicated in vesicular trafficking and endocytosis (McNiven 1998). In summary, the mechanisms of pathogenesis identified to date include: Schwann cell function, disruption of axonal transport or intracellular trafficking, heat shock proteins and ubiquitous housekeeping enzymes.

1.4 DM1+CMT++ family

The DM1+CMT++ family is the first reported DM1 family in which all affected members also have an intermediate type CMT neuropathy. Co-incidence of DM1 and CMT has been observed in a number of patients (Rogoff and Ziegler 1956; Caccia, *et al*. 1972; Panayiotopoulos and Scarpalezos 1976; Borenstein, *et al*. 1977; Cros, *et al*. 1988; von Giesen, *et al*. 1994), but CMT is not a recognized feature of DM1 (Harper 2001). The proband in the DM1+CMT++ family was a 65-year-old woman (II-25, figure 1-2) presenting features of the classical adult form of DM1 such as characteristic myopathic fascies, distal muscle weakness, percussion myotonia and myotonic discharges on EMG. In addition, she presented features of CMT such as reduced motor conduction velocity (~60% of normal) and slight sensory loss in the lower limbs. Subsequently, her eldest son (III-7) was investigated and he also presented symptoms of the classical adult form of DM1 including frontal balding and testicular atrophy, together with features of CMT including *pes cavus*, increased lumbar lordosis and slight distal sensory loss in the legs. Furthermore, signs of axonal degeneration and regeneration, demyelination and remyelination, and many small onion bulbs were detected in a sural nerve biopsy (Spaans, *et al*. 1986).

A family study was undertaken and 12 more cases were identified (detailed information of the DM1 and CMT symptoms are shown in table 1-2 and 1-3 respectively). Previous generations were also investigated but it seems that the disorders have occurred only among the descendants of I-6 and I-9. Both I-6 and I-9 were described by their offspring as having no motor disability and lived apparently healthy lives until 79 and 85 years of age respectively (Spaans, *et al*. 1986). Taking into account the results of nerve biopsy studies in III-7 and III-9 and the values of motor conduction velocity being in the range of 25 to 45 m/s, the disease has been classified as an intermediate type CMT neuropathy (Spaans, *et al*. 2008).

Anticipation is observed in classic DM1 families, however anticipation in the DM1+CMT++ family, was not completely detected. Both individuals in the first generation were asymptomatic, but one of them probably transmitted the mutant allele to five of their daughters who developed late-onset DM1. A decrease in the age of onset and an increase in the severity of DM1 symptoms were observed between the second and third generation. Nevertheless, no congenital cases were detected in the last generation, despite four adult-onset affected maternal transmissions (Spaans, *et al*. 1986; Brunner, *et al*. 1991). All patients in the fourth generation are intelligent, energetic, obtained university degrees and are employed in highly skilled jobs. The first generation has not been investigated, however one of them probably carried the DM1+CMT++ allele but with the number of CTG at the 5'-end not enough to be pathogenic. Unexpectedly for CMT, an increase in the severity of symptoms and decrease in the age of onset was observed between the third and fourth generation (Spaans, *et al*. 2008).

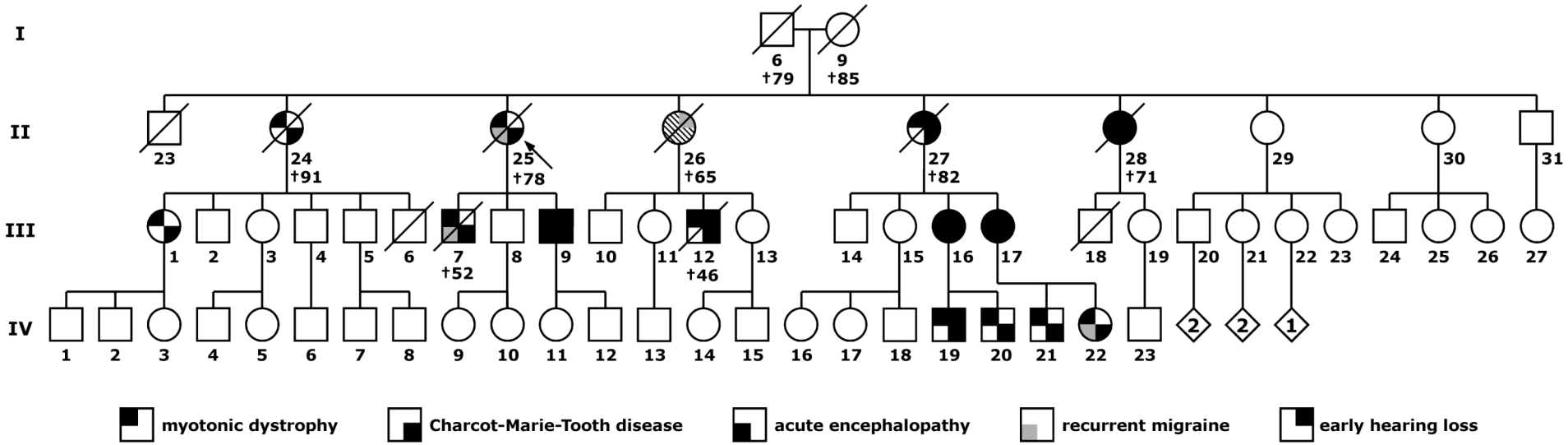


Figure 1-2. Pedigree of the DM1+CMT++ family. The symptoms present in the affected members appeared only among the descendants of I-6 and I-9, both were described by their offspring as having no motor disability and lived apparently healthy lives until 79 and 85 years of age respectively. I-6 worked as butcher and farmer until he died. The individual II-26 refused to be examined but she must be an obligatory carrier. The age of the deceased is indicated (diagram obtained from Monckton, unpublished data).

Table 1-2. Summary of DM1 symptoms

	Cataract	Myotonic discharges	Weakness	Muscular atrophy	Percussion myotonia	Cardiac involvement	Age of Onset (years)
II-24	+	+	NN	-	-	-	NN
II-25	+	++	+	+	+	-	50
II-27	+	++	-	-	-	-	NN
II-28	+	+++	+	+	++	-	55
III-1	±	-	-	-	-	-	NN
III-7	+	+++	++	++	+	-	28
III-9	+	+++	+	+	-	+	25
III-12	+	+++	++	++	+	-	25
III-16	+	-	+	+	-	-	44
III-17	+	+	-	-	-	+	35
IV-19	-	+	+++	+++	+	-	20
IV-20	-	-	+	+	+	-	24
IV-21	-	-	+	±	±	+	17
IV-22	+	-	-	-	±	-	24

NN (not known), + (present), - (absent), ± (moderate).

Table 1-3. Summary of CMT symptoms.

	Distal sensory disturbances	Absent knee jerk	Absent ankle jerk	MNCV (m/s)	SNCV (m/s)	*Pes cavus*	Age of Onset (years)
II-24	+	+	+	40	NR	-	ND
II-25	±	-	+	38	30	-	ND
II-27	+	-	-	45	NR	-	ND
II-28	±	+	+	31	NR	-	ND
III-1	±	-	-	48	NR	-	ND
III-7	+	±	+	27	NR	+	28
III-9	+	±	+	39	NR	+	25
III-12	+	±	+	42	NR	-	25
III-16	+	-	-	44	NR	±	32
III-17	-	-	-	33	NR	-	35
IV-19	++	-	-	39	NR	++	10
IV-20	+	+	+	23	NR	±	17
IV-21	-	-	+	27	NR	-	12
IV-22	-	-	+	44	NR	-	4

MNCV (motor conduction velocity) normal value >40-45 m/s, SNCV (sensory nerve conduction velocity) normal value ≥42m/s. NR (no response with distal nerve stimulation), + (present), - (absent), ± (moderate), ND (not determined).

The mode of inheritance of DM1 and CMT was classified as autosomal dominant and complete co-segregation was shown by linkage analyses. Both diseases co-segregated with the *ApoC2* locus with a maximum LOD score of 7.04 with zero recombination (Brunner et al. 1991). The molecular

diagnosis of DM1 was performed in II-27, III-9, III-16, III-17, IV-19, IV-20, IV-21 and IV-20 by Southern blot analysis of restriction digested genomic DNA by PCR and by repeat primed-PCR (RP-PCR). Small expansions (~200-400 CTG repeats) were detected by Southern blot, but the expanded allele could not be amplified either by PCR or RP-PCR. Mutations in genes most frequently associated with CMT (*PMP22* and in *GJB1*) were investigated (Shy 2004). III-7 was screened for a duplication in *PMP22* and IV-21 for mutations in *GJB1*, but no abnormalities were detected. Standard cytogenetic studies (GTG-banded analysis) and FISH analysis with probes specific for band 19q13 and the subtelomerics regions of chromosome 19 revealed no abnormalities in III-9 and III-16 (Spaans, *et al*. 2008). As a result, a single or two or more closely linked mutations were suggested to be the cause of DM1 and CMT.

A number of patients (II-25, II-26, II-28, III-7, III-9, III-16, III-17 and IV-22) were identified as suffering from migraine-like headache attacks. In addition, II-28, III-9, III-16 and III-17 from about 50 years of age, developed several bouts of acute encephalopathy (figure 1-2). The symptoms observed included headache, confusion, fever, encephalographic abnormalities and impaired consciousness. II-28 and III-9 developed coma on more than one occasion. Disorders presenting similar symptoms include familial hemiplegic migraine type 1 (FHM1) and type 2 (FHM2), and cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). FHM1 can be caused by mutations in the *calcium channel* gene (*CACNA1A*) (Ophoff, *et al*. 1996), FHM2 by mutations in a gene enconding the *alpha-2 subunit of the sodium/potassium pump* (*ATP1A2*) (De Fusco, *et al*. 2003) and CADASIL by mutations in *NOTCH3* (Joutel, *et al*. 1996). Screening of these genes in III-9 did not reveal any abnormalities (Spaans, *et al*. 2008).

Hearing loss at a relatively young age was also observed in most of the patients. Apparently, II-26, II-27, II-28 and III-12 developed hearing loss before 45 years old as reported by their relatives. Audiologic data of III-9, III-16, III-17 and IV-19 revealed severe to profound bilateral sensorineural hearing loss in excess for their age (Spaans, *et al*. 2008). Mutations in more

than 20 different genes have been associated with autosomal dominant impaired hearing (Finsterer and Fellinger 2005).

The complex symptomatology observed in the DM1+CMT++ family has been extensively studied, however the molecular lesion still remains to be defined. Additional studies at the molecular level will be required to delineate the nature of the mutation or mutations responsible for DM1, CMT, acute encephalopathy and hearing loss.

1.5 Aims of the project

The molecular diagnosis of myotonic dystrophy type 1, in most of the cases, is a straightforward procedure where an expansion of $(CTG)_n$ in the *DM1* locus is detected by Southern blot of restriction digested genomic DNA, PCR and/or RP-PCR. In the DM1+CMT++ family an expanded allele was detected by Southern blot analysis, but PCR and RP-PCR failed to detect an expanded allele. It was postulated that an additional lesion might be responsible of the complex symptomatology. Taking into account the evidence obtained in bacteria carrying long CTG repeats. It was postulated that the expanded repeat at the *DMPK* locus may also predispose the repeat tract and flanking regions to further instability, generating a putative break where deletions and/or rearrangements might occur. These novel mutations might not be identified with the standard techniques and could modify the expression of *DMPK* and/or other genes, explaining the symptomatology observed in the DM1+CMT++ family. Additionally, there are a number of DM1 sporadic patients with an unusual molecular diagnosis, which might also present these novel mutations.

A series of molecular approaches will be carried out to investigate the nature of the molecular lesion responsible for the symptomatology observed in the DM1+CMT++ family and in a number of other sporadic DM1 cases with unusual molecular diagnosis. It is hoped that the definition of the mutation in this family will shed light on the pathogenic pathways not only in this family but also in classic DM1. A comparison between the similarities and/or differences of the findings with the DM1+CMT++ family

and the sporadic DM1 patients will be necessary to design novel molecular diagnostic tools for myotonic dystrophy patients and also for patients with a clinical diagnosis of Charcot-Marie-Tooth disease with an unknown aetiology.

2 Materials & Methods

2.1 Materials

Chemicals, molecular biology reagents and plastic ware were obtained from standard suppliers such as: Fisher Scientific, Invitrogen, Merck Ltd., Promega, Sigma, unless stated otherwise. Oligonucleotides were obtained from Sigma-Genosys.

2.2 Techniques

All kits and reagents were used following the manufacturer's instructions unless stated otherwise.

2.2.1 DNA extractions

DNA was isolated from various sources: Nucleon kit (Amersham Biosciences) to extract DNA from blood samples; QIAquick PCR Purification Kit and QIAquick Gel Extraction Kit (Qiagen) to purify DNA from PCRs and from agarose gel slices, respectively; and QIAprep spin miniprep kit and endofree plasmid maxi kits (Qiagen) to isolate plasmid DNA. The concentration of DNA was determined with a Nanodrop ND-100 Spectrophotometer.

2.2.2 PCR

2.2.2.1 Standard reaction

Genomic DNA (20 ng) was amplified in a 10 μl reaction, with 1 μM of each primer, 1U of *Taq* DNA polymerase (Bioline) and 1X PCR buffer (45 mM Tris•HCl pH 8.8, 11 mM ammonium sulphate, 4.5 mM $MgCl_2$, 0.048% 2-mercaptoethanol, 4.4 μM EDTA, 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 1 mM dTTP, 0.113 mg/ml BSA). 10% DMSO was added when it was required. The reactions were performed in a thermal cycler (Biometra) cycled through 28-

30 rounds of (96°C for 45 sec, annealing temperature (AT) for 45 sec, 70°C for 3 min), AT for 1 min and 70°C for 10 min. The sequences of the primers and the annealing temperatures are listed in table 2-1 and table 2-2.

Table 2-1. Forward primers within the *DM1* locus

Name	5'-3' Sequence	Target sequence	AT_1/AT_2 (°C)
DM-T*	AGTCCCAGGAGCCAATCAGAG	5'-end of CTG repeat tract	68
DM-H	TCTCCGCCCAGCTCCAGTCC	5'-end of CTG repeat tract	68/63.5
DM-A	CAGTTCACAACCGCTCCGAGC	5'-end of CTG repeat tract	68/63.5
DM-C	AACGGGGCTCGAAGGGTCCT	5'-end of CTG repeat tract	68/63.5
DM-Cc	GGCTCGAAGGGTCCTTGTAG	5'-end of CTG repeat tract	68
BAB452	CTGCTGCTGCTGCTGGG	within the repeat tract	66
DM-GGC	TGCTGCTGCTGCTGGGCGGCG	repeat in DM1+CMT++ family	66
DM-M	CACGTTTTGGATGCACTGAGAC	3'-end of CTG repeat tract	68
DMEcoF*	ATCTCGTCTGGCAGCAGTTT	*Eco*RI site in *LOC38553* gene	64.5

* Primers were designed using the Primer3 website (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). AT_1 (without DMSO) AT_2 (in presence of 10% DMSO)

Table 2-2. Reverse primers within the *DM1* locus

Name	5'-3' Sequence	Target sequence	AT_1/AT_2 (°C)
DM-CR	AGGACCCTTCGAGCCCCGTTC	5-end of CTG repeat tract	68
DM-ER	AAATGGTCTGTGATCCCCCCA	3-end of CTG repeat tract	68
DM-DR	CAGGCCTGCAGTTTGCCCATC	3-end of CTG repeat tract	68/63.5
DM-BR	CGTGGAGGATGGAACACGGAC	3-end of CTG repeat tract	68/63.5
DM-GR	GCAGGGCGTCATGCACAAGAAA	3-end of CTG repeat tract	68
DM-SR	TCCCCGAAAAAGCGGGTTTGG	3-end of CTG repeat tract	64
DM-MR	GTCTCAGTGCATCCAAAACGTG	3-end of CTG repeat tract	68
DM-QR	CACTGTGGAGTCCAGAGCTTTG	3-end of CTG repeat tract	68/63
3PBR	GGAACGGGCTAGAAAGTTTGCAGC	3-end of CTG repeat tract	68
3PCR	CCATCGGGACAACGCAGAAGGTAAC	3-end of CTG repeat tract	68
3PDR	CAGCGTGTGCTTCTGGCTCAG	3-end of CTG repeat tract	68
SIX-ER*	CAGGCAAGGTAGCCATGTTT	Exon 1 *SIX5* gene	61
SIX-IR*	AGCGCTCCTTGAAGCAGTAG	Exon 1 *SIX5* gene	64.4
DMEcoR*	GCCTATGGGACTTTGCACAT	*Eco*RI site in LOC38553 gene	64.5

* Primers were designed using the Primer3 website (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). AT_1 (without DMSO) AT_2 (in presence of 10% DMSO)

2.2.2.2 Repeat Primed-PCR (RP-PCR)

RP-PCR was performed similarly to standard PCR for 28 cycles with the addition of three primers: TAG, a specific flanking primer and the primer

with a target sequence within a repeat tract coupled with a 5'-end tail (TAG) used at a final concentration of 0.1 μM. The sequences of the primers and the annealing temperatures are listed in table 2-3.

Table 2-3. Primers selected for RP-PCR

Name	Sequence (5'-3')	Target sequence	AT_1/AT_2 (°C)
TAG	TCATGCGTCCATGGTCCGGA	none	68/65 or 67
TAG-GTC	TCATGCGTCCATGGTCCGGATGCTGCTGCTGCTGC	$(CTG)_{\geq 5}$	68/65
TAG-AGC	TCATGCGTCCATGGTCCGGAAGCAGCAGCAGCAGCAGC	$(CTG)_{\geq 5}$	68/65
TAG-CCGCTG	TCATGCGTCCATGGTCCGGACCGCTGCCGCTGCCGCTG	$(CCGCTG)_{\geq 3}$	68/67
TAG-CAGCGG	TCATGCGTCCATGGTCCGGACAGCGGCAGCGGCAGCGG	$(CCGCTG)_{\geq 3}$	68/67
TAG-CCTGCTG	TCATGCGTCCATGGTCCGGACCTGCTGCCTGCTGCCTG	$(CCTGCTG)_{\geq 3}$	68
TAG-GCC	TCATGCGTCCATGGTCCGGACCGCCGCCGCCGCCGCCG	$(CCG)_{\geq 5}$	68/65
TAG-GGC	TCATGCGTCCATGGTCCGGAGCGGCGGCGGCGGCGGCG	$(CCG)_{\geq 5}$	68/65

AT_1 (without DMSO) AT_2 (in presence of 10% DMSO)

2.2.2.3 Small pool PCR (SP-PCR)

To investigate the dynamics of the DM1 alleles SP-PCR was performed as previously described (Monckton, *et al*. 1995b). The DNA samples were digested with *Hin*dIII and serially diluted in (1xTE and 0.1 μM of forward primer). A standard PCR was carried out for 28 cycles in the presence of 10% DMSO with primers DM-A and DM-DR. In the absence of DMSO the following modification was carried out: 1 μl of DNA was amplified in a final volume of 7 μl containing 0.2 μM of each primer (DM-C and DM-DR), 0.35 U of Taq DNA polymerase (SIGMA) in 1X PCR buffer.

To investigate the degree of somatic mosaicism at least 100 single molecules were sized for each sample. To determine the number of input molecules two procedures were followed by amplifying a low concentration of DNA. If the degree of variation was very high, the different alleles were counted directly, assuming that each band arose from a single allele. However, if the degree of variation was low and single alleles were not resolved the average number of input molecules was estimated according to the Poisson distribution. The ratio of positive to negative reactions (f(0))

was calculated and the average number of input molecules could be estimated by $m = -\ln\left(f(0)\right)$(Gomes-Pereira, *et al*. 2004a).

$$f(0) = \frac{\text{(total number of reactions - number of reactions yielding a product)}}{\text{total number of reactions}}$$

2.2.2.4 Vectorette PCR

Genomic DNA (1 μg) was digested overnight with 5U of the restriction enzyme of choice in the presence of 1 mM of spermidine in a final volume of 20 μl. The enzyme was heat inactivated for 20 min either at 65°C or 80°C. The doublestranded linker oligo was annealed as follows: 2.5 μg of two oligos of choice (table 1-5), 167 mM NaCl, 17 mM $MgCl_2$, 17 mM Tris-HCl in 20 μl were placed in a tube and placed in a beaker of boiling water, removed and cooled slowly to room temperature. The choice of linker oligos depended on the enzyme used to digest the genomic DNA (detailed in table 2-5). The digested DNA (1 μg) and 2 μg of linker were ligated at 16°C overnight in the presence of 200 U of T4 DNA ligase and 1X T4 ligase buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 1mM ATP and 25 μg/ml BSA) in a total volume of 100 μl. A control ligation was performed by removing 10 μl of the reaction and adding 500 ng of *Hin*dIII digested λDNA. The control ligation was resolved on an agarose gel to check the ligation efficiency. Excess linker was removed by incubating the ligation reaction at 95 °C and performing an ammonium acetate precipitation immediately. The remaining ligation reaction (90 μl) were mixed with 1/3 volume of 10 M NH_4 acetate (pH 7) and 1 1/3 volume of isopropanol. The DNA was collected by centrifugation in a bench top microcentrifuge (10 min at 13,000 RPM). The supernatant was decanted off, the pellet was rinsed with 1ml of 80% ethanol and the DNA pellet was recovered as above. The DNA pellet was air dried in a laminar flow hood for ~20 min and dissolved in 100 μl of water. A volume of 1 μl of each vectorette library was subjected to 30-cycles of PCR with a specific primer located in the flanking DNA and the vectorette TAG primer (the sequences of the TAG primers are listed in table 2-4). A second

round of 25-cycles hemi-nested or nested-PCR was performed using a 200-fold dilution of the primary PCR.

2.2.3 Restriction Digestion

Restriction digests were performed following the supplier's recommendation unless otherwise stated; the enzymes were obtained either from Promega or New England Biolabs.

2.2.3.1 Restriction digest of genomic DNA

Genomic DNA (5 μg) was digested with 5 U of either *Eco*RI or *Pst*I in 1X recommended buffer and 100 μg/ml of BSA in a final volume of 50 μl. Reactions were incubated overnight at 37°C. The *Eco*RI-digested products were resolved on a 0.7% agarose gel and the *Pst*I-digested products on a 1% agarose gel.

2.2.3.2 Restriction mapping to characterise interruptions within the CTG repeat tract in the *DM1+CMT++* mutant allele

A 10-fold dilution of the PCR product generated with flanking primers was digested with 5 U of *Fnu*4HI, *Aci*I, *Hha*I or *Msp*A1I in a final volume of 20 μl at 37°C for 1 hour. The products were resolved on a 1.5% agarose gel and detected by Southern blot hybridisation either with the DM1 CTG repeat probe or the 5'-end probe.

Table 2-4. Vectorette primers

Name	Sequence (5'-3')	AT (°C)
TAG2	TGGTCATCCGGAGGTGTGAG	68
TAG3	TCCGGAGGTGTGAGCTGATT	68
TAG4	TCATGCGTCCATGAGTGGCCGGAT	68

Table 2-5. Oligos used to generate linkers

Oligo	Sequence (5'-3')	Oligo	Sequence (5'-3')	Enzyme
V-SU	GATCGGACTTGCTACGGTAATCAG	V-NL-TAG2	TGGTCATCCGGAGGTGTGAGCTGATTACCGTAGCAAGTCC	*Bam*HI, *Sau*3A1
V-CG	CGGGACTTGCTACGGTAATCAG	V-NL-TAG2	TGGTCATCCGGAGGTGTGAGCTGATTACCGTAGCAAGTCC	*Taq*I *Aci*I
V-AATT	AATTGGACTTGCTACGGTAATCAG	V-NL-TAG2	TGGTCATCCGGAGGTGTGAGCTGATTACCGTAGCAAGTCC	*Eco*RI
V	GGACTTGCTACGGTAATCAG	V-NL-TAG4-ACGT	TCATGCGTCCATGAGTGGCCGGATCTGATTACCGTAGCAAGTCCTGCA	*Pst*I
V	GGACTTGCTACGGTAATCAG	V-NL-TAG4-GC	TCATGCGTCCATGAGTGGCCGGATCTGATTACCGTAGCAAGTCCCG	*Hha*I
V-TA	TAGGACTTGCTACGGTAATCAG	V-NL-TAG4	TCATGCGTCCATGAGTGGCCGGATCTGATTACCGTAGCAAGTCC	*Mse*I
V-CG	CGGGACTTGCTACGGTAATCAG	V-NL-TAG4	TCATGCGTCCATGAGTGGCCGGATCTGATTACCGTAGCAAGTCC	*Taq*I *Aci*I
V-AATT	AATTGGACTTGCTACGGTAATCAG	V-NL-TAG4	TCATGCGTCCATGAGTGGCCGGATCTGATTACCGTAGCAAGTCC	*Eco*RI
V-ANT	ANTGGACTTGCTACGGTAATCAG	V-NL-TAG4	TCATGCGTCCATGAGTGGCCGGATCTGATTACCGTAGCAAGTCC	*Hin*fI
V-SU	GATCGGACTTGCTACGGTAATCAG	V-NL-TAG4	TCATGCGTCCATGAGTGGCCGGATCTGATTACCGTAGCAAGTCC	*Bam*HI, *Sau*3A1

The underlined base indicates the overhand ends generated with the enzymes.

2.2.4 Electrophoresis

2.2.4.1 Agarose gels

Depending on the size of the products to be resolved different percentages of agarose gels were used in 0.5 X TBE buffer (45 mM Tris, 45mM boric acid, 1 mM EDTA pH8) with 0.5 μM of ethidium bromide and run at a constant voltage (100-130V).

For the resolution of small fragments with small size differences a 2% (w/v) Nusieve, 1% (w/v) agarose gels were used. Before loading the samples in the gel 5X Orange G loading dye (0.2 % (w/v) Orange G 15 % (w/v) Ficoll, 1X TBE) was added. The DNA fragments were visualised using a UV transilluminator and photographed using the UVP ImageStore 7500 system. The bands were sized by comparison to molecular weight markers using Kodak Digital Science 1D 3.5.4 software (Kodak). The molecular weight markers used were 1 kb+ ladder (60 ng/μl 1 kb ladder, 1X DNA loading dye in 1X TBE) and λ*Hin*dIII (10 ng/μl λ*Hin*dIII, 1X orange G loading dye).

2.2.5 Southern "squash" blot

Southern "squash" blots were performed to transfer DNA from gels to nylon membranes. Gels were rinsed in distilled water, incubated with gentle shaking in the following solutions: 10 min in depurinating solution (0.25 M HCl), 30 minutes in denaturing solution (0.5 M NaOH, 1.5 M NaCl) and 30 min in neutralising solution (1.5 M NaCl, 0.5M Tris pH 7.5). The gels were rinsed in distilled water between incubations. Cling film was placed on the bench followed by two sheets of gel blotting paper, the gel inverted, the membrane (Hybond-N form GE Healthcare), two sheets of gel blotting paper, a thick layer of paper towels, a glass and a weight (~1 kg). The membrane was wet first in distilled water and then in neutralising solution; the sheets of gel blotting paper were wet in neutralising solution. Any bubbles trapped in each layer were carefully removed. The DNA was transferred to the membrane

overnight by capillary action. The DNA was fixed to the membrane by 20 min baking at 80°C followed by UV crosslinking at 1,200 J/m^2 with a Stratagene UV crosslinker 2400.

2.2.6 Preparation of DNA probes

Three different probes were used: the DM1 CTG repeat probe, a 5'-end probe and p5B1.4, which is a 1.4 Kb *Bam*HI-digested product of cDNA25 (Buxton*, et al*. 1992). The DM1 CTG repeat probe was generated with primers DM-A and DM-DR using a template containing 56 CTG repeats. The 5'-end probe was generated with primers DM-H and DM-CR, generating a 99 bp product that will hybridise at the 5'-end of the repeat tract. The DNA was purified by precipitation, PCR products were mixed with 1 volume of 5M NH_4 acetate (pH 7) and 1 1/2 volumes of isopropanol. The DNA was collected by centrifugation in a bench top centrifuge (15 min at 13,000 RPM). The pellet was washed with 1ml 80% ETOH, recovered again as above, air dried and dissolved in water. The concentration of DNA was quantified and 10 ng/μl aliquots were made.

Purified DNA (20-30 ng) and the DNA marker (2.5 ng) were radiolabelled with 1.85 MBq [α-^{32}P]dCTP (3,000 Ci/nmol) using Ready-to-go DNA labelling beads (GE Healthcare), incubating the reaction at 37°C for 30 min. The probe was denatured for 5 min at 100 °C and snap cooled on ice prior to addition to the hybridisation buffer. The probe p5B1.4 selected for the analysis of Southern blot analysis of restriction digested genomic DNA was purified with G-50 columns (Microspin G-50, Amersham Biosciences) before addition to the hybridisation buffer.

2.2.7 Southern hybridisation

The membranes were pre-hybridised for 30 min at 65°C with 10 ml of hybridisation buffer (7% (w/v) SDS, 1 mM EDTA, 0.5M Na_2HPO_4), which was then replaced by 5 ml of the same solution containing the radiolabelled probe. To block non-specific binding of p5B1.4 probe to the membrane, 22 μg of single stranded salmon testes DNA (SIGMA) was added in the pre-

hibridisation step. The membranes were hybridised overnight in a rotating oven at 65 °C. Subsequently the blots were washed with washing solution (0.2% (w/v) SDS, 0.2X SSC): 2x30 min at 65°C and 1x 30 min at room temperature. The membranes were dried at 80°C, exposed to X-ray film and developed using an X-Ograph Compact X2 system.

2.2.8 Genotyping of single nucleotide polymorphisms (SNPs)

Polymorphisms were genotyped by PCR following restriction digest (detailed in table 2-6). The SNPs rs207070736 and rs572634 were analysed as previously described (Neville*, et al*. 1994). To design the analysis of the majority of the SNPs DNannotator was used (http://bioinfo.bsd.uchicago.edu/SNP_cutter.htm). PCR products were digested with the appropriate restriction enzyme in a final volume of 20 μl following the supplier's recommendations. The products were resolved on an agarose gel and detected by ethidium bromide staining. Allele-specific primers were designed for the analysis of rs16980013 and rs635299, since no other option was available.

Table 2-6. Polymorphims genotyping assays

Forward primer	5'-3' sequence	Reverse primer	5'-3' sequence	refSNP ID	AT (˚C)	Product size (bp)	Restriction enzyme
839	AGGGCCCCTCATCAAAGTCCACGGTGT	823	ACGGTTCTGCAGAGTGGAAGT	rs2070736	55*	183	*Dra*III
Intron2-For	CTGCCTGTCCAACATGTCAGC	Intron2-Rev	CTCCAGAGTGGTGGCATAGGAC	rs657640§	63	365	*Nla*III/*Taq*I
463	CCGTCTCCACTCTGTCTCACT	600	GCTCTTGTCCCTCTTCCTAGG	rs572634	55*	334	*Ban*I
Intron5-For	GGTCATGGAGTATTACGTGGGC	Intron5-Rev	AGGGTGGGATAAATGAACCTCC	rs1799894§	67	397	*Hin*P1I
I8A	CACCGTGTAACAAGCTGTCATGGC	I8R	ACAGATGTGAGCAGCAGTCGTCAG	rs4646995	64	930/1907	N/A
HinfI-for	ATTGGCCTACCTGGGACTCT	HinfI-rev	AGCCCTCACCTTTTCTCTCC	rs16939§/ rs527221§	63	333	*Hin*fI/*Bpm*I
Fnu-For	GAGTCTTCCAGGAGCCACAG	Fnu-Rev	TTGACTGTGGGGAGGTAAGG	rs915915	63	316	*Fnu*4HI
490	CTCCGATCGGGTCACCTGTC	505TAG	TCACATGGTCCGGACGCAGCTAAGCGGGTGGCAA	rs635299	61	216	N/A
491	CTCCGATCGGGTCACCTGTC	506	CGCAGCTAAGCGGGTGGCAC		61	202	N/A
BAB452	CTGCTGCTGCTGCTGGG	DM-GR	GCAGGGCGTCATGCACAAGAAA	rs10419007	65	186	*Sau*3AI
DM-R	GTCCTCCGACTCGCTGACAG	3PDR	CAGCGTGTGCTTCTGGCTCAG	rs3745803§	68	762	*Taq*I
SIX-L	CACACACAAGGGGACAGAGA	SIX-D*R*	GGGTGGGTAAGAGTAACGGTCA	rs3745802§	62.5	1378	*Sau*3AI
SIX-K	GCCAGCTCTTGCAGACTTTG	SIX-D*R*	GGGTGGGTAAGAGTAACGGTCA	rs3745802§	66	1089	*Sau*3AI
SIX-J	TGGCGCAGTGGACAAGTAT	SIX-D*R*	GGGTGGGTAAGAGTAACGGTCA	rs3745802§	62.5	647	*Sau*3AI
SIX-C	CAGGTCAGCAACTGGTTCAAGA	SIX-DR	GGGTGGGTAAGAGTAACGGTCA	rs3745802§	63	457	*Sau*3AI
SIX-O	TCTCCACTTCTCTCTGGCTTGG	SIX-BR	CAGGCAGAAGGATGTGGTGACT	rs2014377§	60	800	*Pst*I
SIX-A	CTGGAACTAAGCGCAGGAACAG	SIX-BR	CAGGCAGAAGGATGTGGTGACT	rs2341097§	63	336	*Nla*III
SIX-P	TTTGGGTTTCCTTCCTAAGCCT	SIX-KR	TCGATGAACTTGTCCCTCCT	rs10775546§	64	166	*Aci*I
SIX-N	CCCAGGAGGGACAAGTTA	LOC-AR	CACAACCGCATCGTGAAG	rs16980013	61.6	214	N/A
TAG-SIX-M	TCATGCGTCCATGGTCCGGACCCAGGAGGGACAAGTTC	LOC-AR	CACAACCGCATCGTGAAG		61.6	234	N/A
TaqI-For	TCCCCAAGTTACCTCCTGGT	TaqI-Rev	AGGGCCCCTCTAGAACACAT	rs10415988§	63	167	*Taq*I

* Amplification was carried out for 10 cycles (94°C for 1min, 60°C for 1min 30s, 70°C for 3 min), 25 cycles (94°C for 1min, 55°C for 1min 30s,70°C for 3 min), 55°C for 1min and 70°C 10 min (Neville*, et al.* 1994). §DNannotator was used to design the analysis. N/A (not applicable). Underlined base indicates the error introduced to create a restriction site, base in red gives specificity to the primer.

2.2.9 Cloning and sequencing

Purified PCR products were cloned with TOPO TA Cloning® Kit for Sequencing (Invitrogen) with the following modifications: 4 µl of PCR product were ligated to the vector (10 ng), the ligation was incubated for 30 min at room temperature and 4 µl TOPO cloning reaction was added into the vial of one shot TOP10 chemically competent cells. Each transformation (10-200 µl) was spread on a prewarmed Luria Bertani (1% (w/v) Bacto-tryptone, 0,5% (w/v) bacto yeast extract, 1% (w/v) NaCl) containing ampicillin (100 µg/ml) plate and incubated overnight at 37°C. At least 20 different colonies were picked with a sterile tip, streaked first onto a fresh plate to generate a patch plate and then the tip was resuspended in 20 µl of water. After incubation for ~30 min at -70°C, 1 µl of the colony suspension was amplified by standard PCR using specific primers to verify the presence of the insert.

Positive colonies were selected and grown overnight in Luria Bertani (LB) with ampicillin (100 µg/ml) with vigorous shaking at 37°C. If the volume of the culture was between 3 and 5 mls, the cells were harvested by centrifugation at 6,800 x g for 3 min and the plasmid DNA isolated with a Qiaprep spin miniprep kit. To increase the DNA yield, a starter culture of 2–5 ml LB medium with ampicillin was inoculated with a single colony and incubated for ~8 h at 37°C with vigorous shaking. The starter culture was diluted by inoculating 100 ml medium with 100–200 µl of the starter culture and grown at 37°C for 12–16 h with vigorous shaking. The cells were harvested by centrifugation at 4,800 x g for 20 min at 4°C. The plasmid DNA was isolated with endofree plasmid maxi kits (Qiagen). The DNA was sent for sequencing to Geneservice, Cambridge Science Park, Milton Road, Cambridge, CB4 0FE. The sequences were visualised and analysed using ApE-A plasmid editor v1.12.

2.2.10 Methylation studies

To investigate the pattern of DNA methylation 60 ng of genomic DNA were digested overnight with a methylation sensitive restriction enzyme of

choice in a final volume of 20 μl. A control reaction without enzyme was also performed for each sample. The reactions were terminated by heat-inactivation for 20 min at 65°C. The samples were amplified as follows: 30 cycles (96°C for 45 sec, AT for 45 sec, 70°C for 50 sec), AT for 1 min and 70°C for 3 min. The PCR products were resolved on a 1.5% agarose gel and detected by Southern blot hybridisation.

3 Identifying the nature of the genetic lesion in the DM1+CMT++ family using a combination of molecular genetic approaches

3.1 Introduction

DM1 is the most common form of muscular dystrophy in adults, and is characterised by multiple tissue involvement including skeletal myotonia, progressive myopathy, cataracts and abnormalities in the heart, brain and endocrine systems. The DM1 mutation has been identified as an expansion of a pure $(CTG)_n$ repeat located in the 3'-untranslated region (UTR) of the *DM protein kinase* gene (*DMPK*) which maps to 19q13.3 (Aslanidis, *et al*. 1992; Brook, *et al*. 1992; Buxton, *et al*. 1992; Fu, *et al*. 1992; Harley, *et al*. 1992; Mahadevan, *et al*. 1992). The *DM1* locus is highly polymorphic and in normal individuals contains 5 to 35 copies of the CTG repeat. In DM1 patients this repeat is expanded in the range of 50 to 100 in those mildly affected, 200 to 1000 in those with the adult classical form and more than 1000 in those with the congenital form (IDMC, 2000). Expanded alleles are genetically highly unstable and nearly always increase in length when transmitted from parent to child, resulting in an increase of disease severity and earlier age of onset in successive generations, a phenomenon termed genetic anticipation. In the first generation of a classic DM1 family, an individual presents a mild form with only a few symptoms such as cataracts developed in old age but generally with no muscle symptoms. However, an enlarged allele is transmitted to offspring. In the second generation, a classical adult form with an age of onset between the third and fourth decade often occurs. Once again, the repeat expands on transmissions, maternal transmissions will generally result in a congenital form with a very severe phenotype including hypotonia and mental retardation in the third generation (Harper, *et al*. 1992).

CMT is a group of genetically heterogeneous neuropathies affecting motor and sensory nerves of the peripheral nervous system (Shy 2004). The

phenotype is characterised by wasting and weakness of distal limb muscles, generally with distal sensory loss, skeletal deformities (usually *pes cavus*), and decrease or absence of tendon reflexes. In addition, hearing loss is observed in 5% of patients with CMT. Nerve conduction is one of the parameters assessed to define the type of CMT (Pareyson 2003). CMT1 (the demyelinating type) is characterised by a motor median nerve conduction velocity less than 38 m/s, CMT2 (the axonal type) is characterized by normal or slightly reduced nerve conduction velocities (≥ 45 m/s), and intermediate CMT is characterized by nerve conduction velocities in a range of 25-45 m/s (Davis, *et al*. 1978; Harding and Thomas 1980). The inheritance pattern can be autosomal dominant, autosomal recessive, or X-linked. To date, 30 different loci have been mapped, and mutations have been identified in more than 20 different genes. However, there are still patients with an unknown aetiology. CMT1A is the most common type (40-50 % of all CMT patients) and is caused mainly by duplications of the *PMP22* gene (17p11.2-p12)(Matsunami, *et al*. 1992; Timmerman, *et al*. 1992; Valentijn, *et al*. 1992).

The DM1+CMT++ family is the first reported pedigree with three generations in which all the affected patients present clinical and/or electromyographic features of both DM1 and intermediate type CMT. Although there are other reported cases where co-incidence of DM and CMT has been observed (Rogoff and Ziegler 1956; Caccia, *et al*. 1972; Panayiotopoulos and Scarpalezos 1976; Borenstein, *et al*. 1977; Cros, *et al*. 1988; von Giesen, *et al*. 1994), CMT is not a recognised feature of DM1 (Harper 2001). Genetic anticipation is only partially observed in this family. Even though an increase in the severity of symptoms and a decrease in the age of onset are observed between the second and third generation, no congenital cases were observed in the fourth generation despite four female transmissions. On the contrary, all affected members in the fourth generation are intelligent, energetic, obtained university degrees and perform skilled jobs despite their physical disability. Unexpectedly, a decrease in the age at onset and an increase in the severity of the symptoms of CMT were observed between the third and the fourth generation. Genetic anticipation is not a common feature of CMT, but there

are similar reported cases where genetic anticipation was suspected (Hamiel, *et al*. 1993; Marrosu, *et al*. 1997; Dupre, *et al*. 1999; Kovach, *et al*. 2002). The mutations responsible for CMT were point mutations in the *PMP22* gene (Marrosu, *et al*. 1997; Kovach, *et al*. 1999; Kovach, *et al*. 2002), point mutations in the *Po myelin* gene (Bird, *et al*. 1997), and a duplication in the *PMP22* gene (Dupre, *et al*. 1999). The mutation in the family described by Hamiel and collaborators is unknown. The cause of anticipation is still unknown in these reported families and there is insufficient evidence to reject ascertainment bias.

In addition to the DM1 and CMT symptoms, several affected individuals (II-26, II-27, II-28, III-9, III-12, III-16, III-17 and IV19) developed early hearing loss, a symptom that has been observed in 5% of the patients with CMT (Pareyson 2003). Four affected individuals (II-28, III-9, III-16, III-17) have also developed recurrent episodes of acute encephalopathy. The episodes of encephalopathy were characterised by confusion, headache, psychomotor agitation, fever impaired consciousness, and even coma resulting in several hospital admissions. Disorders with similar symptoms include familial hemiplegic migraine types 1 (FHM1) and 2 (FHM2), cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) and acute confusional migraine (ACM). FHM1 can be caused by mutations in the *calcium channel* gene (*CACNA1A*) (Ophoff, *et al*. 1996), FHM2 by mutations in a gene encoding the *alpha-2 subunit of the sodium/potassium pump* (*ATP1A2*) (De Fusco, *et al*. 2003), and CADASIL by mutations in *NOTCH3* (Joutel, *et al*. 1996). Screening of the most common mutations in these genes was performed in either one or two affected members of the DM1+CMT++ family, but was negative.

Spaans and colleagues showed by linkage analysis that DM1 and CMT co-segregate with the *APOC2* locus on chromosome 19 (~2 cM from the *DMPK* gene) with a maximum LOD score of 7.04 with zero recombination, but do not segregate with chromosome 1 or 17 (Spaans, *et al*. 1986; Brunner, *et al*. 1991). Furthermore, Southern blot analysis of restriction digested genomic DNA has revealed a fragment equivalent to a small CTG expansion at the *DM1* locus in all affected members. However, Spaans and colleagues

have not been able to PCR across the expanded allele, nor to visualise the expanded allele by repeat primed-PCR (RP-PCR), which allows the detection of long expansions (explained in detail in section 3.2.1.3). Finally, standard cytogenetic studies and FISH using probes for a pericentric or paracentric inversion of chromosome 19, revealed no abnormalities (Spaans, *et al*. 2008). These data suggest that either a single or multiple closely linked mutations located near the *APOC2 marker* cause the DM1 and CMT symptoms in this family. There are three different types of CMT that have been mapped to chromosome 19. Dominant intermediate type B (DI-CMTB) associated with mutation in *DNM2* (19q12-13.2), recessive type 3 (CMT3D or CMT4F) associated with mutation in the *PRX* gene (19q13.1-13.2) and an axonal recessive form, CMT4C3, that maps to 19q13.3 (Delague, *et al*. 2000; Leal, *et al*. 2001; Zuchner, *et al*. 2005). The *DNM2* gene is probably not mutated since it is located ~40Mb distal to the *DMPK* gene and no abnormalities were observed by FISH. However, the *ARCMT2B* locus or the *PRX* gene located respectively at 2Mb and 5Mb from the *DMPK* gene might be involved.

The $(CTG)_n$ expansion is the cause of DM1, however the DM1+CMT++ family has an unusual molecular diagnosis suggesting that another modification is present in addition to this expansion. It is known that once the number of CTG repeats passes a certain threshold, the repeats become very unstable in the soma and in the germline. Although transmissions from patients usually involve expansions, rare reverse mutations into the normal range have occurred on transmission to offspring in a number of DM1 families (Shelbourne, *et al*. 1992; Brunner, *et al*. 1993b; Hunter, *et al*. 1993; O'Hoy, *et al*. 1993; Shelbourne, *et al*. 1993). Strikingly, in one case a gene conversion event occurred in the father, two regions of the mutated chromosome including the CTG repeat tract were exchanged with the normal chromosome and the individual inherited an allele within the normal range (O'Hoy, *et al*. 1993). These data suggest that the instability within the repeats could also encompass the flanking DNA. A deletion in the flanking DNA has not yet been reported in DM1 patients. However in bacteria, long CTG repeat tracts inserted into a plasmid promoted the formation of inversions and long deletions that removed part or all of the

repeats as well as some of the flanking sequence (Wojciechowska, *et al*. 2005). Furthermore, in a number of fragile X syndrome and Jacobsen syndrome patients, diseases that are both caused by expansion of CGG triplet repeats, the expansions predispose to additional DNA instability leading to deletions encompassing the repeat tract and the flanking DNA (Meijer, *et al*. 1994; de Graaff, *et al*. 1995; Jones, *et al*. 1995; Quan, *et al*. 1995).

It appears logical to speculate that the expanded $(CTG)_n$ repeat at the *DMPK* locus may predispose the repeat tract and the flanking DNA to further instability, generating a putative break where insertions, deletions and/or rearrangements might occur (figure 3-1). These novel mutations might not be identified with standard diagnostic techniques, and they could modify the expression of *DMPK* and/or other genes and so explain the complex symptomatology observed in the DM1+CMT++ family. A combination of molecular approaches was performed to investigate the molecular lesion in the DM1+CMT++ family.

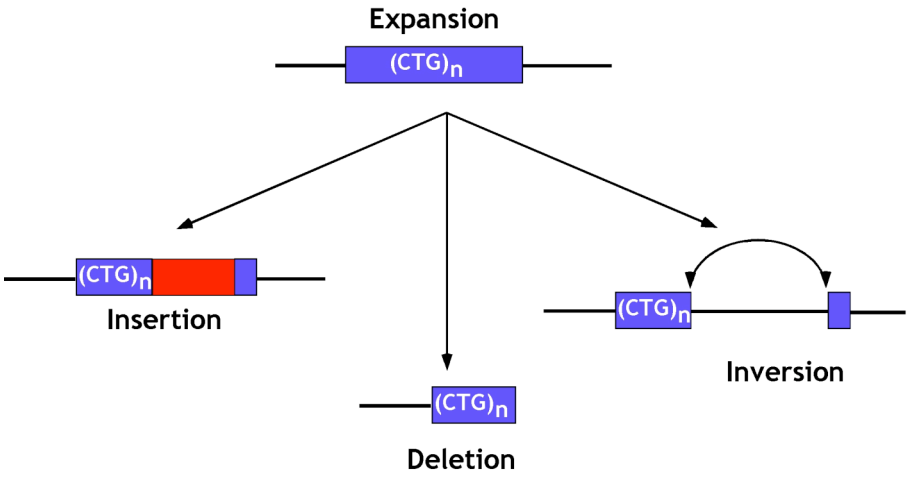


Figure 3-1. Three possible novel types of mutation in the *DM1* locus. As observed in fragile X syndrome and Jacobsen syndrome, the instability of the expanded repeat at the *DM1* locus could generate novel mutations such as insertions, deletions and/or inversions that might explain the complex symptomatology observed in the DM1+CMT++ family.

3.2 Results

3.2.1 Genotyping the *DM1* locus by traditional methods

In spite of the fact that Southern blot of digested DNA, PCR and RP-PCR had already been performed on patient DNA from the DM1+CMT++ family by the Spaans group, this project began by genotyping the *DM1* locus by PCR in order to confirm the previous data.

3.2.1.1 PCR

The genotype of the *DMPK* locus was assessed by standard PCR amplification across the repeat array with the primers DM-A and DM-DR (figure 3-2). As expected, only normal size alleles were detected in unaffected controls (U1 and U10), and both the normal and expanded alleles were observed in classic control DM1 patients (A2, A3 and A4). Large expanded alleles were only detected after Southern blot hybridisation (figure 3-2C). The mutant allele was observed as a heterogeneous smear comprised of fragments containing different number of CTG repeats, revealing the somatic mosaicism observed at the *DM1* locus (Monckton, *et al*. 1995b; Wong, *et al*. 1995). As expected, individuals IV-11 and IV-12 from the DM1+CMT++ family, who are both unaffected, present as heterozygous (38/5) carrying two normal sized alleles. However, III-9, III-16, IV-19, IV-20, III-17 and IV-21 appeared as homozygous, since only a single band corresponding in size to a normal allele was detected. The band detected in III-9 corresponds to an allele containing ~38 repeats. This allele is below the pathogenic range but occurs at a very low frequency (0.07%) in Caucasian populations (Martorell, *et al*. 2001), hence it would be extremely unlikely (1 in 2 million) to find a homozygous individual for an allele of that size. Consequently, a ‘null allele’ (failure to detect the product by the technique) was suggested. A 'null allele' could not be confirmed because DNA from the unaffected parent was unavailable.

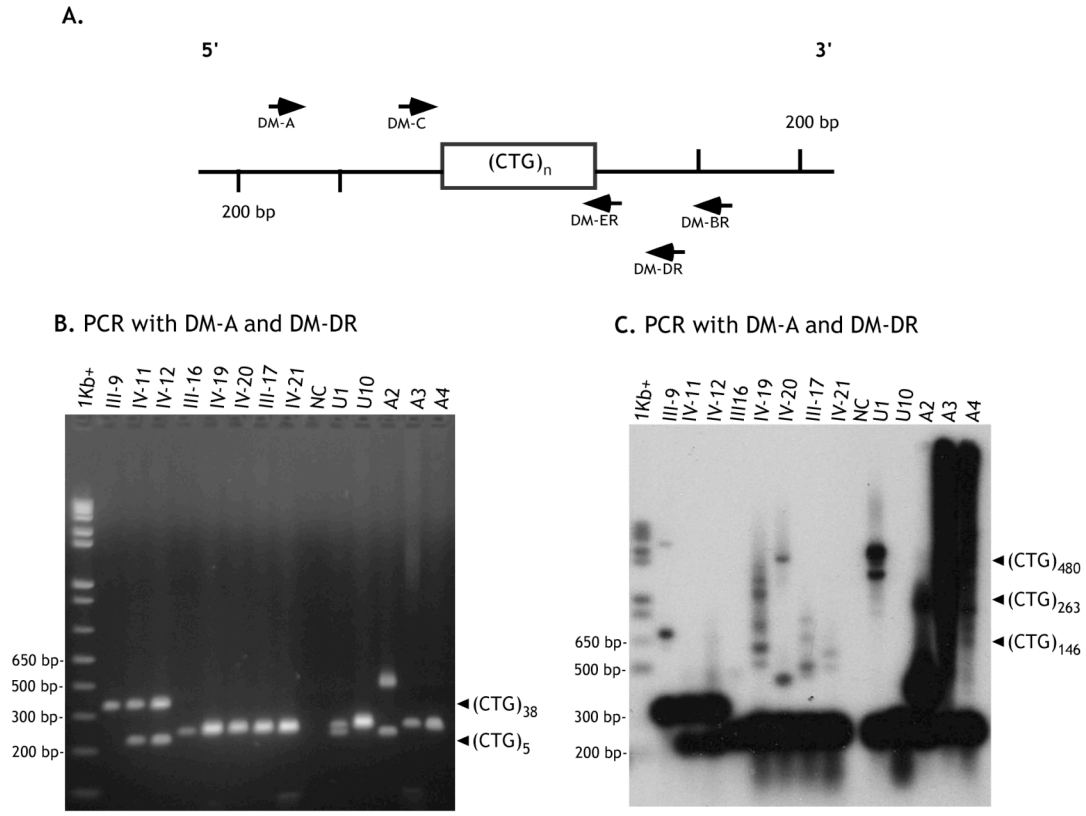


Figure 3-2. Genotyping of the *DMPK* locus by standard PCR. A. Schematic diagram of the *DM1* locus. The location of the primers used to PCR across the (CTG) tract is indicated. B. PCR products generated with primers DM-A and DM-DR, which amplified 211 bp of flanking DNA, were resolved on a 1.5% ethidium bromide-stained agarose gel. C. The products were subsequently detected by Southern blot hybridisation with a DM1 CTG repeat probe. U1, U10 (normal individuals), A2, A3, A4 (classic DM1 patients), NC (negative control, PCR without DNA). The sizes of the marker are indicated on the right side and the estimated number of repeats is indicated on the left side. The expanded allele can be visualised as a smear in A2, A3 and A4. Additional bands are also visualised in III-9, IV19, IV-20, III-17, IV-21 and U1. However, the intensity is not as strong as in the classic DM1 patients so they could be either heteroduplex DNA or contamination.

A 'null allele' could be caused by either a technical artefact, a single nucleotide variation within the location of the primers in the expanded allele, or a putative deletion, insertion or rearrangement in the flanking DNA disrupting several priming sites. The affected members in the DM1+CMT++ family showed either the late onset form (second generation) or the adult form of the disease (third and fourth generations) so the number of $(CTG)_n$ repeats was not expected to exceed 1,000 repeats in any patient. A technical artefact does not seem likely since alleles up to 2,000 CTG repeats (as observed in the congenital form) can be readily amplified by PCR (Monckton, *et al*. 1995b; Martorell, *et al*. 1998). To test for a single nucleotide variation in the primer site, three different combinations of primers (DM-C/DM-DR, DM-A/DM-BR, DM-C/DM-ER) were assessed by PCR, but an expanded allele in the DM1+CMT++ family members was still not

detected (data not shown). Therefore, it seemed reasonable to suggest that an insertion within the CTG repeat tract, a deletion or rearrangement in the flanking DNA might explain the absence of an expanded allele detected by PCR.

3.2.1.2 Southern blot analysis of restriction digested genomic DNA

Southern blot analysis of restriction digested genomic DNA was carried out with *Eco*RI and *Pst*I. Both enzymes only have restriction sites in the flanking DNA so the length of the repeat array in the *DM1* locus can be assayed (figure 3-2A). The probe p5B1.4 in *Eco*RI-digested DNA detects a fragment of 8.7 Kb and a larger fragment of 9.6 Kb. Both fragments contain the variable polymorphic CTG trinucleotide tract and a 0.9 Kb insertion/deletion polymorphism located 4.7 Kb 5'-end distal to the CTG repeat tract (figure 3-3A). The 977 bp difference between the fragments corresponds to a deletion of three Alu repeats in the smaller fragment (Alu-). The analysis of the structure of the two alleles and the presence of the Alu+ allele in primates suggested that the ancestral allele is Alu+ whilst Alu- is a derived deletion (Mahadevan, *et al*. 1993b; Rubinsztein, *et al*. 1994). The Alu+ allele has been shown to be in complete linkage disequilibrium with the DM1 mutation in Eurasian populations (Harley, *et al*. 1992; Mahadevan, *et al*. 1992; Yamagata, *et al*. 1992). Although, the DM1 mutation was associated with the Alu- allele in one Nigerian kindred (Krahe, *et al*. 1995), this is not usually the case. Therefore, we would expect that the DM1+CMT++ mutation would most likely be found on an Alu+ 9.6 Kb+ *Eco*RI fragment.

Two classic DM1 patients, A7 (~200/13) and A2 (~50/9), were included as controls. Southern blot analysis in *Eco*RI-digested DNA revealed two fragments in A7. The larger fragment (>9.6 Kb) contained ~200 CTG repeats and the Alu+ allele. The bottom band (8.6 Kb) contained 13 CTG repeats and the Alu- allele (figure 3-3B). In A2, only a single band with a size of 9.6 Kb was observed. This patient was probably homozygous for the Alu+ allele, but the resolution of the gel was not sufficient to differentiate the 9 repeat allele from the 50 repeat allele. III-16, IV-19 and IV-20 were heterozygous.

Two fragments were observed, the 8.7 Kb band that presumably contained the normal size allele on an Alu- fragment, and a larger fragment. Only a slight increase in size was observed in the larger fragment relative to the normal 9.6 Kb Alu+ allele, suggesting a small expansion in the range of ~150-200 CTG repeats on a Alu+ fragment. Only a single ~9.6 Kb band was detected in IV-11 and IV-12, indicating that both individuals were probably homozygous (Alu+/Alu+) and the difference in the number of repeats between the alleles (38/5) was insufficient to be resolved as two bands in this percentage of agarose gel. In III-9 two larger fragments were obtained, presumably the bottom fragment contained the normal size allele (38 repeats) whilst the top fragment contained the expanded allele (~230 repeats), both on an Alu+ background.

The probe p5B1.4 detected two fragments in *Pst*I-digested DNA: the top one containing the variable polymorphic CTG trinucleotide repeat (1.2 kb), and the bottom one of 886 bp without the variable tract. The use of *Pst*I increased the resolution of the test by providing smaller fragments to be resolved in a higher percentage of agarose gel (1% agarose gel), but the level of background was high (data not shown). Only the fragment corresponding in size to the normal CTG allele was visualised in III-9, III-17 and IV-19. However, two discrete bands were detected in IV-20, the normal size allele and the mutated allele. The low level of somatic mosaicism present in IV-20, probably because of his young age, may explain this result. In III-9, III-17 and IV-19 the level of somatic mosaicism was probably higher than in IV-20 and consequently the fragments were spread out in the gel preventing their detection.

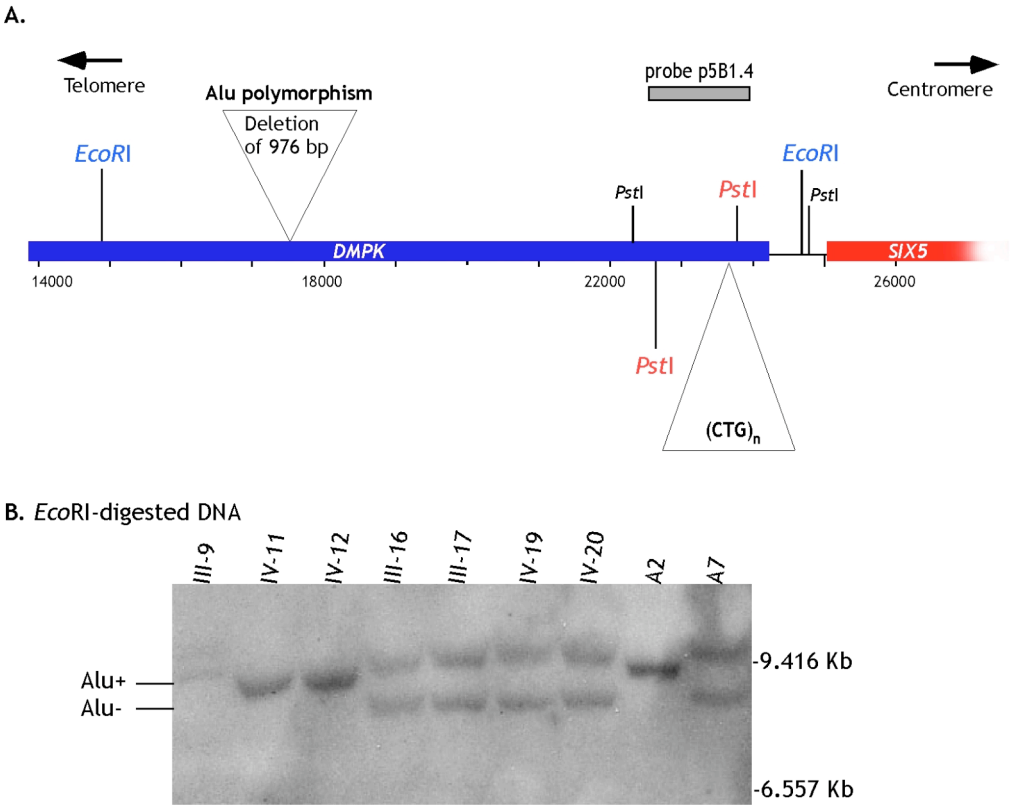


Figure 3-3. Southern blot of restriction digested genomic DNA. A. Restriction map of the *DMPK* locus showing the sites of *Eco*RI and *Pst*I and the probe p5B1.4. B. *Eco*RI-digested DNA resolved in 0.7% agarose gel and probed with p5B1.4. The large (>9.8-kb) band varies in size and defines the DM1 mutation. In A7 (classic DM1 patient) the >9.8Kb and 8.6 Kb fragments are detected. A2 (another classic DM1 patient), who is probably homozygous for the Alu+ allele, the expanded allele could not be resolved from the normal size allele due to the low resolution of the gel and the small size difference between both alleles. Two faint bands are present in III-9. The sizes of the marker are indicated on the right, and the expected sizes of the Alu+ (9.6 Kb) and Alu- (8.6Kb) alleles on the left.

3.2.1.3 Repeat primed-PCR (RP-PCR)

RP-PCR was first described in the context of minisatellite DNA repeat amplification (Jeffreys, *et al*. 1991), and then it was adapted for the diagnosis of neurodegenerative diseases when long repeats could not be amplified by using flanking primers (Warner, *et al*. 1996). The strategy involves a PCR containing three types of primers: a specific primer flanking the repeat array (P1 or P4), and a second primer (P2+TAG or P3+TAG) targeted to the repeat tract containing a 5' tail to be used as an anchor for the third primer (TAG) (figure 3-4). The sequence of the tail is carefully selected so that it does not contain sequence identity to the genome being investigated. The concentration of the primer targeted to the repeat tract is used at 1/10 of the other primers so that it is exhausted in the earlier cycles, reducing the shortening in the size of the PCR product generated by

priming in internal sites. In alleles within the normal range, a characteristic ladder of fragments corresponding to the number of repeats will be obtained, in contrast in expanded alleles the ladder will continue higher up (> 50 repeats) allowing a distinguishable pattern. Accurate estimations of the number of repeats are not possible. An absence of product at either end of the repeat tract (5'-end or 3'-end) could be suggestive of a local rearrangement, deletion or insertion.

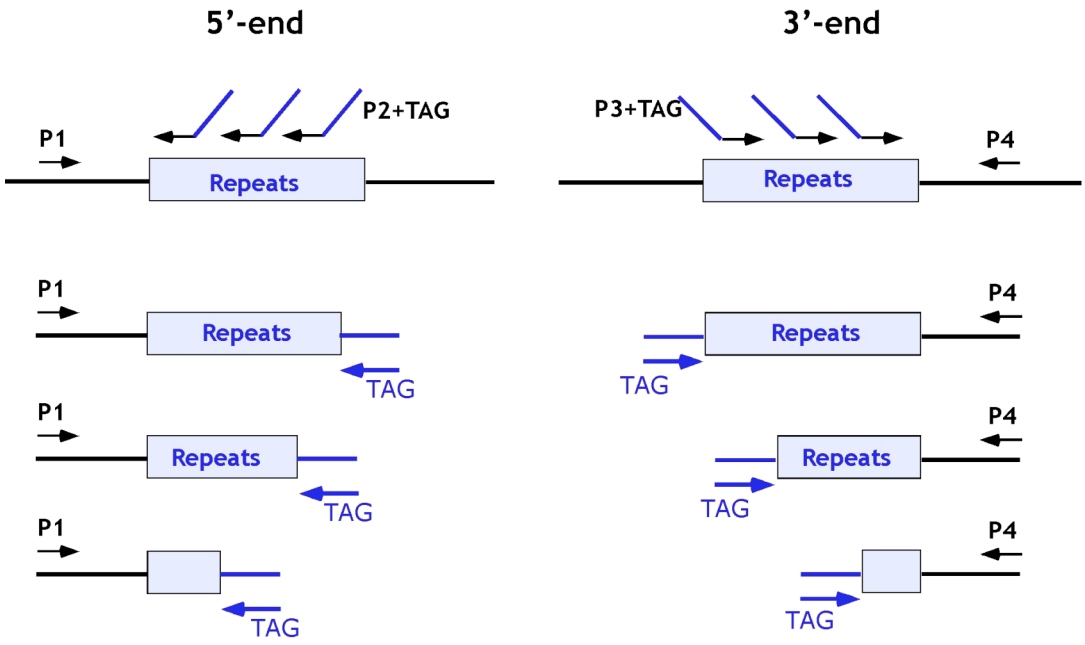


Figure 3-4. Schematic representation of RP-PCR from both ends. P1 and P4 are the primers that give specificity to the reaction. P2+TAG or P3+TAG are targeted to the repeat tract and therefore will prime in multiple positions within the repeat tract generating fragments of different lengths that will be amplified with TAG.

When using RP-PCR to investigate the DM1+CMT++ family, a ladder of expanded alleles was detected at the 5'-end using multiple flanking primers (DM-T, DM-A, DM-H and DM-C) located up to 400 bp from the repeat in the affected family members investigated as well as in the classic DM1 patients (A6 and A8)(figure 3-5). As expected, no expansions were detected in the unaffected family members (IV-11 and IV12) and in the normal controls (U10 and U11). These data suggest that the 5'-end seems to be intact up to ~400 bp. A different scenario was observed at the 3'-end of the repeat array. Multiple flanking primers (DM-ER, DM-DR, DM-BR, DM-QR, DM-3PBR, DM-3PCR and DM-3PDR) up to 1 kb distal from the CTG repeat array failed to detect expanded alleles in the DM1+CMT++ patients.

It seemed reasonable to conclude from this data that the expansion at the *DM1* locus in the affected individuals was coupled with a deletion or a rearrangement at least up to 1kb at the 3'-end, since multiple PCR primers sites did not appear to be present. If so, the *Eco*RI fragments detected by the probe p5B1.4 on the Southern blot analysis could be generated with more distal *Eco*RI sites.

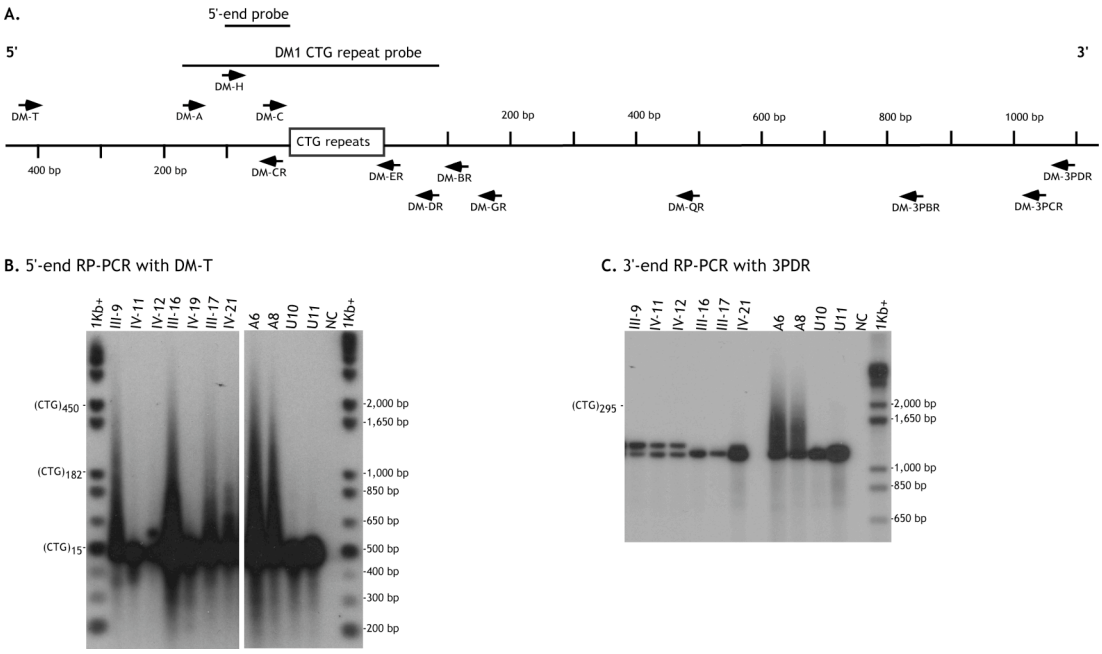


Figure 3-5. RP-PCR analysis. A. Schematic representation of the *DM1* locus showing the locations of the different primers (arrows) and the CTG repeat tract (box). The two probes used in the analysis are also indicated. B. 5'-end RP-PCR analysis using primer DM-T located at 434 bp from the repeat tract (P1 from figure 3-4). C. 3'-end RP-PCR analysis using primer 3PDR located at 1093 bp from the repeat (P4 from figure 3-4). The products were resolved on a 1.5% agarose gel and subsequently detected by Southern hybridisation. The DM1 CTG repeat probe was used to detect the 3'-end products and a 5'-end probe (generated with primers DM-H and DM-CR) was used to detect the 5'-end products. Non-specific products were visualised on the negative control at the 5'-end so a probe was designed to hybridise only at the 5'-end of the repeat tract and the problem was solved. A6 and A8 (classic DM1 patients); U10 and U11, (normal individuals); NC (negative control). The sizes of the marker are indicated on the right and the estimated number of repeats is indicated on the left.

Assuming the 5'-end is intact, the next *Eco*RI site must be very close. Indeed, the next *Eco*RI site is located in the *LOC388553* gene (~11Kb). If a large deletion or rearrangement occurred, a fragment generated with a primer located at the 5'-end of the repeat and a primer located after the next *Eco*RI site (such as DMEcoR) should be obtained (figure 3-6). Two different forward primers (DM-A and DM-C) were used with DMEcoR but a product was not obtained in either case. In addition, RP-PCR was performed using the primer DMEcoR as the specific primer but the mutant

allele could not be visualised. These data seemed to argue against a deletion or rearrangement of that size.

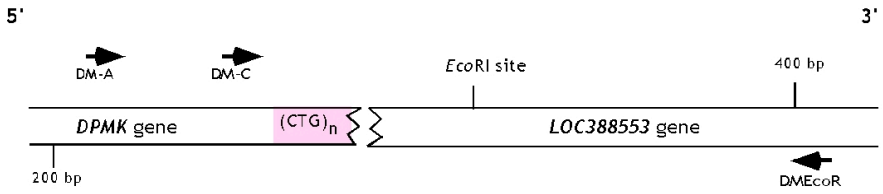


Figure 3-6. Schematic diagram of a putative deletion at the 3'-end of the $(CTG)_n$. The primers selected to investigate the putative deletion are indicated. The *Eco*RI site indicated is located in the *LOC388553* gene at ~11Kb distal from the $(CTG)_n$ in the reference sequence. If a deletion or rearragement occurred in the DM1+CMT++ mutated chromosome by using forward primers (DM-A or DM-C) with the reverse DMEcoR primer a product might be generated.

It was possible that the extent of any deletion was not as large as ~11Kb, so another attempt at PCR and RP-PCR was performed but using primers located in the *SIX5* gene. The primer DM-C was used in combination with multiple reverse PCR primers (SIX-ER, SIX-IR and SIX-DR) (figure 3-7). Both the expanded and normal alleles in the classic DM1 patients were amplified. However, in the DM1+CMT++ family only the normal size alleles could be detected. In addition, RP-PCR at the 3'-end was performed with SIX-DR. The expanded allele in the classic DM1 patients was visualised, but only the normal size allele was visualised in the DM1+CMT++ cases. These data seem to argue against a deletion of ~1Kb.

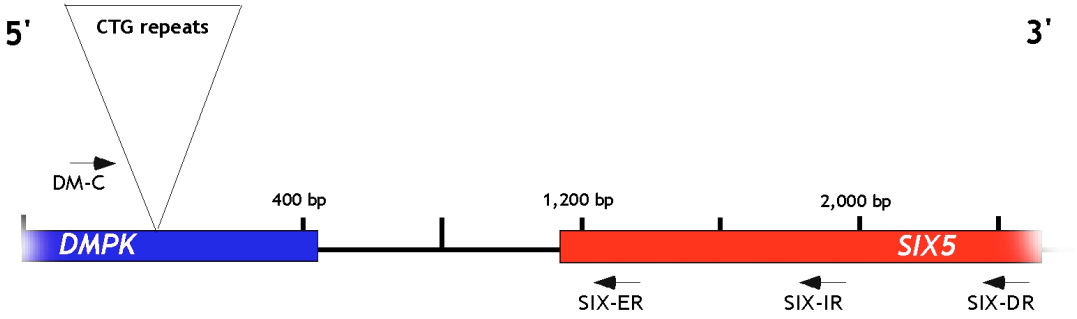


Figure 3-7. Schematic diagram of the 3'-end of the repeat tract. The locations of multiple reverse primers are indicated. If there is a deletion of ~1Kb at the 3'-end, a product could be amplified by PCR using DM-C and a primer located after the deletion.

3.2.2 Defining the boundaries of a putative deletion

In an effort to narrow down the extent of any putative deletion, a loss of heterozygosity deletion analysis was performed. All patients should present as pseudohomozygous for all markers within a deleted region. Conversely, any SNPs that present as heterozygous in any patient will indicate that such a marker is not contained within a deletion (figure 3-8).

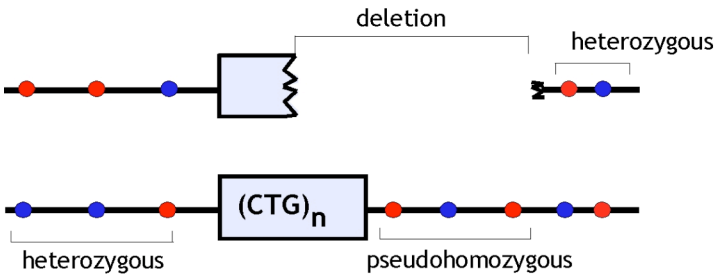


Figure 3-8. Loss of heterozygosity deletion analysis. Several SNPs located at either end of the repeat tract were investigated. If a deletion event is present, the patients should present as pseudohomozygous for one or more SNPs depending on the size of the deletion. The limit of a deletion event could be defined at the SNP where the individuals present as heterozygous.

The insertion/deletion polymorphism located in intron 8 of the *DMPK* gene was one of the first polymorphisms described in the literature at the *DM1* locus. The Alu+ allele was shown to be in linkage disequilibrium with an expanded (CTG) tract. Subsequently, eight more polymorphic sites spanning an ~40 Kb region around the *DMPK* gene were genotyped and 9-11 different haplotypes were defined in Caucasian and Japanese populations. However, only one of these haplotypes has been shown to be associated with the DM1 mutation (Neville, *et al*. 1994; Yamagata, *et al*. 1998). An exception was described in a Nigerian kindred, where the DM1 mutation was not associated with the classic DM1 haplotype (Krahe, *et al*. 1995).

To determine the extent of a putative deletion and to investigate whether the DM1+CMT++ mutation arose in the classic DM1 associated haplotype, nine polymorphic sites were genotyped to define the haplotype of each individual within a region of ~40 Kb on chromosome 19, extends from the 3'-end of the *DMWD* gene to the intron 1 of *LOC388553* gene. The polymorphisms were genotyped by restriction digests assays following PCR.

The insertion/deletion polymorphism (rs4646995) was detected by PCR using primers flanking the insertion site (figure 3-9). As expected, III-16, IV-19, IV-20, III-17 and IV-21 are heterozygous whilst III-9, IV-11 and IV-12 are homozygous for the Alu+ allele. These data confirm the previous data obtained by *Eco*RI-digested DNA and suggest the DM1+CMT++ mutation arose on the common Alu+ DM1 haplotype.

The construction of the haplotypes was performed manually with the data obtained from the genotype analysis (table 3-1). Unfortunately, the DNA of unaffected individuals was not available so it was not possible to unambiguously assign all haplotypes (figure 3-11). III9, IV-11 and IV-12 present as homozygous for all the SNPs genotyped except for the number of CTG repeats so there were only two possible haplotypes. III-9 is homozygous for the classic DM1 haplotype so the DM1+CMT++ mutation must have arisen in the common DM1 haplotype. At least one DM1 haplotype could be constructed in the other affected members investigated, suggesting that the mutation was also present on the DM1 classic haplotype.

Most interestingly, III-16, IV-19, IV-20, III-17 and IV21 were genotyped as heterozygous at several SNPs at the 5'-end of the CTG repeat tract (rs1799894, rs4646995, rs16939, rs915915 and rs635299) and at the rs1045988 located at ~30Kb at the 3'-end. These data indicated the 5'-end region was present and presumably intact. Heterozygosity at rs10415988 appeared to limit a simple 3'-end deletion to less than 40Kb.

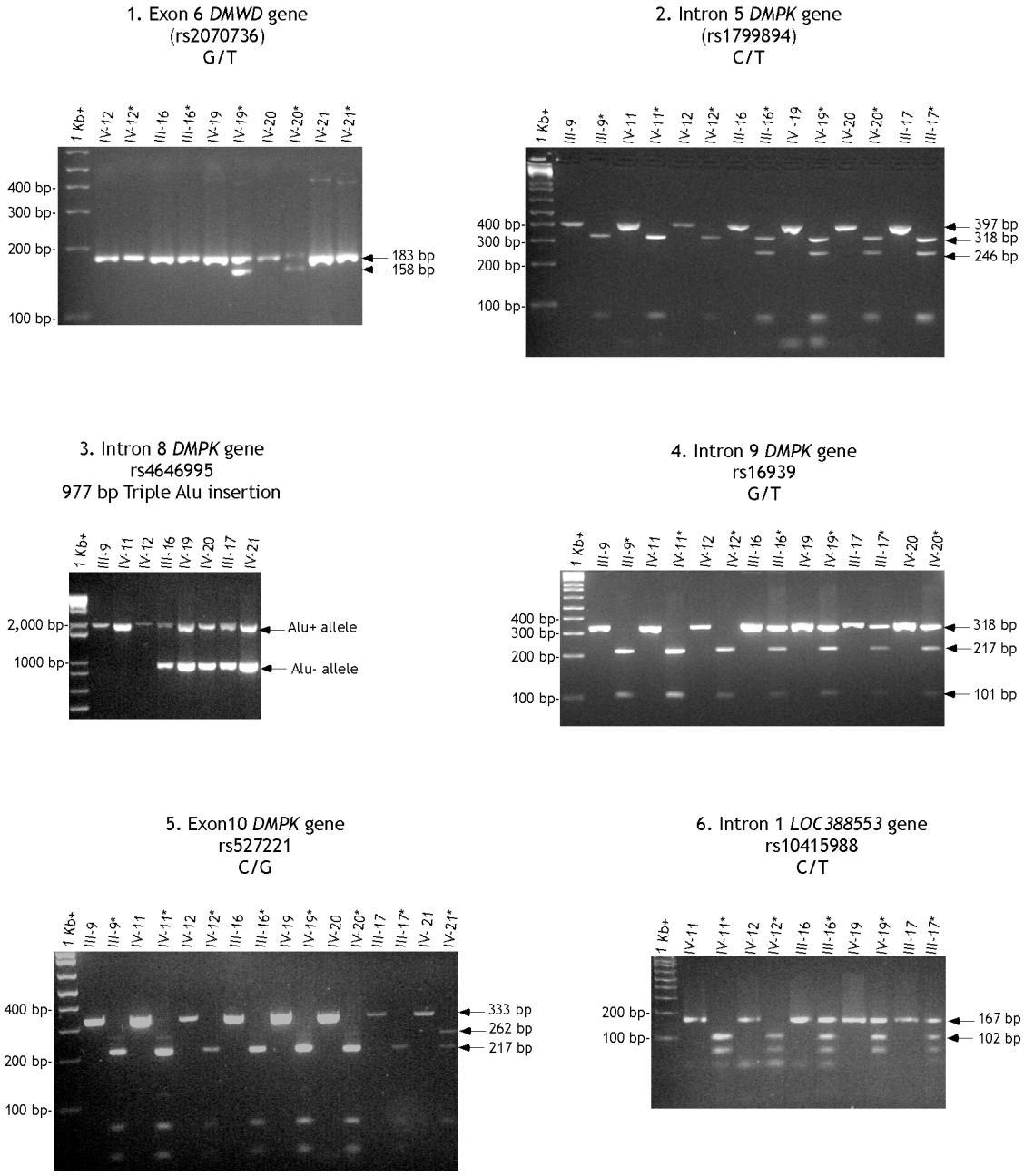


Figure 3-9. PCR based assays which genotype 6 polymorphisms in a region ~40 kb around the *DMPK* gene. PCR products were resolved on 2% Nusieve, 1% agarose gel, excluding number 3 that was resolved on 1.5% agarose gel. In each panel, the first lane contains the DNA marker followed by PCR products before or after (*) digestion. The sizes of the marker, as well as the sizes of the uncut and digested products are indicated.

Eight more SNPs with a known high value of heterozygosity were selected from the Entrez SNP database (http://www.ncbi.nlm.nih.gov/sites/entrez) to be genotyped at the 3'-end in an effort to reduce the boundary of a putative deletion (figure 3-10). Again several affected individuals were genotyped as heterozygous at variable SNPs (table 3-1). The closest polymorphic site to the 3'-end of the CTG tract is rs3745802, which is located ~2.3 Kb downstream of the CTG repeat in the *SIX5* gene and this SNP was heterozygous in several patients (figure 3-12A and B). Therefore,

the boundary of a putative deletion was reduced even more. However, it was not clear from this data if a deletion could be present upstream the rs3745802 SNP. The G allele seemed to be associated with the DM1+CMT+ mutation since all affected members carried at least one G allele.

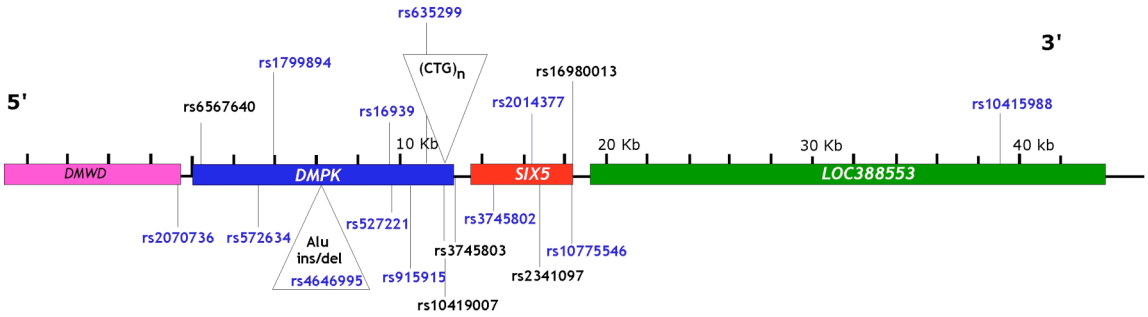


Figure 3-10. Diagram of the *DM1* locus. Shown are the relative positions of the different SNPs investigated, the SNPs, which at least one affected individual from the DM1+CMT++ family presented as heterozygous (in blue), the insertion/deletion polymorphism and the CTG repeats (in triangles).

Table 3-1. Genotype of several polymorphisms in the *DM1* locus

refSNP ID	Position (bp)*	H	III-9	IV-11	IV-12	III-16	IV-19	IV-20	III-17	IV-21
rs2070736	-978	0.408	1/1	1/1	1/1	1/1	1/2	1/2	1/1	1/1
rs6567640	+378	0.418	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
rs572634	+3232	0.226	2/2	2/2	ND	2/2	2/2	2/2	2/2	1/2
rs1799894	+3991	0.495	1/1	1/1	1/1	1/2	1/2	1/2	1/2	1/2
rs4646995	+6099	0.498	+/+	+/+	+/+	+/-	+/-	+/-	+/-	+/-
rs16939	+9680	0.498	2/2	2/2	2/2	1/2	1/2	1/2	1/2	1/2
rs527221	+9760	0.320	2/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2
rs915915	+10764	0.495	1/1	1/1	1/1	1/2	1/2	1/2	1/2	1/2
rs635299	+11565	NN	T/T	T/T	T/T	G/T	G/T	G/T	G/T	G/T
$(CTG)_n$	+12223		38/N	38/5	38/5	14/N	14/N	14/N	14/N	14/N
rs10419007	+12278	NN	2/2	2/2	2/2	2/2	2/2	2/2	ND	2/2
rs3745803	+12773	NN	ND	1/1	ND	1/1	1/1	1/1	ND	1/1
rs3745802	+14559	0.429	1/1	1/1	1/1	1/2	1/2	1/2	1/2	1/2
rs2014377	+16423	0.260	2/2	1/2	1/2	1/2	1/2	1/2	ND	1/2
rs2341097	+16834	0.386	1/1	1/2	1/2	1/1	1/1	1/1	1/1	1/1
rs10775546	+18202	0.303	1/2	1/1	1/2	1/2	1/2	1/2	1/2	1/2
rs16980013	+18283	0.460	G/G	G/G	G/G	G/G	G/G	G/G	ND	G/G
rs10415988	+39031	0.500	2/2	2/2	2/2	1/2	1/2	1/2	1/2	1/2

*distance from the beginning of the *DMPK* gene. H (heterozygosity), N (null allele), NN (not known), ND (not determined).

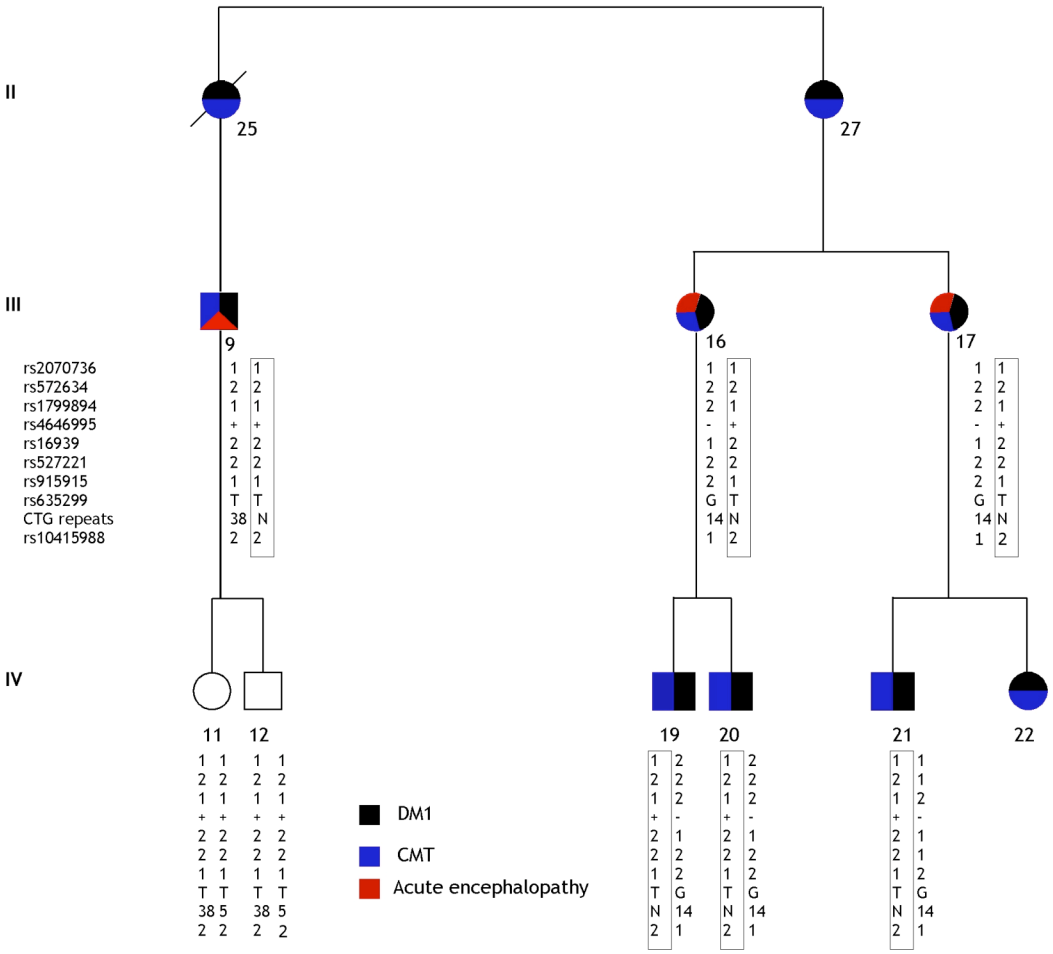


Figure 3-11. Possible haplotypes in the DM1+CMT++ family. The name of each SNP genotyped is indicated on the left and the alleles are indicated under each individual. DNA samples were only available for the individuals genotyped. III-9, IV-11 and IV-12 present as homozygous for all the SNPs, except for the number of repeats so only one haplotype was possible. In contrast, III-16, IV-19, IV-20, III-17 and IV-21 present as heterozygous in more than one SNP, so more than one haplotype could be constructed. The DM1 associated haplotype is boxed. N (null allele).

3.2.2.1 Analysing the integrity of the 3'-end up to intron 1 of the *SIX5* gene

To investigate the extent of any putative deletion event up to ~2.32 kb from the 3'-end of the repeat, the rs3745802 affected heterozygotes were genotyped with a series of forward primers located at different positions distal to *SIX5* and the reverse primer SIX-DR (3-11C). With primers SIX-J, SIX-K, SIX-L and SIX-C, the polymorphic site was assayed with *Sau*3AI, and heterozygosity was preserved. However, the direct identification of the polymorphic site was not possible with the primers BAB452 and DM-R because fragments of similar sizes were obtained from other *Sau*3AI restriction sites located in the amplicon. Performing a nested-PCR with

primers SIX-C and SIX-DR solved this problem. All individuals remained heterozygous using primers as distal as BAB452 (targeted to the end of the repeat array) (figure 3-12). These data provide evidence that the sequence within the first 2.3 kb at the 3'-end of the repeat tract is intact.

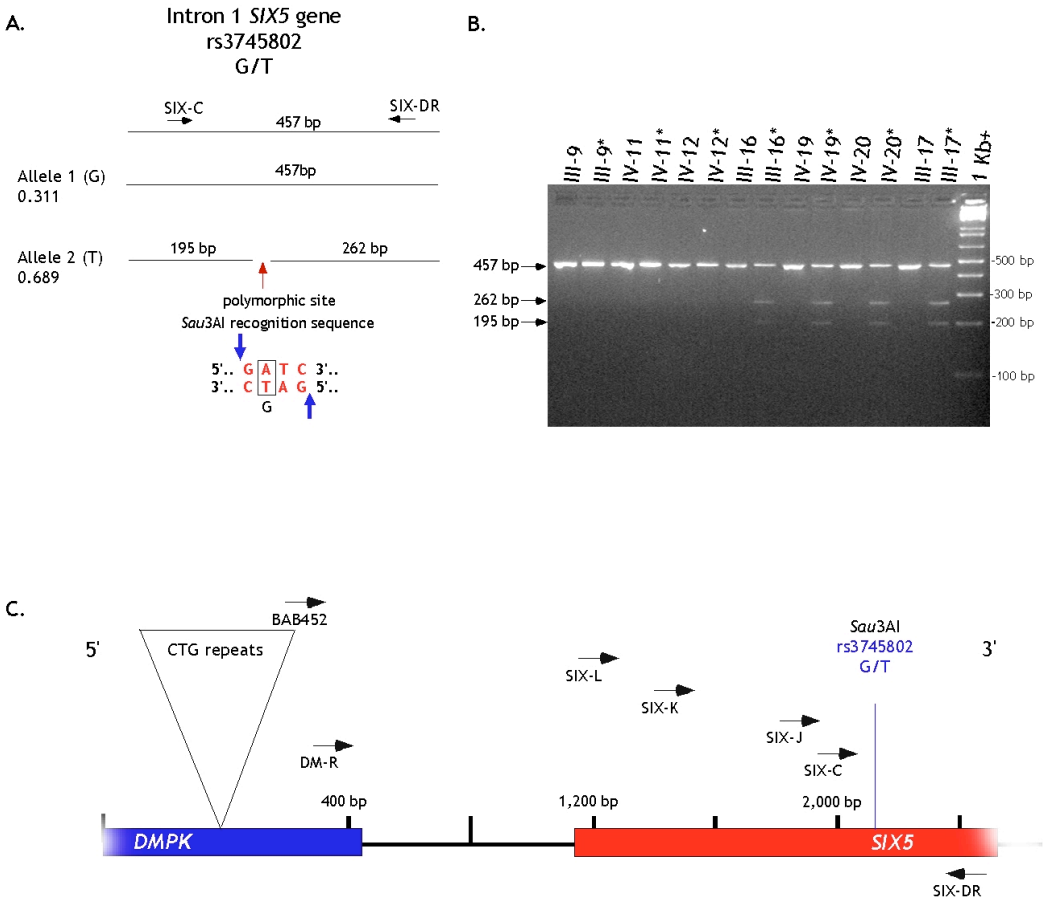


Figure 3-12. Investigating the 3'-end. A. Diagram of the amplified region showing the sizes of the PCR product and the location of the polymorphic site. B. The products were resolved in 2% Nusieve, 1% agarose gel. The size of the marker as well as the size of uncut and digested products are indicated. A (*) indicates the products were digested with *Sau*3AI. C. Schematic diagram of the *DMPK* region and flanking sequence at the 3'-end. The location of primers used to analyse the integrity of the 3'-end of the repeat tract up to ~2.32 Kb are indicated.

An alternative approach performed in parallel was to sequence the region at the 3'-end. The aims were to search for possible SNPs located before (rs3745802) and additionally to analyse the sequence at this end. PCR products generated with primers BAB452 and SIX-DR from III-16 and IV-19 were cloned into the vector pCR®4-TOPO® using the TOPO TA cloning kit for sequencing. III-16 and IV-19 were selected because both present as heterozygous at rs3745802 and it was possible to identify the different chromosomes. The clones obtained were screened by PCR following

digestion with *Sau*3AI. Clones carrying either the G or T allele were selected, plasmid DNA was extracted and sequenced (figure 3-13). The sequencing data revealed that the sequence contained within the location of the primers BAB452 and SIX-DR was intact in III-16 and IV-19, and no additional variants were observed.

It was clear from this data that a simple deletion event was not present in the DM1+CMT++ family. The next step was to determine whether the presence of a rearrangement at the 3'-end or an insertion within the $(CTG)_n$ could explain why an expanded allele was visualised by Southern blot. However, it was not possible to amplify by PCR and the 3'-end RP-PCR failed.

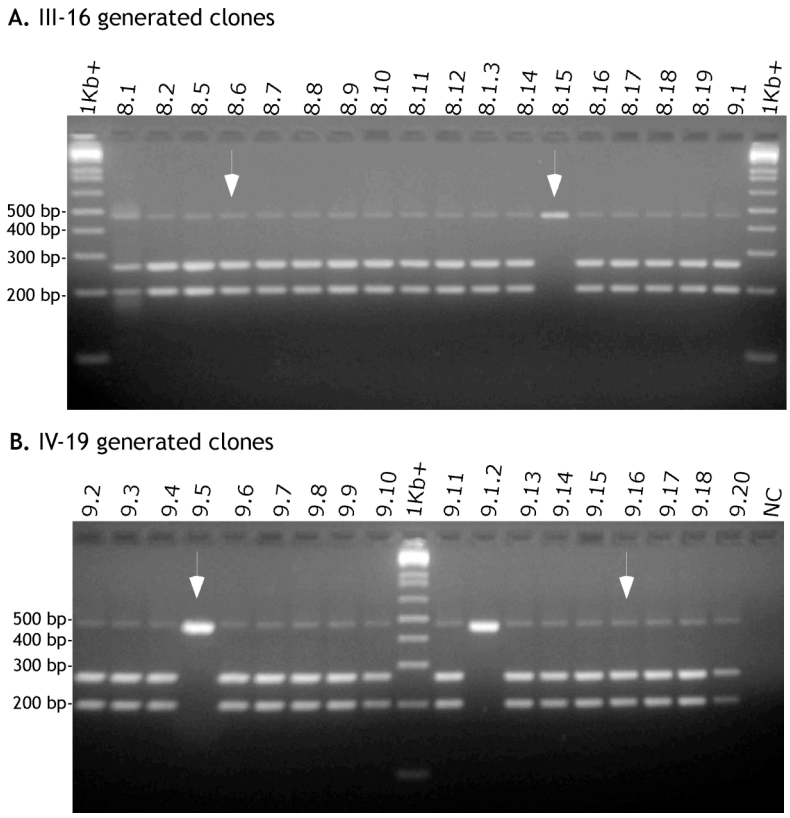


Figure 3-13. Cloning ~2Kb at the 3'-end of the $(CTG)_n$ tract. A. PCR products generated with BAB452 and SIX-DR were cloned. Screening was performed by PCR using the primers SIX-C and SIX-DR and *Sau*3AI digestion A. III-16 generated clones B. IV-19 generated clones. The white arrows indicate the clones sequenced, one *Sau*3A+ and one *Sau*3A- from each patient. NC (negative control).

3.2.3 Investigating a rearrangement at the 3'-end of the $(CTG)_n$

Vectorette PCR was selected to "walk" across the chromosome to clone a putative breakpoint at the 3'-end of the array (figure 3-14). DNA vectorette libraries were generated by digesting genomic DNA with a suitable restriction enzyme followed by the ligation of a linker containing unique PCR primer sites. PCR analysis was carried out using one specific primer annealing to the known region, and a vectorette primer identical to a part of the linker that will only prime after the first cycle. This strategy prevents the amplification of products with a linker at each end so that only fragments that are recognised by the specific primer will be amplified (Riley*, et al.* 1990; Arnold and Hodgson 1991). A high level of background was observed so, in an attempt to reduce the background, firstly an internal primer was designed (TAG3) followed by the design of another primer (TAG4). A slight reduction in background was obtained.

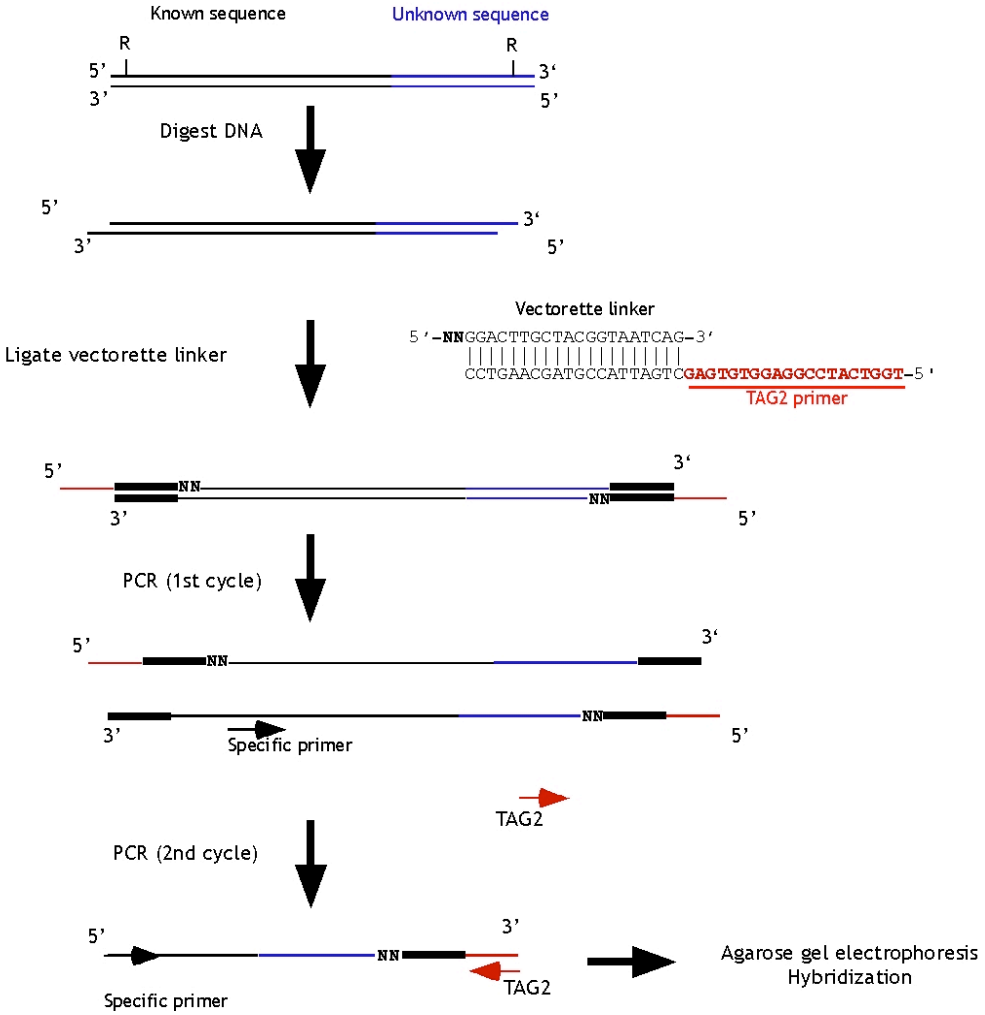


Figure 3-14. Schematic representation of vectorette PCR. Genomic DNA was digested with the enzyme of choice; equimolar amounts of the vectorette linker and the digested DNA were ligated to generate the vectorette libraries. The sequence of one of the vectorette linkers generated is indicated. The vectorette library was amplified with a primer targeted to a region within the known region and the vectorette primer (targeted to the vectorette linker). Two rounds of PCR were generally performed to obtain single bands.

Several restriction enzymes were used to generate vectorette libraries from III-9, III-16, IV-20, two unaffected controls and two classic DM1 patients. Both normal and expanded alleles were distinguished with the DM1 CTG repeat probe in all the vectorette libraries screened from the DM1 classic patients using either primers located at the 5'-end or 3'-end of the repeat tract. However, in the DM1+CMT++ patients, only the normal size allele could be identified in the vectorette libraries generated with a variety of four, five and six-cutter restriction enzymes: *Taq*I, *Sau*3AI, *Hin*fI, *Bam*HI, *Eco*RI, *Pst*I and *Mse*I (figure 3-15). Unexpectedly, the mutant DM1+CMT++ allele could not be amplified even with *Eco*RI or *Pst*I libraries. This result was not consistent with the previous data generated by Southern blot of restriction digested genomic DNA, where both enzymes generated

restriction fragments identified by the probe p5B1.4 (section 3.2.1.2) indicating the presence of an *Eco*RI or *Pst*I restriction site with ~ 200 bp of the 3'-end of the repeat.

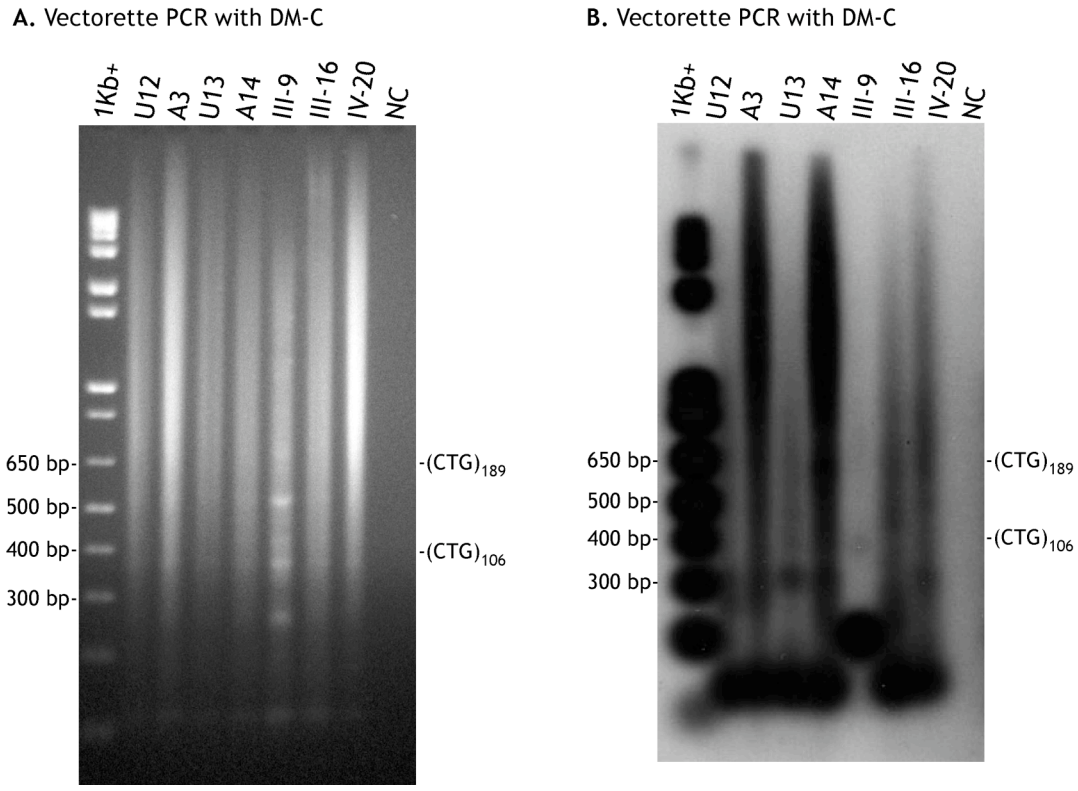


Figure 3-15. Vectorette PCR. PCR with primers DM-C and TAG-4 were performed on *Sau*3AI vectorette libraries. A. The products were resolved on a 1.5 % agarose gel and stained with ethidium bromide (A). The products were detected by DM1 CTG repeat probe by Southern blot hybridisation (B). Strong smears were detected with DM1 CTG repeat probe in two classic DM1 patients (A3 and A 14). However, smears similar to those obtained in unaffected controls (U12 and U13) were observed in three members of the DM1+CMT++ family (III-9, III-16 and IV-20). NC (negative control). The sizes of the marker are indicated on the left and the estimated number of repeats is indicated on the right.

*Aci*I and *Hha*I, enzymes with high GC rich recognition sequences, were selected for a further attempt. Finally, the mutant DM1+CMT++ allele could be detected when using a forward primer located at the 5'-end of the CTG repeat tract and TAG4. Smears with a length similar to the smears detected in the classic DM1 patient were observed with both libraries. Unexpectedly, the "38" repeat allele of III-9 was not visualised at the expected size in the *Aci*I vectorette library, suggesting that this allele must be digested with *Aci*I (figure 3-16C). The normal alleles, but not the expanded alleles were detected as discrete bands by ethidium bromide staining with *Hha*I vectorette libraries. Nevertheless, fragments of agarose located on top of the normal alleles from III-9, III-16 and IV-20 were cut and the DNA purified

(figure 3-16B). The DNA was cloned into the vector pCR®4-TOPO® using the TOPO TA cloning kit for sequencing. It should be noted that fragments of different sizes were purified together including non-specific products generated by vectorette PCR. To identify clones carrying the DM1 expanded allele and not other non-specific products, positive clones were detected by Southern blot hybridisation using the DM1 CTG repeat probe. Indeed, some of the inserts were not detected with the probe.

Only two clones from III-9 were detected with the DM1 CTG repeat probe and both were sequenced (figure 3-17). The sequence revealed that both ends (5'-end and 3'-end) of the repeat tract were intact and that vectorette PCR was working properly. The sequence of the repeat array consisted of 38 repeats with the structure $(CTG)_{5-6}$ $(CCGCTG)_{14}$ $(CTG)_{5-6}$, and the *Hha*I site was located at the 3'-end of the array as in the reference sequence. However, the sequence obtained probably corresponds to the normal size "38" repeat allele previously detected by PCR and not the mutant allele. These data explained why a band corresponding in size to the "38" repeat allele was not visualised in the *Aci*I library. This allele was digested with *Aci*I and only a small smear was amplified containing the 5'-end and 5 CTG repeats. The "38" repeat allele from patient III-9 was transmitted to his unaffected offspring (IV-11 and IV-12). To confirm the sequence of the allele in IV-11 and IV-12, the allele was amplified by PCR, cloned and sequenced. The "38" repeat allele in IV-11 and IV-12 was the same as the one present in III-9. Interestingly, a similar repeat structure has also been observed in a 37 repeat allele from a study of more than 300 sperm donor samples (Leeflang and Arnheim 1995).

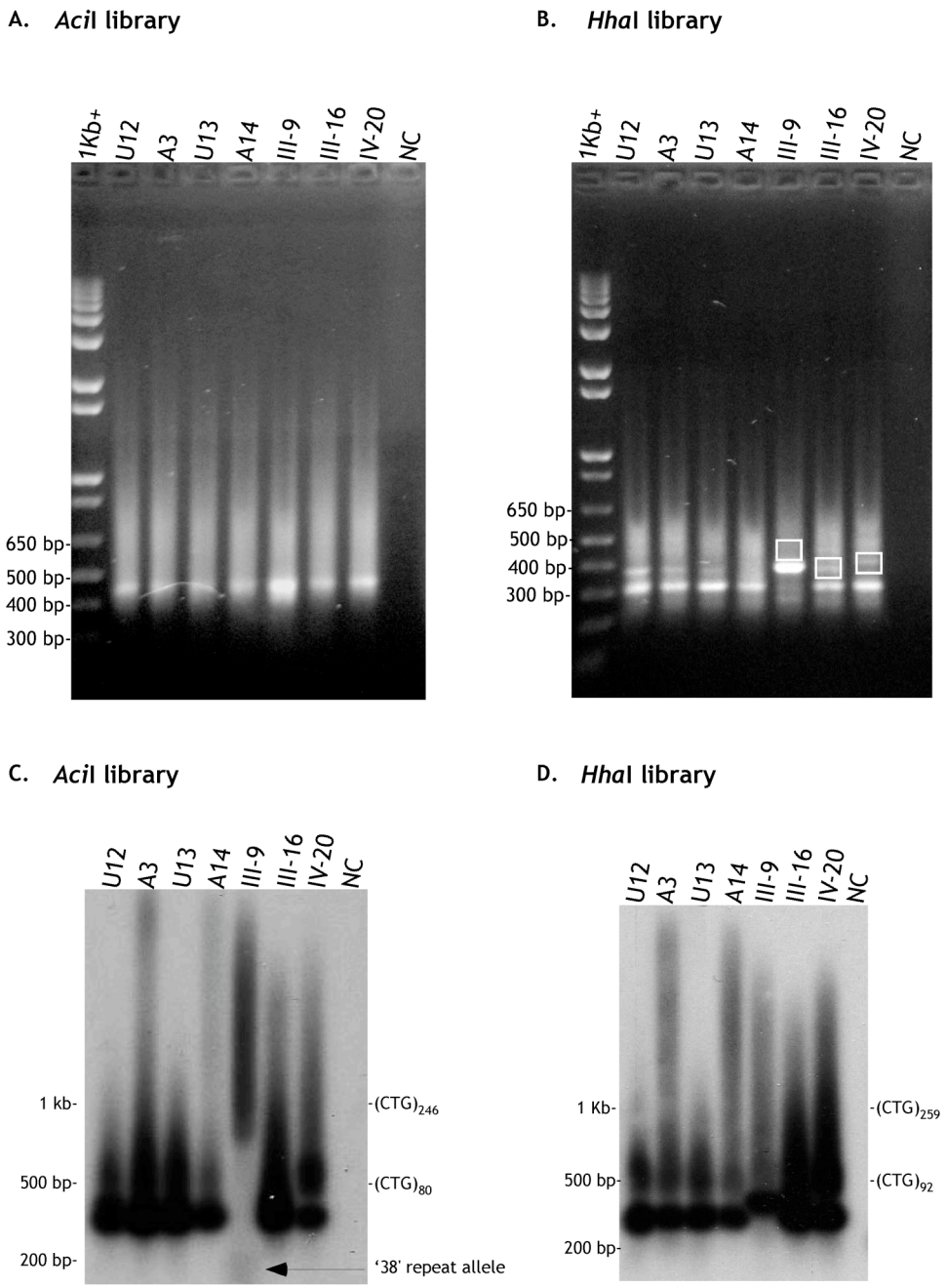


Figure 3-16. Vectorette PCR. A. and B. Hemi-nested PCR using a primer located at the 5'-end of the CTG repeat tract, DM-Cc with *Aci*I vectorette libraries and DM-C with *Hha*I libraries. The products were resolved on a 1.5% agarose gel and stained with ethidium bromide. C. and D. The gels were blotted by Southern 'squash' blot and then detected by [α-^{32}P] dCTP DM1 CTG probe. The sizes of the marker are indicated on the left and the estimated number of repeats is indicated on the right side. Although short smears, were observed in the two unaffected controls (U12 and U13), long strongly hybridising smears similar to those observed in the affected controls were presented in III-9, III-16 and IV-20. Strikingly, the "38" repeat allele of III-9 seems to be digested with *Aci*I. The boxes indicate the fragments of agarose gel from where DNA was purified and cloned. NC (negative control).

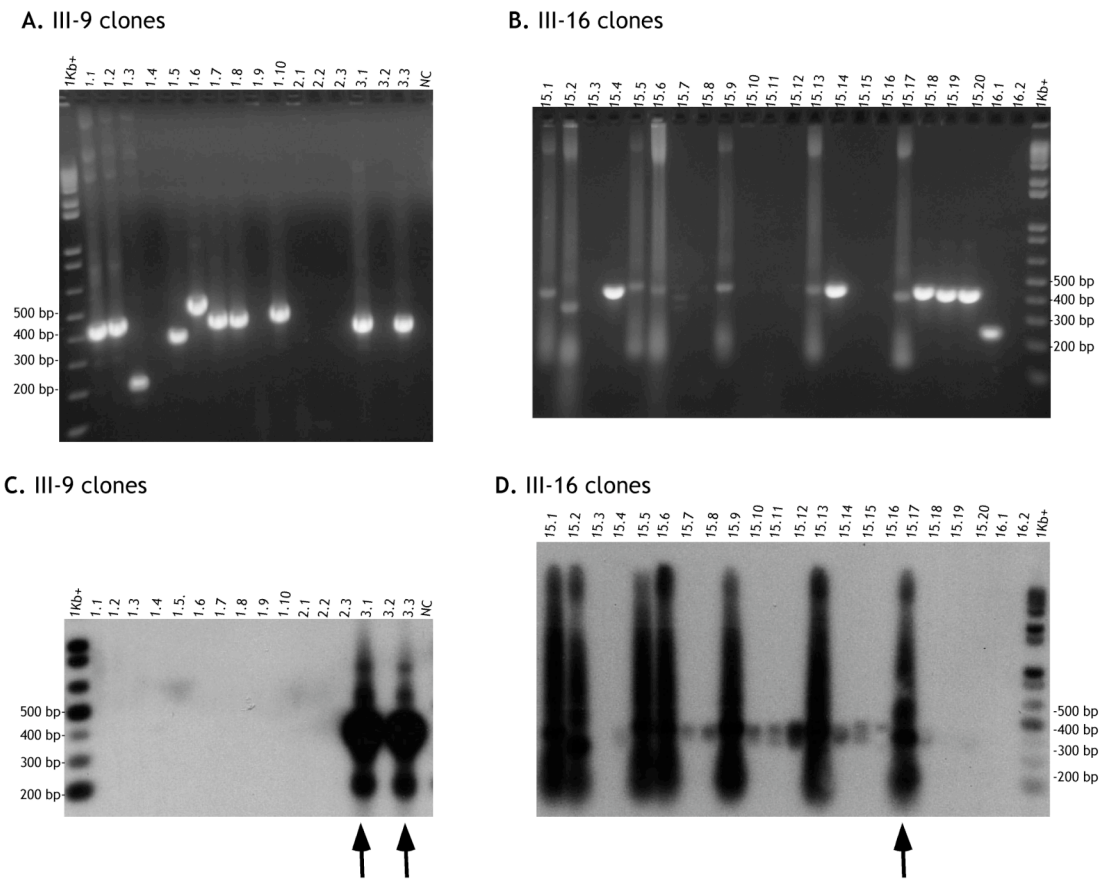


Figure 3-17. PCR screening of clones. PCR using primers DM-C and TAG4 was performed to evaluate the presence of the insert in the clones obtained. A. Clones obtained from vectorette products generated from III-9 vectorette libraries in A and from III-16 in B. C. and D. Auto-radiographs of gels A. and B. respectively. The arrows indicate the clones selected for sequencing. The sizes of the marker are indicated either on the left or the right side of each panel. NC (negative control).

The clones obtained from III-16 were comprised of fragments with a wide variety of sizes. Only one clone was sequenced just to determine whether a sequence could be obtained from it. A partial sequence was read revealing that it was comprised of $(CTG)_{31}$ $(CCGCTG)_1$ $(CTG)_{65}$ $(GGC)_2$ GC (appendices A and B showed examples of sequencing chromatograms). A new type of repeat was detected (GGC), and evidence for at least one *Hha*I site (GCGC) within the CTG array was confirmed. The flanking sequence at the 5'-end of the repeat tract was intact. It seemed reasonable to conclude that an increase of the GC content within the repeat array may be why a PCR product could not be generated by the standard procedure. In addition, it also potentially explained why the expanded allele could only be visualised from vectorette libraries generated with *Aci*I and *Hha*I libraries. Both restriction enzymes have sites within the CTG repeat tract in the DM1+CMT++ allele so a PCR product could be generated by a standard PCR because it was not necessary to amplify across the GC rich repeat tract.

Presumably, there are no sites for the other restriction enzymes used within the CTG repeat tract since no PCR products were obtained.

3.2.4 Investigating the presence of interruptions within the $(CTG)_n$ repeat by a modified PCR

Several additives can be added to a PCR to improve the amplification of high GC content regions. Dimethyl sulfoxide (DMSO) is a denaturing reagent that it is believed to decrease the melting temperature of the template DNA (Chakrabarti and Schutt 2002), and it was selected to be added to the PCR. Temperature gradient PCR was performed to obtain the annealing temperature to be used with each combination of primers selected (work performed by Mahajan Navdeep, MSc project student under my supervision).

Genomic DNA was amplified with primers DM-C and DM-DR in the presence of 10% DMSO. The expanded allele was visualised as a heterogeneous smear of fragments in all the affected members investigated with an estimated size between 200-400 triplets (figure 3-18). These data indicate that the molecular lesion in the DM1+CMT++ family seems to be an insertion of a GC-rich region into the CTG repeat array. Slight increases in the length of the repeat were detected in the offspring of III-16 and III-17 giving evidence for the presence of germline mosaicism. SP-PCR analysis is required to characterise the degree of somatic and germline mosaicism in detail. Characterisation of the interrupted CTG tract would be necessary to investigate the effect on somatic and germline mosaicism, and to compare it with classic DM1 patients with a pure CTG repeat tract.

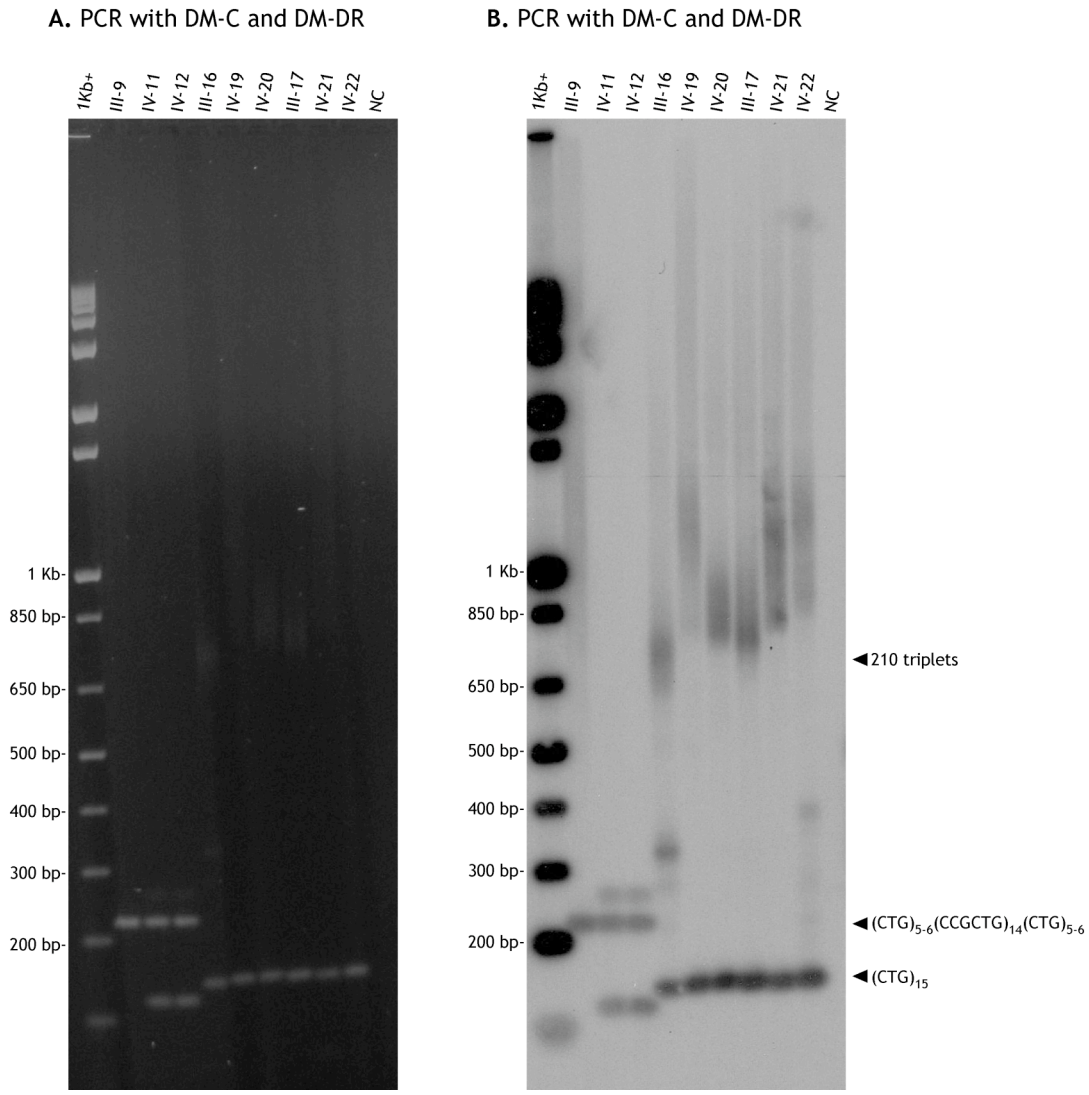


Figure 3-18. Genotype of the *DM1* locus by a modified PCR. A. Products of 28 cycle PCR using primers DM-C and DM-DR in the presence of 10% DMSO were resolved on 1.5% agarose gel and visualised on UV. B. The products were detected by Southern blot hybridisation. The size of the markers is indicated on the left and the estimated numbers of repeats is indicated on the right. An extra band is visualised above the band corresponding to the "38" repeat allele in IV-11 and IV-12, and probably corresponds to heteroduplex DNA. NC (negative control).

3.3 Discussion

A combination of molecular approaches was performed to identify the molecular lesion in the DM1+CMT++ family. Small expansions were visualised by Southern blot analysis of restriction digested genomic DNA, but it was not possible to amplify across the array using the standard approach. Further investigation of the flanking DNA by RP-PCR showed an expanded allele at the 5'-end but was negative at the 3'-end of the CTG repeat. As a result, it was postulated that either a deletion, a rearrangement or an insertion within the $(CTG)_n$ might explain the

molecular lesion in the DM1+CMT++ family. The most plausible explanation was thought to be a deletion at the 3'-end, which could disrupt one or more than one gene causing the complex symptomatology observed in the DM1+CMT++ family. Therefore, the investigation was mainly focused on finding a deletion at the 3'-end.

Several SNPs were genotyped in an attempt to determine the extent of a putative deletion, but it turned out that several individuals presented as heterozygous for many SNPs flanking the repeat providing evidence against a deletion. Similarly, the 3'-end of the CTG tract spanning a distance of ~2 Kb in III-16 and IV-19 was cloned and sequenced, but no differences between the normal allele and the DM1+CMT++ allele could be detected. Therefore, it was concluded that a deletion was not the molecular lesion in the DM1+CMT++ family.

In an effort to identify a breakpoint at the 3'-end, vectorette PCR was performed to walk across the unknown flanking DNA. Both normal and expanded alleles were clearly distinguished in classic DM1patients' vectorette libraries. However, only the normal allele could be identified in the DM1+CMT++ patients' vectorette libraries generated with a variety of four, five and six-cutter restriction enzymes (*Taq*I, *Sau*3AI, *Hin*fI, *Bam*HI, *Eco*RI, *Pst*I and *Mse*I). Unexpectedly, the mutant allele was not amplified from *Eco*RI libraries, even though a product was visualised by Southern blot analysis of restriction digested genomic DNA (section 3.2.1.2). However, the mutant DM1+CMT++ allele was finally detected as a smear with vectorette libraries generated with *Aci*I and *Hha*I (enzymes with high GC rich recognition sequences). The expanded alleles were cloned and sequenced. It was revealed that both ends (5'-end and 3'-end) of the repeat tract were intact and that vectorette PCR was working properly. One of the sequences revealed the structure of the "38" repeat allele of III-9, which consists of $(CTG)_5$ $(CCGCTG)_{14}$ $(CTG)_5$. Similar interrupted alleles were previously obtained (Leeflang and Arnheim, 1995; Monckton *et al* unpublished). Single sperm analysis showed that the instability of the "37" interrupted repeat allele was comparable to an allele of 27 repeats without interruptions, revealing a decrease of instability in the presence of

interruptions (Leeflang and Arnheim 1995). Indeed, the "38" repeat allele was transmitted unaltered from III-9 to both his daughter (IV-11) and son (IV-12). In addition, an incomplete sequence obtained from a clone from III-16 revealed that it consists of at least one hexamer (CCGCTG), $(GGC)_2$ and a *Hha*I site within the CTG tract. Subsequently, it was clear that the failure to PCR across the repeat tract might be due to an increase of the GC content within the CTG tract. By adding DMSO to the PCR, the mutant allele was detected and expansions between 200-300 triplet repeats similar to those obtained by *Eco*RI-digested DNA were visualised. Additional bands were previously obtained in IV-19 and III-17 by standard PCR (figure 3-2), and these bands were interpreted as heteroduplex DNA and/or contamination. However, the bands probably corresponded to the expanded allele that was not being efficiently amplified because of the high GC content. Only slight changes in the number of (CTG) were observed through transmissions (figure 3-18) giving molecular evidence for the absence of genetic anticipation observed between the third and fourth generations.

In conclusion, the molecular lesion in the DM1+CMT++ family seems to be the insertion of several interruptions, which increases the level of GC content in the region preventing the amplification by standard PCR. Even though the *ARCMT2B* locus or *PRX* or *DNM2* genes associated with three forms of CMT were not investigated, it would be extremely unlikely that 14 affected members carry an imperfect CTG allele together with a mutation in one of these loci. Further investigations would be necessary to identify the nature of the insertion before any conclusion could be drawn about the pathogenesis of the CTG repeat tract with insertions. Meanwhile, we could speculate about possible mechanisms a novel RNA gain-of-function and/or a novel effect on the downstream genes. The evidence obtained so far indicates that at least one GGC repeat is present. If several repeats are present a novel RNA gain-of- function might explain the CMT symptoms by analogy to what happens in fragile-X associated tremor/ataxia syndrome (FXTAS).

FXTAS is a neurodegenerative disorder that affects more than one third of the adult male carriers of premutation alleles (55<CGG<200) in the *FMR1* gene. The phenotype and the pathogenesis are completely different to fragile X syndrome (Hagerman, *et al*. 2001; Berry-Kravis, *et al*. 2003; Jacquemont, *et al*. 2003; Leehey, *et al*. 2003; Hagerman and Hagerman 2004). An RNA gain-of-function model similar to the one postulated for DM1 and DM2 seems to be the cause (Willemsen, *et al*. 2003). Ubiquitin-positive intranuclear inclusions containing *FMR1* mRNA were observed in neurons and astrocytes of FXTAS patients (Greco, *et al*. 2002; Tassone, *et al*. 2004). The protein composition of the intranuclear inclusions was investigated, and more than 20 different proteins were identified. Some of these proteins identified such as hnRNP A2, a number of neurofilaments and lamin A/C proteins have been associated with different forms of CMT. In addition, MBNL1, which is one of the RNA binding proteins interacting with the CUG repeat that is recruited in DM1 and DM2 nuclear foci was also found co-localised in the intranuclear inclusions (Iwahashi, *et al*. 2006). Therefore, if several CGG repeat are present inside the $(CTG)_n$ tract, one or several of these proteins could be recruited explaining the CMT symptoms observed in the DM1+CMT++ family.

4 Defining the imperfect CTG repeat allele in the DM1+CMT++ family and quantifying germ line and somatic instability

4.1 Introduction

Several factors have been shown to affect the stability of simple repeat loci. These factors are the sequence and the number of repeats, the presence of interrupting repeat units, the content of the flanking sequences, and *trans*-acting elements. Only some specific triplet repeat sequences are unstable. The most common motif is $(CNG)_n$, and this is probably because these repeats can form unusual stable DNA structures that may interfere with normal processes such as replication and repair leading to genetic instability (Chen, *et al*. 1998; Pearson and Sinden 1998; McMurray 1999; Sinden, *et al*. 2002). Shorter alleles are more stable than longer ones, and interrupted alleles are more stable than pure repeat tracts (Ashley and Warren 1995). High GC content in the flanking DNA has been associated with more unstable alleles (Brock, *et al*. 1999). However, it is still not clear how the level of GC content in the flanking DNA affects the instability of the repeats. Possible mechanisms involve alterations in the chromatin structure and/or methylation status. A number of proteins participating in DNA repair pathways were identified as *trans*-acting elements. The absence of *Msh2*, *Msh3* and *Pms2* (proteins involved in mismatch repair) has been shown to reduce the levels of instability in several mouse models (Manley, *et al*. 1999; van den Broek, *et al*. 2002; Savouret, *et al*. 2003; Wheeler, *et al*. 2003; Gomes-Pereira, *et al*. 2004b).

Normal interrupted repeat alleles are commonly observed at the *FRAXA* (Eichler, *et al*. 1994), *SCA1* (Chung, *et al*. 1993), *SCA2* (Choudhry, *et al*. 2001), *SCA3* (Kawaguchi, *et al*. 1994), *SCA8* (Moseley, *et al*. 2000; Silveira, *et al*. 2000), *SCA10* (Matsuura, *et al*. 2006), *SCA17* (Koide, *et al*. 1999; Nakamura, *et al*. 2001), *FRDA* (Cossee, *et al*. 1997; Montermini, *et al*. 1997) and *DM2* loci (Liquori, *et al*. 2001). Generally, expanded alleles have

lost the interruptions and become unstable. At the *SCA17* and *SCA2* loci, pure expanded CAG repeats were associated with germline instability, but uninterrupted alleles of equal size are stably transmitted through the generations (Gao, *et al*. 2008). At the *SCA1* locus, interrupted intermediate alleles are stable and non-pathogenic, in contrast to pure repeat alleles of similar sizes that are unstable and associated with the SCA1 phenotype (Chong, *et al*. 1995; Zuhlke, *et al*. 2002). In SCA2, intermediate size interrupted CAG alleles are associated with autosomal dominant parkinsonism, while pure alleles are associated with cerebellar ataxia (Charles, *et al*. 2007). Similarly, in SCA10, differences in the pathogenesis were observed between expanded uninterrupted and interrupted alleles. In addition, absence of anticipation was observed in a family carrying an interrupted SCA10 expanded allele. It was postulated that the sequence configuration of the repeat tract could be a disease modifier not only in SCA10 but also in other repeat diseases associated with an RNA gain-of-function (Matsuura, *et al*. 2006).

An imperfect CTG allele at the *DM1* locus was demonstrated to be the molecular lesion in the DM1+CMT++ family. This finding was unexpected since in myotonic dystrophy type 1 it was assumed that the expanded CTG repeat tract-causing disease was pure. Nevertheless, an imperfect CTG repeat tract within the normal size was found in a study of 383 sperm donors (Leeflang and Arnheim 1995) and Cockburn and colleagues presented data at the International Myotonic Dystrophy Consortium (IDMC4) of at least three patients with an interrupted CTG repeat allele. Kang and collaborators showed that interrupted CTG repeats are more stable than pure repeats in *E. coli* (Kang, *et al*. 1996). It is not known if the interruptions within the CTG repeat tract at the *DM1* locus will affect the instability of the tract and/or have an effect on the pathogenesis. It could be postulated that the CTG repeat at the *DM1* locus will probably show the same behaviour as the other types of repeats associated with diseases.

Here we present the combination of approaches performed in order to characterise the nature and extent of the interruptions in the DM1+CMT++ allele. The stability of the imperfect CTG repeat allele was investigated

and compared to the stability of CTG repeats in classic DM1 patients. In a preliminary attempt to investigate whether the interruptions modify the activity of the *DMPK* and/or *SIX5* gene, the methylation pattern of the repeat tract on the DM1+CMT++ allele was also investigated.

4.2 Results

4.2.1 Cloning and sequencing the imperfect CTG allele in the DM1+CMT++ cases

To identify the type and number of interruptions at the imperfect CTG allele in the DM1+CMT++ family, cloning and sequencing was performed. Genomic DNA from DM1+CMT++ cases was amplified with primers DM-A and DM-DR in the presence of 10% DMSO. The expanded alleles were visualised as smears with a size range from ~850 bp up to ~1.6 Kb (figure 4-1). DNA from the smears was purified with the Quiaquick gel extraction kit and cloned into the pCR®4-TOPO® vector using the TOPO TA cloning kit for sequencing. The kit has a very high efficiency, but only a few colonies were obtained. An increased efficiency was obtained by increasing the ligation reaction time and the amount of ligated-vector used to transform the cells. The clones obtained were screened by PCR. A heterogeneous population of fragments of different sizes was obtained due to the fact that the DNA was purified from a smear and not from a single band (figure 4-1). Unexpectedly, two main fragments were constantly obtained even within each clone: an expected large fragment containing ~213 triplets (~ 850 bp) and a smaller fragment of ~40 triplets (~330 bp). The smaller fragments probably correspond to large deletions generated by *E. coli* as Kang and colleagues previously observed. The authors showed that long pure CTG repeat tracts in bacteria are very unstable with a bias towards large deletions reducing the size of the repeats down to 20-40 repeats which are stable (Kang, *et al.* 1995).

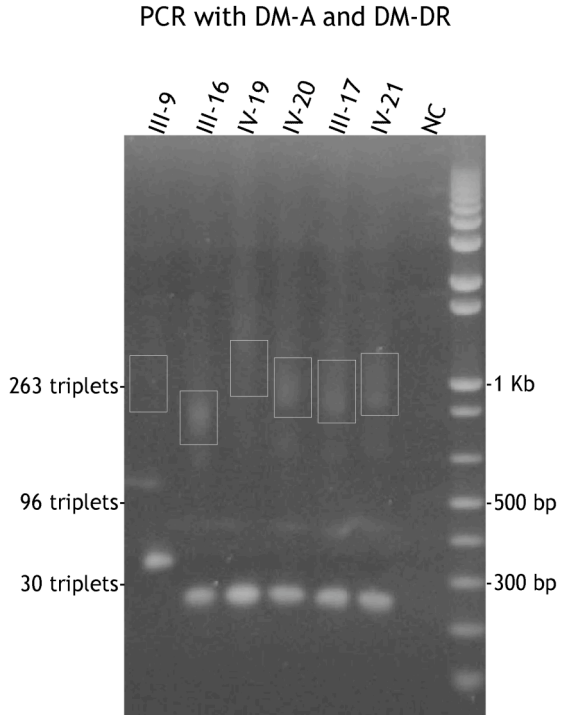


Figure 4-1. Amplifying expanded alleles by a modified PCR. Genomic DNA was amplified with primers DM-A and DM-DR in the presence of 10% DMSO. The products were resolved on a 1.5% agarose gel and detected with ethidium bromide staining. The expanded alleles were visualised as a smear. DNA from the smears (indicated with white boxes) were purified and cloned using the TOPO TA cloning kit for sequencing. NC (negative control). The position of the molecular weight marker is indicated on the right and the estimated number of triplets is indicated on the left.

Two clones, 2.17 (figure 4-2A) and 5.19 (data not shown), were selected to sequence with vector primers flanking the insert (T3 and T7F). In both clones, the flanking DNA at both end was intact. In 5.19, $(CTG)_{30}$ was read, which corresponded in size to the smaller insert. Only a partial sequence was obtained from 2.17 but, very interestingly this sequence contained two copies of a variant CCGCTG repeat:$(CTG)_{35}$ $(NNN)_?$ $(CCGCTG)_2$ $(CTG)_{35}$. The sequence beyond the CTG repeats was not possible to read from the 5'-end. Only two CCGCTG hexamers were read after the CTG repeats at the 3'-end. A complete sequence was probably not read because the sample contained a heterogeneous population of inserts and/or due to the complexity of the insert. DNA from the fragment containing ~213 triplets was purified from the gel (figure 4-2A), but a cleaner sequence could not be obtained.

In an attempt to obtain clones with a more homogeneous population of plasmids, a number of selected clones (containing the large insert) were streaked on LB plates. The population was still very heterogeneous in these primary subclones so a second round of single colony subcloning was

performed. Unfortunately, the fragment with ~213 triplets was constantly lost and only the fragment with ~40 triplets remained (figures 4-2B and C). Plasmid DNA from four secondary subclones was sequenced. Once again the flanking DNA at both ends was intact. One clone (indicated in figure 4-2C) contained just $(CTG)_{33}$. Two clones had the following structure: $(CTG)_{13}$ $(CCGCTG)_6$ $(CTG)_{34}$. In the last clone, 57 and 58 CTG repeats were read from both ends, but again the sequence beyond the CTG repeats was impossible to read. The sequence of these collapsed subclones appear to confirm the presence of variant CCGCTG hexamers within the array.

In an attempt to generate clones with longer inserts, the cloning was carried out using Max efficiency Stbl2 competent cells, which are more suitable for cloning some unstable inserts (Trinh 1994). These cells were previously used to clone an insert containing ~162 CTG repeats and flanking DNA of the *DMPK* gene. The yield was low but stable clones were generated (Colm Nestor personal communication). Nevertheless, the attempt in our case was unsuccessful; only one clone from 20 colonies investigated contained just the smaller fragment with ~40 triplets.

Another attempt at direct sequencing was performed with single molecule separated expanded alleles obtained by SP-PCR analysis. The use of single molecule separated alleles removes both the complicating factor of starting with a heterogeneous population of fragments and the cloning step. *Hin*dIII-digested genomic DNA from III-9, III-16, IV-19 and IV-20 was serially diluted and amplified with primers DM-A and DM-DR. The products were resolved on agarose gels and detected by Southern blot hybridisation with the DM1 CTG repeat probe. Single molecule separated alleles containing an expanded allele were selected for sequencing. PCR products from three single molecules (two from III-16 and one from IV-20) were sequenced directly with primers DM-A and DM-DR. However, once again a clear sequence from within the repeat array was not obtained, presumably because of sequence misalignment within the array caused by *Taq* polymerase mediated slippage events within the PCR product. These data highlight the technical difficulties in cloning and sequencing large repeat tracts. Nonetheless, the sequence data from the collapsed subclones

suggest the array is primarily comprised of CTG repeats with some CCGCTG hexamers.

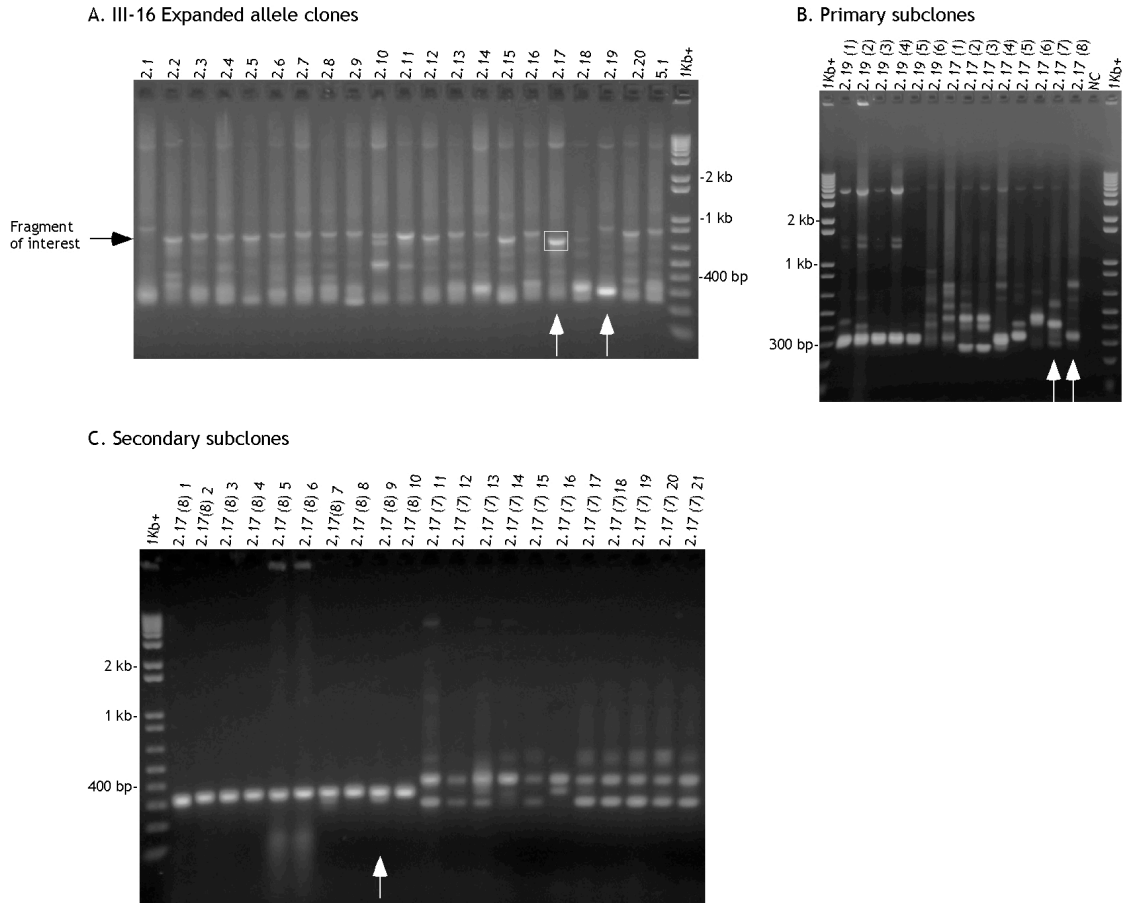


Figure 4-2. Representative data from PCR screening. A. Expanded alleles amplified with primers DM-A and DM-DR of III-16 were purified from the gel and cloned with TOPO TA cloning kit. The clones were screened by PCR using primers DM-A and DM-DR. The size of the expected insert is indicated with a black arrow. The white arrows indicate the clones selected for the primary subcloning. The white box indicates the DNA that was purified and sequenced. B. Screening of clones from the first streaking. The white arrows indicate the clones that were selected for the secondary subcloning. C. The repeats are quite stable but the correct insert was lost. The arrow indicates the clone that was selected for sequencing. Both ends of the repeat tract were intact and the clone contained $(CTG)_{33}$. NC (negative control). The sizes of the molecular weight markers are indicated either on the right or the left side of the panels.

4.2.2 Restriction site mapping

Vectorette PCR data and sequence data from vectorette library fragments (in section 3.2.3) revealed that the expansion is comprised of at least one *Hha*I site (GCGC), a number of hexamers (CCGCTG), and two (GGC). Restriction site mapping was performed in an attempt to define the array. Four different restriction enzymes were selected: *Hha*I, *Msp*A1I, *Aci*I and *Fnu*4HI. *Msp*A1I (CMGCKG) recognizes and cleaves the CCGCTG hexamers; *Aci*I (CCGC) recognizes and cleaves the CCGCTG hexamers as well as GGC;

and *Fnu*4HI (GCNGC) recognizes and digests CNG repeats. Diluted PCR products generated with primers DM-A and DM-DR were digested independently with one of the restriction enzymes of choice. The products were resolved on an agarose gel and detected with either the DM1 CTG repeat that detects fragments containing the CTG repeat tract and the flanking DNA, or a 5'-end probe that will only detect fragments containing the 5'-end flanking DNA. By sizing the generated fragments, the location of the different restriction sites, the size and type of interruptions could all be estimated.

Analysis of PCR products derived from the genomic DNA of III-16, IV-19 showed that *Fnu*4HI digested the repeat tract completely so the repeat tract must be comprised primarily of runs of CNG (figure 4-3). The expanded allele was digested with *Aci*I and *Msp*A1I and two products were obtained: a small band of ~170 bp (denominated PB) and a smear. PB was detected only with the DM1 CTG repeat probe so it must be comprised of the 3'-end flanking DNA (72 bp) plus ~35 CTG. The smear must contain the 5' flanking DNA plus a number of repeats that were not digested with either *Aci*I or *Msp*A1I. The sizes of the smears obtained with *Aci*I are smaller than the smears obtained with *Msp*A1I. The difference must be due to the presence of a series of CCGCM (M represents C or A) digested with *Aci*I but not with *Msp*A1I. More than one site of both *Aci*I and *Msp*A1I must be present because the size of PB and the smear is less than the size of the undigested expanded allele. *Hha*I digested the expanded allele into a band of ~330 bp (denominated PA) and a smear. The PA band was only detected with the DM1 CTG repeat probe indicating that it must contain just the 3' flanking DNA, whereas the 5' flanking DNA must be in the smear that is detected with both probes. There is only one *Hha*I site because if we add the size of the smear plus the size of PA, the size of the undigested expanded allele is reproduced. By subtracting the size of the 3'-flanking DNA (72 bp) from the size of PA, we can postulate that the *Hha*I site is located at ~258 bp. Similar data including PA and PB was obtained from samples III-9, IV-20, III-17, IV-21 and IV-22 (data not shown), suggesting that the structure of the 3'-end of the repeat tract seemed to be very similar in all the DM1+CMT++ cases. In contrast, the 5'-end of the repeat

tract appeared to be the variable part. Therefore, instead of obtaining discrete bands, a smear was visualised revealing the presence of somatic mosaicism.

Somatic mosaicism gives rise to variant allele length within the samples complicating the estimation of the location of the restriction sites, and consequently the structure of the interruptions could not be more accurately postulated.

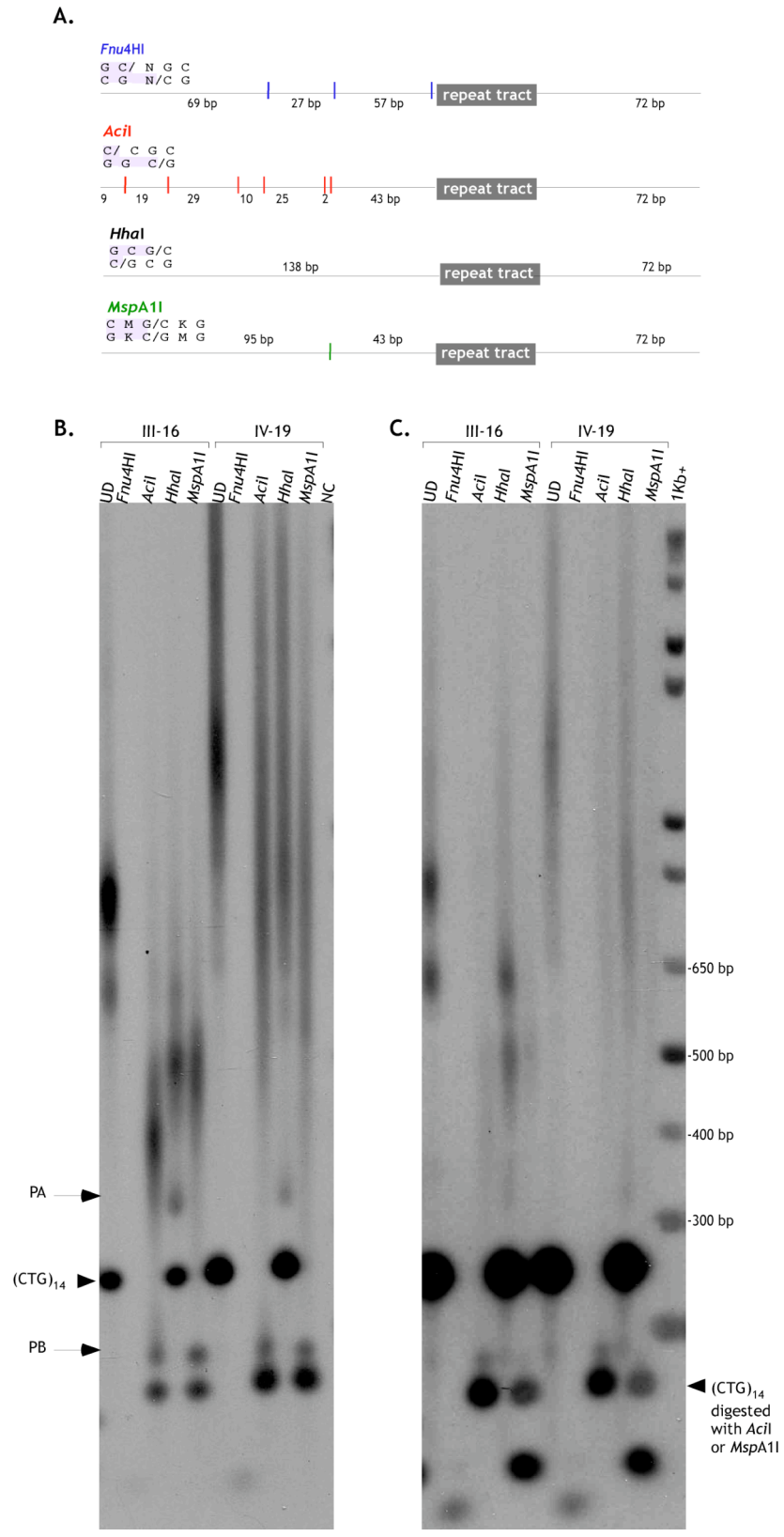


Figure 4-3. Restriction site mapping. A. Diagram illustrating the location of the restriction sites of each enzyme in the flanking DNA of a PCR product generated with the primers DM-A and DM-DR. There are no *Hha*I restriction sites in the flanking DNA. K (represents G or T), M (represents A or C). B. Representative autoradiographs. Genomic DNA was amplified with DM-A and DM-DR, following restriction digest. The products were resolved on a 1.5% agarose gel detected with the DM1 CTG probe (B) or with the 5'-end probe (C). UD (undigested product), NC (negative control). The size of the marker is indicated on the right side. PA band detected constantly with *Hha*I digestion, PB band detected constantly with *Aci*I and *Msp*A1I digestions.

4.2.2.1 Analysing single molecule separated alleles

To decrease the complexity of the analysis, single molecule separated alleles containing expanded alleles from III-9, III-16, IV-19 and IV-20 were obtained by SP-PCR and assessed by restriction mapping following the procedure described previously. The analysis reproduced the data generated with bulk genomic DNA, but mainly single bands were obtained instead of the smears, except PA which remained as a smeary blob (figure 4-4). An estimation of the number and types of repeats was made possible by sizing the fragments (table 4-1). The number of CCGCM repeats was estimated by the difference between the top fragment from *Aci*I and *Msp*A1I. By adding the size of the top *Msp*A1 fragment and the size of PB minus the size of the undigested product, we estimated the number of CCGCKG repeats. The number of CTG repeats at the 5'-end was estimated by subtracting the flanking DNA (43 bp) from the *Aci*I top fragment. The number of CTG repeats at the 3'-end was estimated by subtracting the flanking DNA (72 bp) from the size of PB (table 4-1). The structure of the expanded allele of III-16 was estimated to consist of $(CTG)_{120\text{-}140}$ $(CCGCM)_{21\text{-}25}$ $(CCGCKG)_{29\text{-}32}$ $(CTG)_{38}$. Similar data was obtained with single molecule separated alleles of III-9, IV-19 and IV-20 (table 4-1). The difference between the single molecule separated alleles was the number of CTG at the 5'-end. Slight changes in the numbers of CCGCM (± 6 repeats), CCGCKG (± 6 repeats) and the 3'-end CTG (± 6 repeats) were obtained in different single molecule separated alleles, both within the same patient and between different patients. The length of 9C (single molecule separated allele from III-9) was more than 2 Kb, so the resolution of the digested products on 1.5% agarose gel was not as good as the other alleles. Therefore, it was not considered in the comparisons.

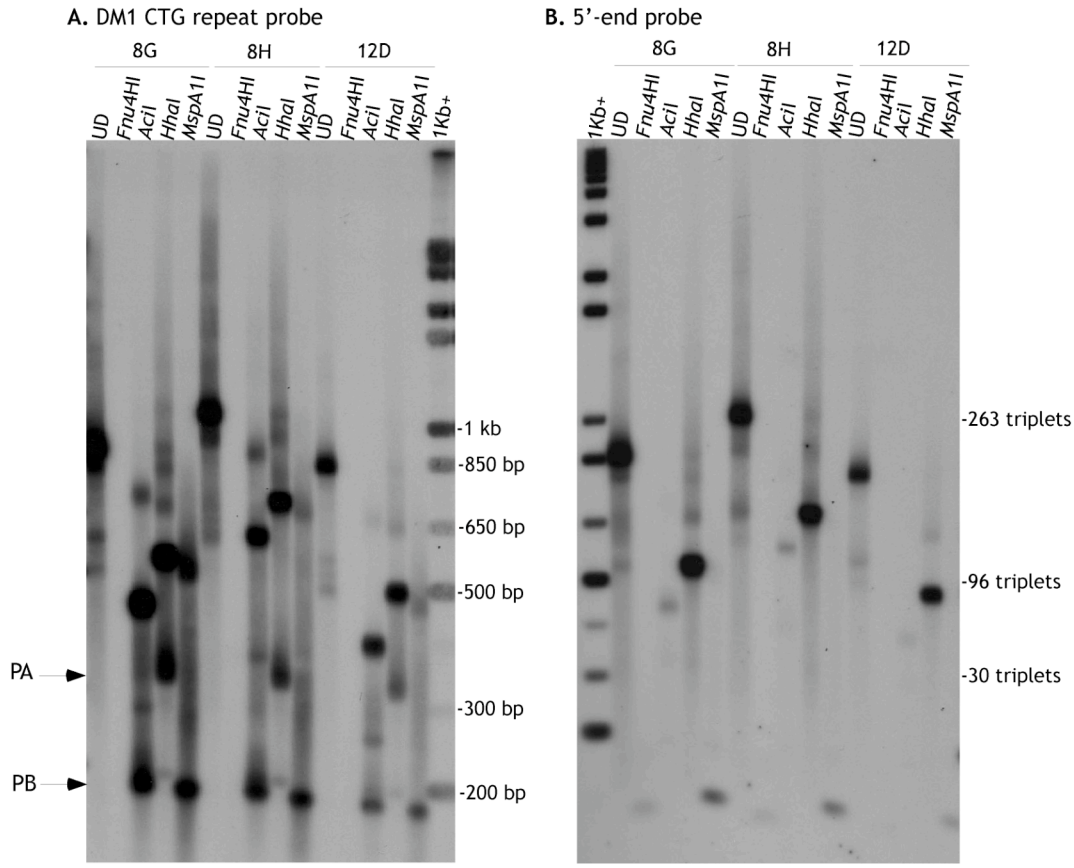


Figure 4-4. Restriction site mapping using single molecule separated alleles. Three single molecule separated alleles of III-16 obtained by SP-PCR were investigated. Diluted PCR products generated with primers DM-A and DM-DR were digested independently with: *Fnu*4HI, *Aci*I, *Hha*I and *Msp*A1I. The products were resolved on a 1.5% agarose gel and detected with either the DM1 CTG probe (A) or the 5'-end probe (B). The size of the marker and the estimated number of triplets are indicated. PA product was generated with *Hha*I, PB product was generated with *Aci*I and *Msp*A1I.

Table 4-1. Sizing the different type of repeats in single molecule separated alleles

	III-16 8G	III-16 8H	III-16 12D	III-9 9D	III-9 9C	IV-19 11E	IV-19 10A	IV-20 6C
5' CTG	121	140	133	395	617*	267	169	147
CCGCM	25	23	21	18	17*	16	12	20
CCGCKG	29	35	32	26	65*	37	35	28
3' CTG	38	38	38	35	35*	36	30	36

*Estimates might not be accurate. The length of the digested products from 9C was more than 1.65 Kb so the resolution on a 1.5% agarose gel was very low.

4.2.3 RP-PCR

4.2.3.1 Detecting CCGCTG

To investigate if runs of the CCGCTG hexamer are located within the CTG repeat tract, CCGCTG repeat specific primers were designed to perform RP-PCR (figure 4-5). The CCGCTG repeat specific primer contains three CCGCTG hexamers, and the CTG repeat specific primer contains five CTG repeats. If the primers were highly specific, the ladders obtained with CTG and CCGCTG primers should have different sizes and partially complementary gaps should be expected in the ladders. The "38" repeat allele of IV-11 and IV-12 has previously been confirmed by sequencing to contain 14 copies of the CCGCTG hexamer, so these alleles were used as positive controls. A ladder with a size corresponding to 14 CCGCTG hexamers was obtained in IV-11 and IV-12 from both ends with the CCGCTG repeat specific primers (figure 4-6). Two discrete bands were obtained with the CTG repeat specific primers. The gap between those two bands coincides with the CCGCTG hexamer smear. These data indicated that each primer detected each type of repeat with a high specificity.

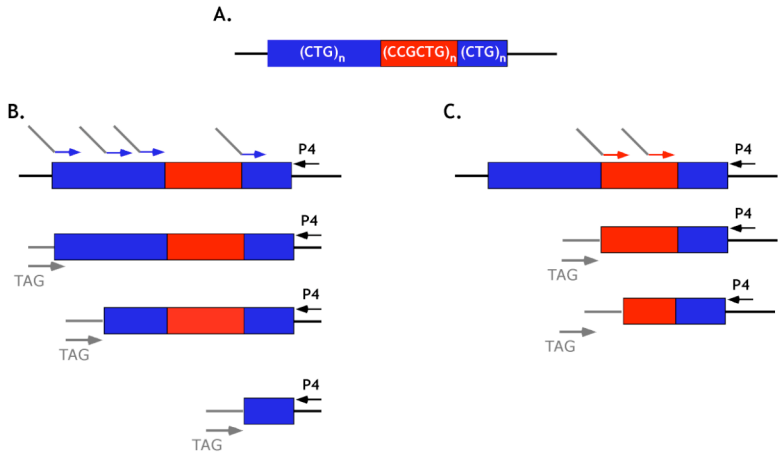


Figure 4-5. 3'-end RP-PCR designed to detect two different types of repeats. A. Possible structure of the DM1+CMT++ allele. Assuming there is an insertion of a number of copies of CCGCTG hexamers within the CTG repeat tract, two different types of primers could be used: one targeted to the CTG and another targeted to the CCGCTG. B. Products generated with the CTG repeat specific primer. C. Products generated with the CCGCTG repeat specific primer.

A long run of CTG repeats were visualised at the 5'-end in III-9, III-16, IV-19, IV-20, III-17 and IV-21, but only small faint smears around 650-850 bp were obtained with the CCGCTG specific primers. At the 3'-end, a smear was

obtained with a size between 200-300 bp, detected in all cases with the CCGCTG specific primer. Two ladders with a gap in the middle were detected with the CTG repeat specific primer (figure 4-6). Most of the gap complemented the ladder detected with the CCGCTG repeat specific primer except the top, indicating that another type of repeat must be present. By sizing the bottom and the top part of the ladders, the number of repeats was estimated. The DM1+CMT++ allele was estimated to comprise of $(CTG)_X$ $(NNN)_x(CCGCTG)_{14}$ $(CTG)_{35}$. The CTG repeats at the 5'-end seemed to be the variable part of the repeat tract.

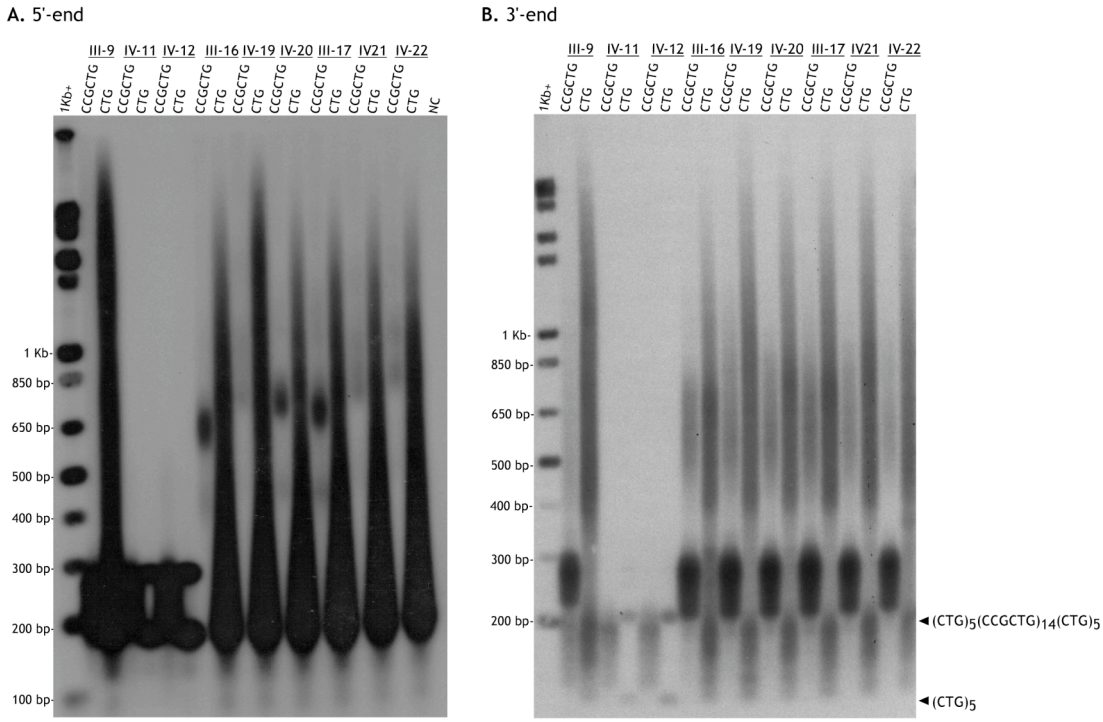


Figure 4-6. RP-PCR. A. DNA was amplified by RP-PCR using primers DM-A and primers TAG-CAGCGG (to detect the CCGCTG hexamer) or TAG-AGC (to detect CTG). The products were resolved on a 1.5% agarose gel and detected with the 5'-end probe. B. DNA was amplified by RP-PCR using primers DM-DR and primers TAG-CCGCTG (to detect the CCGCTG hexamer) or TAG-GTC (to detect CTG). The RP-PCR products were resolved on a 1.5% agarose gel and detected with the DM1 CTG repeat probe. The estimated sizes of triplets are indicated on the right. Note a gap between 300-400 bp where a ladder was not detected with either the CTG repeat specific primer or the CCGCTG repeat specific primer, so other type of repeats should be present. NC (normal control).

Given that small products were generated by RP-PCR at the 3'-end with the CCGCTG specific repeat primer, it appeared logical to clone and sequence those fragments. The DNA from the top part of the CCGCTG hexamer RP-PCR ladder was purified from an agarose gel and cloned into pCR®4-TOPO® using a TOPO TA cloning kit for sequencing. Relatively stable clones were obtained and plasmid DNA from two clones for each sample was sequenced.

Slight differences were observed between the sequences of the clones corresponding to the same patient. Nevertheless, the sequences confirmed the estimated sequence previously postulated by sizing the smears $(CCGCTG)_{10\text{-}15}$ $(CTG)_{35}$ (data generated by Navdeep Mahajan MSc student project under my supervision).

4.2.3.2 Cloning the gap

A gap between 300 bp to 400 bp was detected in 3'-end RP-PCR when using the CTG repeat specific primer, suggesting that another type of repeat must be present in the DM1+CMT++ allele. GGC repeats were observed by sequencing vectorette clones, so it could be possible that the gap corresponds to the presence of GGC repeats. In order to investigate the presence of GGC by RP-PCR, specific primers were designed containing five GGC repeats. Unexpectedly, a ladder for the variant GGC repeats was not obtained, but in contrast the complementary repeat CCG was detected. Ladders at the 5'-end and 3'-end were detected in III-9, III-16, IV-19 and IV-20 using CCG specific primers; as expected IV-11 and IV-12 were negative (figure 4-7). The ladder coincides with the gap previously detected with the CTG repeat specific primer. By sizing the top and bottom part of the ladder, the number of CCG repeats was estimated. The sequence of the repeat seemed to be comprised of $(CTG)_x$ $(CCG)_{20}$ $(CCGCTG)_{14}$ $(CTG)_{35}$. Similar data was obtained with III-17, IV-21 and IV-22 (data not shown).

DNA from the top part of the RP-PCR ladder obtained with the CCG repeat specific primer at the 3'-end was purified cloned and sequenced. Four clones were sequenced, even though slight differences between the different clones were observed, the sequence confirmed the estimated sequence. In addition, DNA from the top part of the RP-PCR ladder at the 5'-end with the CCG repeat specific primer was also purified, cloned and sequenced. One clone was sequenced and the following structure $(CTG)_X$ $(GGC)_3$ G $(CCG)_5$ $(CCA)_1$ $(CCG)_7$ was read. The *Hha*I site within the CTG tract was revealed, together with GGC and CCG. Unexpectedly, a new type of CCA repeat was also observed. The presence of this novel variant could be either a mutation generated during PCR or by *E. coli* or even a sequencing

error, and so it should be verified. The GGC repeats were not detected by RP-PCR but this could be because the primer contained five repeats and the alleles appear to contain only three repeats.

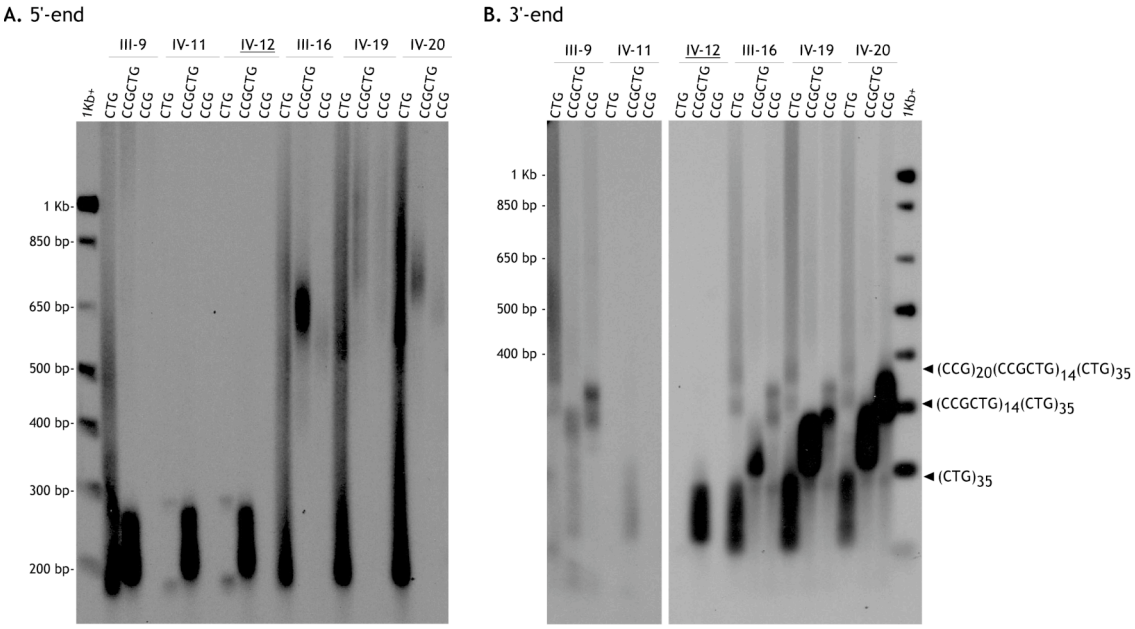


Figure 4-7. Representative RP-PCR autoradiographs. The specific primer was targeted to the 5'-end (A) or to the 3'-end (B). CTG (CTG repeat specific primer), CCGCTG (CCGCTG repeat specific primer) and CCG (CCG repeat specific primer). The products were run on a 1.5% agarose gel and detected with either the 5'-end probe (A) or the DM1 CTG repeat probe (B). Estimated sequences of the repeat tract are indicated on the right and the position of the molecular weight marker is indicated on the left.

4.2.3.3 Single molecule separated alleles investigated by RP-PCR

Single molecule separated alleles were used to test which repeat was responsible for somatic mosaicism observed when amplifying bulk genomic DNA. Ladders with different lengths located at different positions were visualised. By sizing the top and bottom part of each smear, the number of each type of repeat was estimated. The sequence of the repeat seemed to be comprised of $(CTG)_x$ $(CCG)_{20}$ $(CCGCTG)_{14}$ $(CTG)_{35}$. The numbers of CCG, CCGCTG and CTG at the 3'-end seemed to be very similar between single molecule separated alleles containing the expanded allele of the same patient and between different patients. In contrast, the CTG tract at the 5'-end seemed to be the variable part of the array (figure 4-8). This data indicates that the different number of CTG repeats at the 5'-end causes the somatic mosaicism detected in each patient.

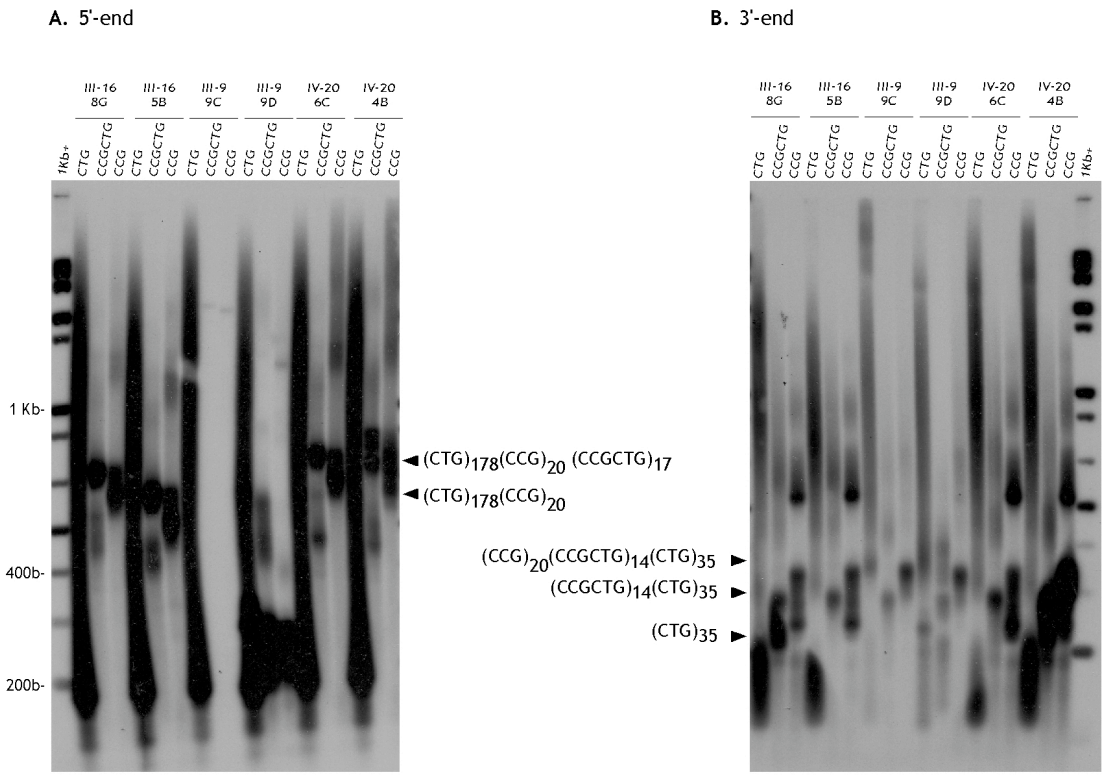


Figure 4-8. Representative RP-PCR autoradiographs. Single molecule separated alleles were investigated by RP-PCR. The specific primer was targeted to the 5'-end (A) or to the 3'-end (B). CTG (CTG repeat specific primer), CCGCTG (CCGCTG repeat specific primer) and CCG (CCG repeat specific primer). The products were run on a 1.5% agarose gel and detected with either the 5'-end probe (A) or the DM1 CTG repeat probe (B). Estimated sequences of the repeat tract are indicated. The size of the molecular weight marker is indicated on the left.

4.2.4 Amplifying the 3'-end of the DM1+CMT++ allele by PCR

Three GGC triplets seemed to be at the end of the CTG repeat at the 5'-end. The primer DM-GGC (TG CTG CTG CTG CTG GGC GGC G) was designed to detect the boundary. DM-GGC was used in combination with DM-DR in an effort to amplify the 3'-end of the repeat tract, which seemed to be similar in all the patients. Single molecule separated alleles containing expanded alleles were amplified with primers DM-GGC and DM-DR and a product of ~350 bp was constantly obtained. The products were digested with *Hha*I and a single site was detected in all the single molecules investigated. The DNA from the bands was purified, cloned and sequenced. The sequence revealed that the product consisted of: GCG $(CCG)_{20}$ $(CCGCTG)_{14}$ $(CTG)_{35}$ (worked performed by Rhoda Stefanatos, undergraduate student project under my supervision).

Slight differences in the sequence obtained from the independent single expanded alleles were obtained. To test if these differences were real or cloning errors, genomic DNA from the DM1+CMT++ cases was amplified with DM-GGC and DM-DR and resolved by agarose gel electrophoresis. Two rounds of PCR were needed to detect amplification products. The first round was performed with primers DM-C and DM-DR in the presence of 10% DMSO. A hemi-nested PCR was performed with primers DM-GGC and DM-DR using 100-fold dilution of the first PCR. The size of the band was estimated in each patient. The average size of the band was 344 bp ± 5 bp (figure 4-9). These data along with the results described above, indicated that the 3'-end of the repeat tract is the same in all the DM1+CMT++ cases investigated. Therefore, the 3'-end was transmitted unmodified from the third to the fourth generation. The only difference between the DM1+CMT++ alleles was the number of CTG repeats at the 5'-end, which is also responsible for somatic mosaicism present in the patients.

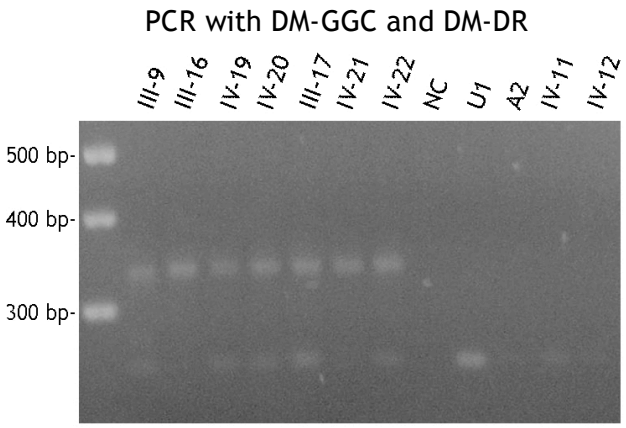


Figure 4-9. Amplifying the 3'-end of the repeat tract. Genomic DNA was amplified with primers DM-C and DM-DR in the presence of 10% DMSO. Hemi-nested PCR was performed with primers DM-GGC and DM-DR using 100-fold dilution of the primary PCR reaction. The products were resolved on a 2% nusieve, 1% agarose gel. NC (negative control), U1 (normal individual), A2 (classic DM1 patient). The size of the marker is indicated on the left. The bottom band is a non-specific band that was not detected by Southern blot hybridisation with the DM1 CTG repeat probe.

In summary, the combination of cloning and sequencing, RP-PCR, restriction mapping and the PCR performed with the DM-GGC primer were necessary to obtain the best-estimated structure of the DM1+CMT++ allele. The allele is comprised of three novel repeat types and a *Hha*I site (figure 4-10). The CTG repeats at the 5'-end are unstable and responsible for both somatic and germ line variation. In contrast, the interrupted 3'-end is somatically and intergenerationally stable.

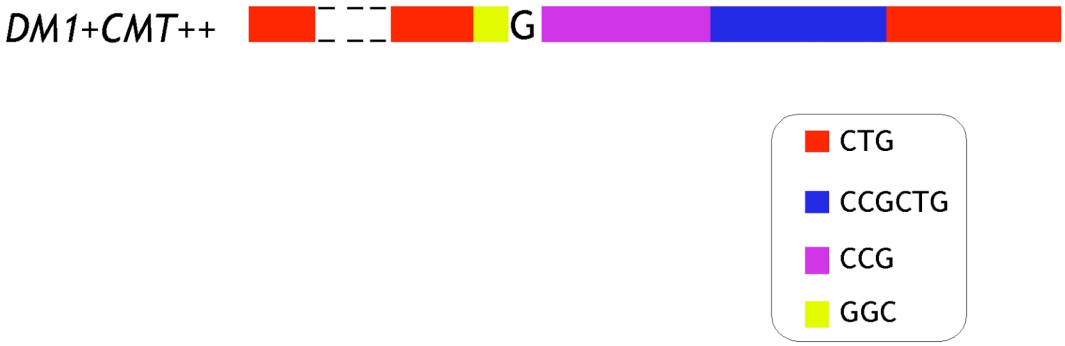


Figure 4-10. Structure of the DM1+CMT++ allele. The best estimated sequence of the imperfect CTG repeat tract is shown. The variable tract of the structure is the CTG run at the 5'-end, followed by $(GGC)_3$ G $(CCG)_{20}$ $(CCGCTG)_{14}$ $(CTG)_{35}$.

4.2.5 Investigating the dynamics of the DM1+CMT++ allele by SP-PCR

The expanded CTG repeat at the *DM1* locus has been demonstrated to be somatically and intergenerationally unstable. The somatic instability is expansion-biased, age-dependant and tissue specific (Monckton, *et al*.

1995a Martorell, 1998; Wong, *et al*. 1995 Martorell, 1998). Given the stabilising effect of variant repeats at other loci, we thought it likely that the presence of interruptions could affect the stability of the CTG repeat.

4.2.5.1 Analysing the degree of somatic mosaicism

To investigate in detail the dynamics of the DM1+CMT++ allele, SP-PCR was performed with DNA derived from blood cells. SP-PCR has been demonstrated to be a powerful technique to resolve and amplify products derived from single molecules, allowing a detailed analysis of repeat length variation to be performed (Monckton, *et al*. 1995).

*Hin*dIII-digested genomic DNA was serially diluted and amplified with primers DM-A and DM-DR. The products were resolved on 1.5% agarose gels and detected with the DM1 CTG repeat probe. The progenitor allele (inherited allele transmitted by the father or the mother) was estimated as the lower boundary of the distribution when amplifying ~150 genomic equivalents. To measure the somatic mosaicism, an average of 3 genomic equivalents per reaction were amplified and more than 100 molecules were sized for each patient (figure 4-11). The degree of somatic mosaicism was estimated as the difference between the 90^{th} percentile and the 10^{th} percentile in the distribution (figure 4-12). The CTG repeats at the 5'-end of the DM1+CMT++ allele was somatically unstable and biased toward expansions; alleles of different sizes were visualised in each DM1+CMT++ patient.

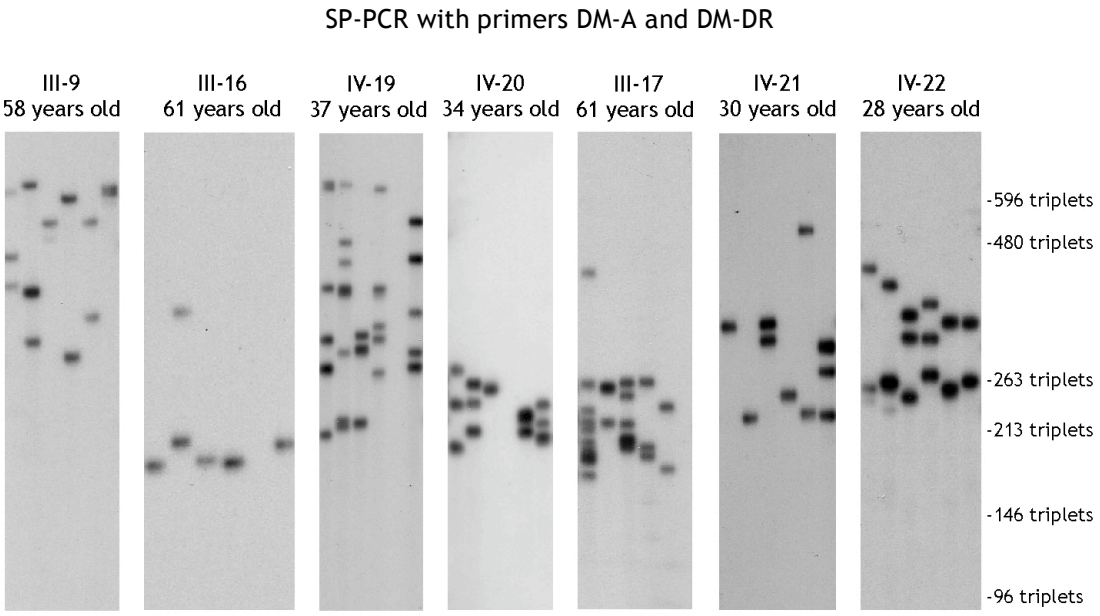


Figure 4-11. Repeat length variation in blood cells from the DM1+CMT++ cases. SP-PCR was used to assess the repeat length variation. Representative SP-PCR autoradiographs derived from ~3 genomic equivalents. The scale on the right displays the molecular weight marker converted into number of triplets. The age at sampling is indicated above each panel.

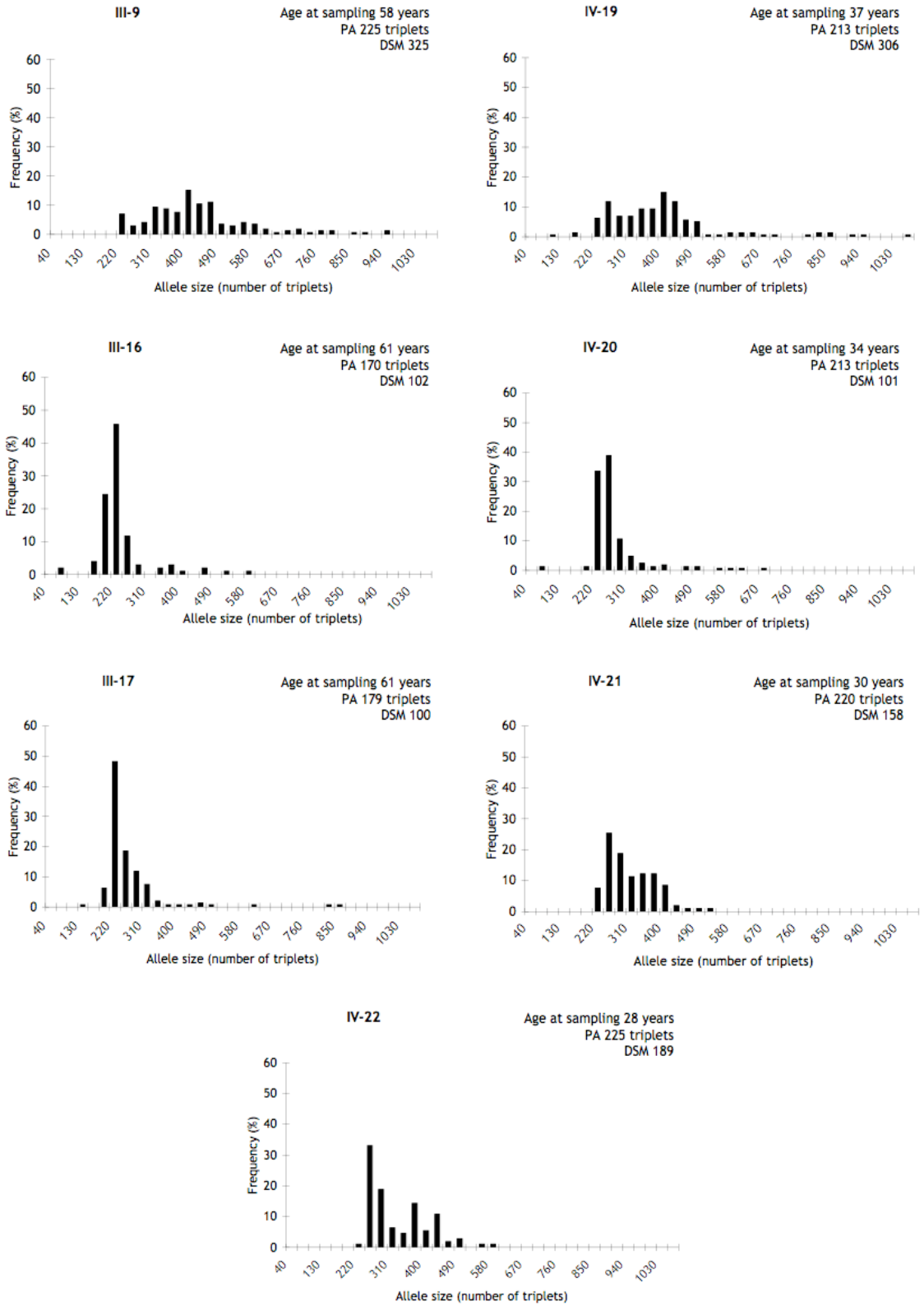


Figure 4-12. Distribution of the imperfect CTG allele in blood cells of seven DM1+CMT++ cases. More than 100 expanded molecules were sized for each individual. The somatic mosaicism of samples with similar progenitor allele and similar age at sampling was very similar as expected (III-16 and III-17, IV-21 and IV-22). PA (total length of the progenitor allele), DSM (degree of somatic mosaicism). The size of the interruptions at 3'-end for all the DM1+CMT++cases was 84 triplets.

4.2.5.2 Understanding the intergenerational dynamics of the DM1+CMT++ allele

Slight increases in the length of the CTG repeats were observed through maternal transmissions, for example the mutant allele of III-16 increased ~43 CTG repeats when transmitted to the offspring. Similar increases in size where observed in the DM1+CMT++ allele of III-17 (table 4-2). These data indicated the interrupted CTG repeat allele was intergenerationally unstable and explained the slight anticipation observed in the DM1 phenotype between the two generations. Nevertheless, these changes are far smaller than those observed in classic DM1 patients. For instance, combining the data generated in two studies of 66 mother-child pairs, the average size of the inherited allele in affected children was 496 CTG repeats ranging from 65 to 1760. The average size of the progenitor allele in the mother was 75 ranging from 56 to 95 (Barcelo*, et al*. 1993; Redman*, et al*. 1993).

The imperfect CTG repeats seemed to be more intergenerationally stable than the pure CTG repeat, consistent with what was previously observed with the "38" repeat allele in sperm (Leeflang and Arnheim 1995). Strikingly, in the DM1+CMT++ family, patients with similar length of the progenitor allele presented similar age of onset of DM1 symptoms (table 4-2).

Table 4-2. Comparison of the progenitor allele length with the age of onset of DM1 and CMT observed in the DM1+CMT++ cases

	Progenitor allele	Transmitted allele	Length change (CTG)	Age of onset of DM1 (yrs)
III-9	225 (141)	-	-	25
III-16	170 (86)	213	43	44
IV-19	213 (129)	-	-	20
IV-20	213 (129)	-	-	24
III-17	179 (95)	220/25	41/46	35
IV-21	220 (136)	-	-	17
IV-22	225 (141)	-	-	24

The total length of the progenitor allele is displayed in number of triplets; the length of the CTG repeats at the 5'-end is shown in brackets. The age of onset is quite difficult to define so it is only an estimation of when the first symptom of the disease could be detected.

4.2.5.3 Age effects on the level of somatic mosaicism

The repeat length of CTG repeats in blood cells of classic DM1 patients increases in size over the lifetime of an individual (Martorell, *et al*. 1995; Wong, *et al*. 1995; Martorell, *et al*. 1998). A detailed study of 111 DM1 patients over a period of time from one to seven years showed that when the size of the progenitor allele was >200 repeats an increase in average repeat length was detected in a period of more than two years. However, if the number of repeats was < 200 repeats, an increment in the modal allele size was not generally detected. From 34 patients with <200 repeats, only one showed a detectable change in average repeat length (Martorell, *et al*. 1998).

To investigate if the somatic mosaicism of the DM1+CMT++ allele changes over time, repeat blood samples from III-9, III-16 and IV-21 taken with time intervals ranging from 3 to 8 years were investigated by SP-PCR. A statistically significant increase in the allele length was observed in IV-21 over 8 years (Mann-Whitney, $p<0.0001$), with an increase of 34 CTG repeats in the median (figure 4-13 and 4-14). In contrast, no obvious differences in the repeat length distributions were observed between the two samples obtained 3 and 4 years apart from III-9 and III-16 respectively. IV-21 was sampled very young, while III-9 and III-16 were sampled after 55 years of age so this could explain the differences. Additionally, probably the age effects are not as dramatic as in pure CTG repeat tracts.

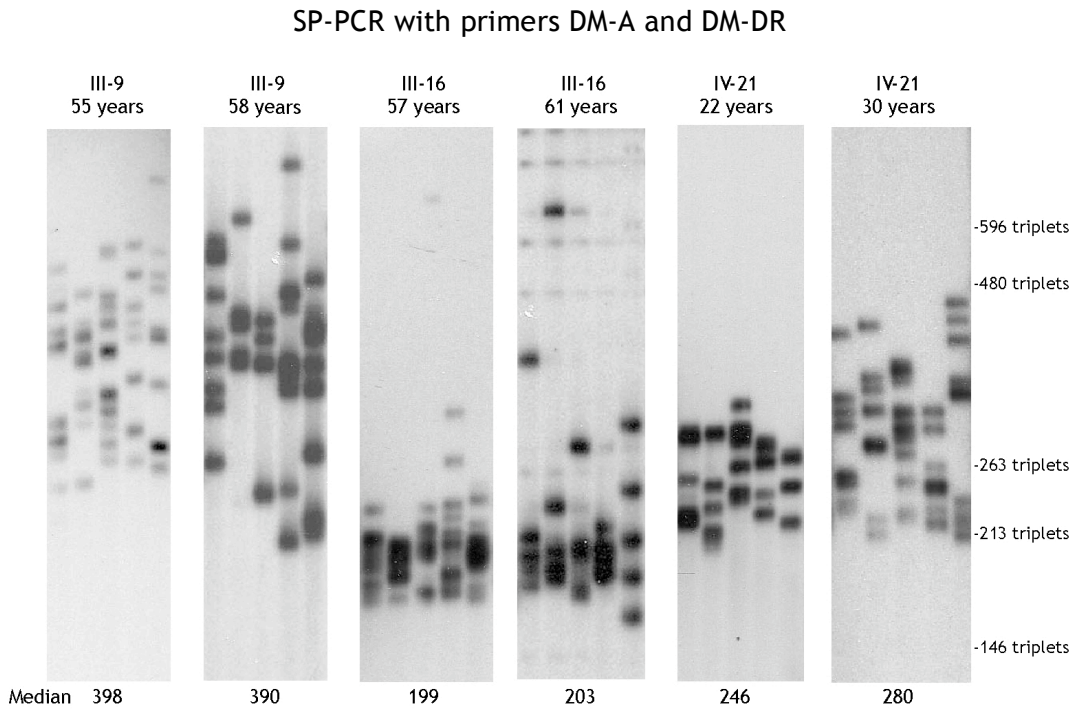


Figure 4-13. SP-PCR analysis of the repeat length in repeat samples from the same patient. Each panel shows 5 lanes of SP-PCR derived from ~15 genomic equivalents. The scale on the right displays the molecular marker converted into number of triplets. The age at sampling is indicated above and the median of the total repeat length below each panel.

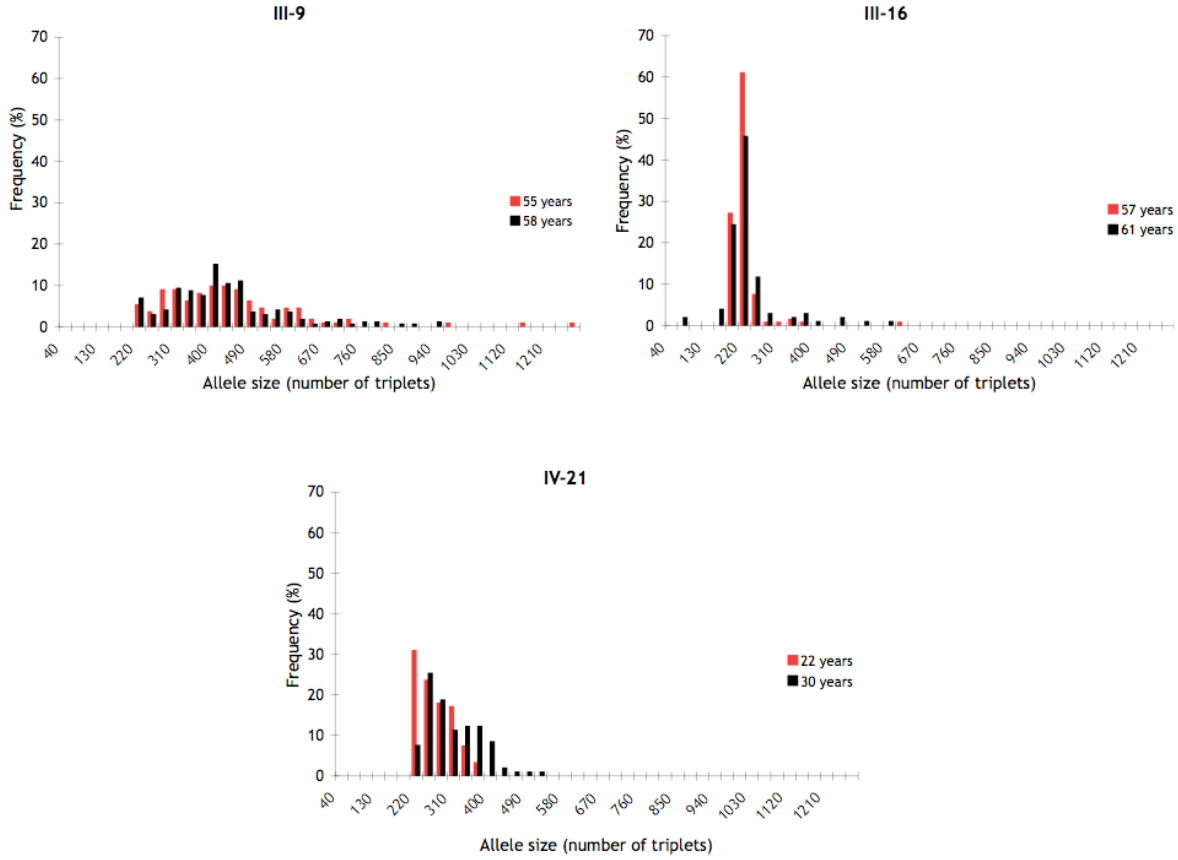


Figure 4-14. Somatic mosaicism of the DM1+CMT++ allele in repeat samples. The histograms represent the frequencies of allele lengths at two different time points. A statically significant variation in the allele lengths was observed in IV-21 over 8 years (Mann-Whitney, $p<0.0001$).

4.2.5.4 Comparison of the level of somatic mosaicism in the DM1+CMT++ family with classic DM1 patients

The progenitor allele length and the age at sampling were identified as the major modifiers of somatic mosaicism in DM1 in an extended study, which comprised the analysis of samples from Costa Rica (50), Texas (30), Scotland (23) and Uruguay (4). Individuals in the extreme of the distribution as congenital cases and individuals with the mild form investigated later in life were not included in the analysis of that study. Linear regression analysis showed that the age at sampling and the progenitor allele length presented significant positive correlations with the degree of somatic mosaicism. To clarify further these interrelationships, the data was analysed using a General Linear Model ANOVA with interaction between the age at sampling and the progenitor allele length. The model revealed that ~66% of the variation in somatic mosaicism is caused by complex interactions between the progenitor allele length and the age at sampling (GLM ANOVA, r = 0.81, p =0.005). The model was: log SV = 2.2228log PA + 3.4945 log AS - 1.0206 log PA x log AS, where SV is somatic variation, PA is progenitor allele in number of CTG repeats and AS is age at sampling in years (Morales 2006).

To determine if the DM1+CMT++ allele affects the degree of somatic mosaicism, the data from the DM1+CMT++ cases was compared with the previous study. The level of somatic mosaicism in the DM1+CMT++ cases was always lower than the expected level for the age at sampling (figure 4-15A) and for the progenitor allele length (figure 4-15B), except for IV-19 and III-9 who both presented a level of somatic mosaicism slightly higher than expected for the length of the progenitor allele. The values of somatic mosaicism calculated with the formula predicted by the General Linear Model were almost all noticeably higher than the observed values (table 4-3). These data strongly suggest that the presence of interruptions at the 3'-end of the repeat tract decreased the instability of the CTG repeats at the 5'-end.

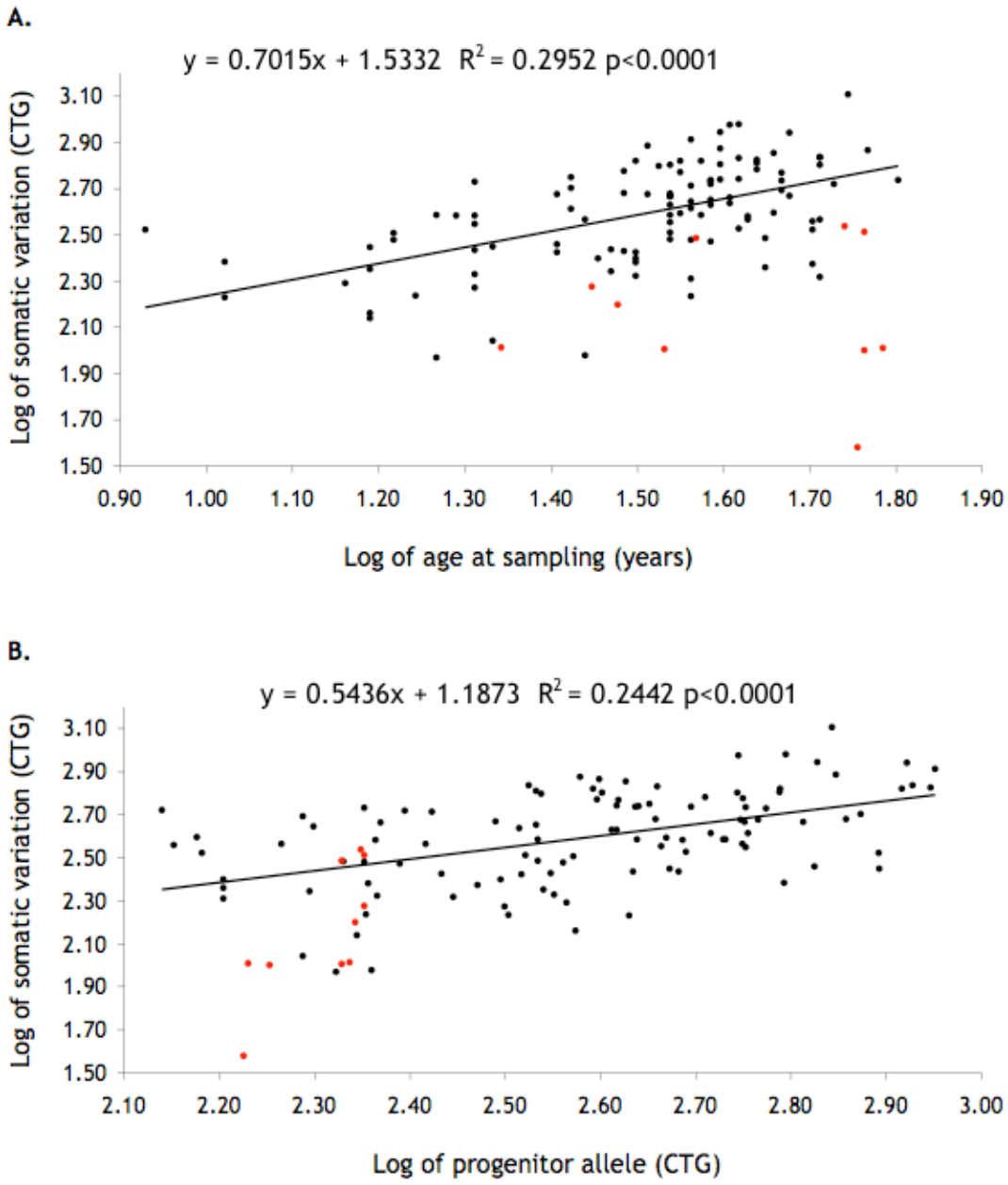


Figure 4-15. Age at sampling and progenitor allele effects on the degree of somatic mosaicism. A. Linear regression analysis between the age at sampling and somatic variation. B. Linear regression analysis between the progenitor allele and somatic variation. Black dots represent data from classic DM1 patients (Morales 2006) and red dots represent the DM1+CMT++ cases. Half of the points should be expected to be located above the tendency line and the other half below. However, all the points from the DM1+CMT++ cases were located below the tendency line in the analysis of age at sampling versus somatic variation. By Fisher exact test, it was shown that the points were statistically different from the tendency line obtained with the DM1 classic patients (p=0.0017), giving evidence for a reduced instability of the DM1+CMT++ alleles.

Table 4-3. Comparing the degree of somatic variation

	Progenitor allele	Age at sampling (years)	Somatic variation observed	Somatic variation estimated
III-9	225 (141)	58	325	511 (419)
III-9	223 (139)	55	344	480 (389)
III-16	170 (86)	61	102	483 (367)
III-16	168 (84)	57	38	442 (328)
IV-19	213 (129)	37	306	302 (221)
IV-20	213 (129)	34	101	275 (197)
III-17	179 (95)	58	100	464 (355)
IV-21	220 (136)	30	158	244 (173)
IV-21	217 (133)	22	103	172 (113)
IV-22	225 (141)	28	189	230 (163)

The total length of the progenitor allele is displayed in number of triplets, with the length of the CTG repeats at the 5'-end shown in brackets. The somatic variation observed was calculated as the difference between the 90th percentile and the 10th percentile. The estimated somatic variation was calculated using the model generated with a GLM (log SV = 2.2228log PA + 3.4945 log AS - 1.0206 log PA x log AS), where SV is somatic variation, PA is progenitor allele in number of CTG repeats and AS is age at sampling in years. The estimated somatic variation taking into account only the CTG repeats at the 5'-end is indicated in brackets.

4.2.6 Investigating DNA methylation in the DM1+CMT++ allele

The CTG repeat at the *DM1* locus cannot be methylated as it does not contain any CpGs. However, the repeat is in a CpG island and is flanked by multiple CpG sites in the vicinity. Methylation of these flanking CpGs may affect the chromatin structure and gene expression. Steinbach and colleagues analysed part of the CpG island at the 3'-end of the *DMPK* gene for methylation status using methylation-sensitive restriction enzymes and Southern blot. In the expanded alleles containing ≥1,000 CTG repeats in congenital cases, the CpG island was hypermethylated, whereas it was unmethylated in normal alleles or expanded alleles of adult onset cases (Steinbach, *et al*. 1998). The interruptions present in the DM1+CMT++ allele provide new CpG methylation sites within the repeat tract; methylation of the repeat tract may similarly affect chromatin structure and gene expression so we investigated whether those sites were methylated.

To determine the pattern of methylation in the DM1+CMT++ allele, genomic DNA of the DM1+CMT++ cases was digested with two methylation sensitive enzymes, either *Aci*I or *Hha*I. The digested DNA was amplified by PCR with specific primers to detect whether the site of interest was methylated or not. If the site was methylated, the enzyme will not recognize the site and a PCR product could be amplified. In contrast, if the site was unmethylated the enzyme will recognize and digest the site and in consequence no PCR product will be generated. Several controls are needed including a no enzyme control. Ideally, three pair of primers should be used: one pair targeted to a region where there is no restriction site; a second pair to amplify a product only in the undigested control; and a third to generate a product only if the restriction site under study is methylated. The first pair will act as an internal control of the PCR, the second will act as an internal control of the digestion and the third is to investigate whether the site is methylated or not.

The *Aci*I sites and the *Hha*I site located within the CTG repeat tract in the DM1+CMT++ allele were investigated by amplifying the digested products with primers DM-C and DM-DR. In our case, it was not necessary to use three pairs of primers because the normal allele does not have a methylation site so it will act as an internal control of the PCR. The *Hha*I site seemed to be unmethylated in the majority of the expanded alleles. The expanded alleles were not detected when the DNA was digested, and the normal alleles always amplified as expected (figure 4-16). In contrast, the majority of the *Aci*I sites seemed methylated, and several expanded alleles were detected when the DNA was digested. It should be pointed out that analysis with *Aci*I enzyme requires a control of the digestion reaction. This data was generated at the end of the project and there was insufficient time left to perform the digestion reaction control.

Several *Aci*I sites within the CTG repeat tract in the DM1+CMT++ allele were methylated. Consequently, the activity of *SIX5* and *DMPK* genes might be altered. It will be interesting to test whether the pattern of methylation in the flanking CpG sites at either end of the repeat tract is also modified.

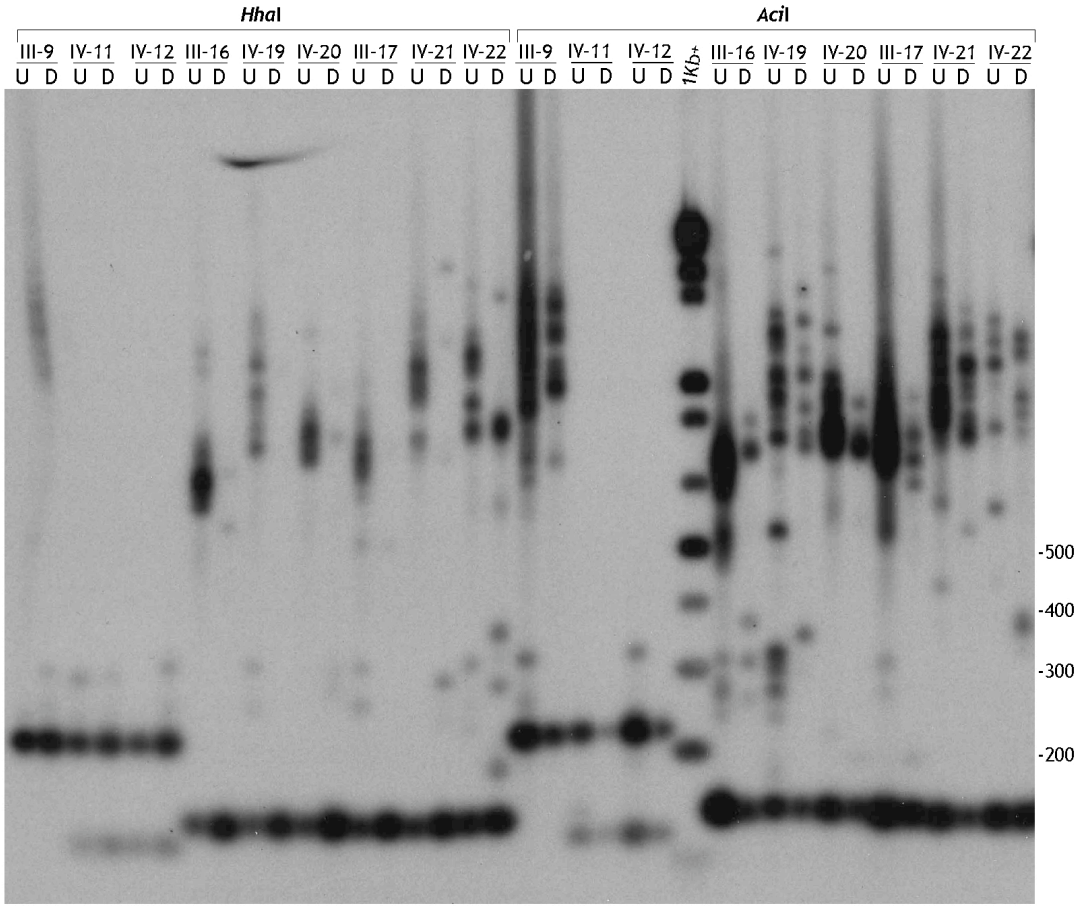


Figure 4-16. Methylation pattern of the DM1+CMT++ allele. Genomic DNA was digested either with *Hha*I or *Aci*I, and amplified with primers DM-C and DM-DR. The digested products and the undigested controls were resolved on a 1.5 % agarose gel and detected by Southern blot hybridisation with the DM1 CTG repeat probe. The position of the molecular weight marker is indicated on the right. There are several non-specific bands with sizes between 250-350 bp with a weak signal, which are probably just contamination.

4.3 Discussion

It has been demonstrated that the combination of cloning and sequencing, restriction mapping and RP-PCR together with the use of single molecule separated alleles has been extremely successful in defining the structure of the imperfect CTG allele in the DM1+CMT++ family. It should be noted that each technique has advantages and disadvantages, and it is necessary to use a combination of the techniques to obtain a reliable structure due to the instability of the repeat structures. The 3'-end of the repeat was amplified by PCR using the primer DM-GGC and DM-DR, and it was verified that the 3'-end of the repeat tract was practically unchanged in the different patients. The structure of the DM1+CMT++ allele has been characterised. It is comprised of a variable number of CTG repeats at the 5'-end followed by: GGC, CCG, CCGCTG and 35 CTG repeats at the 3'-end.

The dynamics of the CTG repeat at the 5'-end of DM1+CMT++ allele was investigated by SP-PCR analysis. It was shown that the CTG repeats at the 5'-end were somatically unstable and biased toward expansions, but to a lower degree than the CTG repeats in classic DM1 patients. This gives a molecular explanation for the mild DM1 phenotype observed in the patients. The presence of interruptions within the CTG repeats at the *DM1* locus seemed to stabilise the repeat tract.

Generally, expanded CTG repeats increase in size when transmitted. For instance, a mother with 100 repeats has a 62% risk of transmitting a congenital allele (Redman*, et al.* 1993). However in the DM1+CMT++ family, only small changes in the number of CTG repeats at the 5'-end were observed through maternal transmissions. The small increases in size in the CTG repeats explain the reduced anticipation in the DM1 phenotypes observed in the DM1+CMT++ family. An increase in the severity of symptoms and a reduced age of onset was observed in the CMT phenotype through the generations in the DM1+CMT++ family. If CMT is caused by the presence of interruptions, the anticipation could be explained by the small changes in the number of CTG repeats at the 5'-end. The CCG repeats are more toxic than the CTG repeats, since only 55 repeats are enough to express a phenotype in FXTAS (Hagerman and Hagerman 2004). Therefore, the small changes in the number of CTG repeats at the 5'-end could be enough to retain more alleles in the nucleus and consequently more proteins would be retained, explaining the anticipation in CMT. In order, to probe this hypothesis, it will be necessary to demonstrate first that the interruptions are definitely the cause of CMT.

The presence of imperfect CTG alleles has not been taken into account before, and it could be one of the factors responsible for the variability observed in the DM1 patients. It will be interesting to investigate the presence of interruptions in patients with a phenotype not as severe as expected for the number of CTG repeats. In addition, a similar mechanism may explain the CMT phenotype in families with an unknown aetiology.

The expansion of CTG repeats in the *DM1* locus has been shown to alter the adjacent chromatin structure establishing a region of condensed chromatin (Otten and Tapscott 1995). *In vitro* studies showed that the CTG repeats are strong nucleosome positioning elements and the efficiency of the formation increased with the number of triplets, indicating that the expansion may repress the transcription of the *DMPK* gene and/or neighbouring genes (Wang, *et al*. 1994). Furthermore, the CTG repeats in the *DM1* locus are inserted in a CpG island that controls the activity of the downstream gene *SIX5* (Boucher, *et al*. 1995). Two CTCF binding sites flank the CTG repeat tract and form an insulator element between *DMPK* and *SIX5* (Filippova, *et al*. 2001). In congenital DM1 cases, the CpG island is methylated so the enhancer element of *SIX5* is lost. In addition, the CTCF protein cannot have access to the sites and consequently the activity of the insulator is compromised. Reduced levels of *SIX5* transcript were observed in DM1 patients (Klesert, *et al*. 1997; Thornton, *et al*. 1997). The interruptions in the DM1+CMT++ allele provide new CpG sites, of which at least some seemed to be methylated (section 4.2.6). It will be interesting to investigate the methylation pattern in the 5' and 3' flanking DNA of the repeat tract in the DM1+CMT++ allele. Two possible outcomes can be envisioned. The first possible outcome is that the CpG island is methylated as in congenital DM1 cases. Alternatively, by analogy to the mechanism observed in FXTAS, the CpG island may remain unmethylated and the level of *SIX5* increased (Tassone, *et al*. 2000b). If indeed the level of *SIX5* increased, it could be a possible explanation for the complex phenotype observed in the DM1+CMT++ family. Furthermore, the CTG repeats are strong nucleosome positioning elements (Wang, *et al*. 1994) in contrast to CCG repeats, which are strong inhibitors (Wang and Griffith 1996). It will be interesting to investigate the structure of the chromatin at the DM1+CMT++ allele where CTG repeats are next to CCG repeats, and to analyse whether the activity of the *DMPK* and *SIX5* genes may be altered.

Future investigations will focus on trying to explain how the other symptoms in the DM1+CMT++ family can be explained by the presence of this novel repeat. As was postulated in the previous chapter, the most plausible explanation is by a mechanism similar to FXTAS. The presence of

interruptions in the *DMPK* transcript probably alters its secondary structure and consequently novel proteins may be trapped and a novel phenotype could be expressed.

5 Investigating the molecular lesion in a number of sporadic DM1 cases with an unusual molecular diagnosis

5.1 Introduction

The DM1 mutation was identified as an expansion of a CTG repeat at the 3'-end of the *DMPK* gene in 1992. Evidence obtained from cloning and sequencing 50-80 repeat alleles and hybridisation with CTG probes revealed that the expanded allele was comprised of pure CTG repeats (Brook, *et al*. 1992; Fu, *et al*. 1992; Mahadevan, *et al*. 1992). An imperfect CTG repeat allele within the normal range (37 repeats) was identified in a sperm sample while studying the male germline mutation rate of triplet repeats (Leeflang and Arnheim 1995). The sequence of the "37" repeat allele revealed that it consisted of $(CTG)_4$ $(CCGCTG)_{16}$ (CTG). The presence of the CCGCTG hexamer can be easily identified by restriction digestion with *Aci*I (CCGC). The presence of the *Aci*I sites was also investigated in 71 individuals carrying between 30 and 54 CTG repeats in the DM1 locus, but no other imperfect CTG repeat alleles were identified (Martorell, *et al*. 2001). These data suggested that imperfect CTG repeat alleles must not be common within the premutation alleles. Nevertheless, the Cockburn group presented preliminary data at the International Myotonic Dystrophy Consortium (IDMC4) that three DM1 patients presented an interrupted CTG repeat allele detected by RP-PCR (Cockburn, *et al*. 2003). We had demonstrated in chapter four that the molecular lesion in the DM1+CMT++ family consisted of an imperfect CTG allele at the *DM1* locus. Therefore, it could be possible that the presence of interruptions within the CTG repeats at the *DM1* locus might not be such a rare event.

Interestingly, our collaborators Jean-Louis Mandel and his colleagues had identified a number of sporadic DM1 patients, who were referred for genetic testing and then presented with an unusual molecular diagnosis. The patients could be classified into two groups. The patients in the first

group (DM1-UC1, DM1-UC2, DM1-UC3, DM1-UC4, DM1-UC5, DM1-UC6, DM1-UC7 and DM1-UC8) presented an interrupted pattern on RP-PCR, and by partially sequencing the 3'-end CTG repeat tract a number of interruptions (CCGCTG or CCTGCTC) were detected. Two patients are related, DM1-UC1 is the mother of DM1-UC2. DM1-UC7 has two normal alleles, with one large interrupted allele containing "40" repeats. In the second group of patients (DM1-UC9, DM1-UC10, DM1-UC11, DM1-UC12, DM1-UC13, DM1-UC14 and DM1-UC15), discrepancies between Southern blot of restriction digested genomic DNA and 3'-end RP-PCR studies were found. In DM1-UC9, a deletion of 10 bp was sequenced at the 3'-end of the repeat tract. DM1-UC10 was the father of DM1-UC11 (a foetus). It was postulated that probably a large deletion occurred between the father and foetus. An expanded allele was detected by Southern blot of restricted digested genomic DNA in both patients, but the expanded allele was only detected in the father (DM1-UC10) by 3'-end RP-PCR. In the patient DM1-UC12, two alleles within the normal range were detected by PCR but only one allele was detected by RP-PCR.

In this chapter, the series of molecular analyses performed to further characterise the DM1 mutation in these unusual DM1 cases are presented. The dynamics of the DM1 alleles was investigated by SP-PCR analysis to test whether the instability of the CTG repeats was modified in the unusual cases. To gain an insight into the possible effects of the unusual mutation on the pathogenesis, the pattern of DNA methylation at the 3'-end flanking DNA was also investigated.

5.2 Results

5.2.1 Genotyping the *DM1* locus by traditional methods

5.2.1.1 PCR

PCR, Southern blot digested genomic DNA and RP-PCR was previously performed by Mandel's group, however the analysis of the unusual cases

began by genotyping the *DM1* locus by a standard PCR to confirm the previous results.

To genotype the *DM1* locus, a standard PCR was performed with primers DM-C and DM-DR (figure 5-1). DM1-UC7, DM1-UC12 and DM1-UC15 presented as heterozygous carrying two normal alleles. One normal allele and an expanded allele were detected in DM1-UC3, DM1-UC5, DM1-UC6, DM1-UC8 and DM1-UC9. The expanded alleles were smeary, consistent with somatic mosaicism as is generally observed in classic DM1 patients. In DM1-UC1, DM1-UC2, DM1-UC4, DM1-UC10, DM1-UC11, DM1-UC13 and DM1-UC14, the expanded allele was not detected and only one normal allele was identified. A polymorphism within the primer site could be one reason why an expanded allele was not detected in the latter patients. However, three independent combinations of primers (DM-C/DM-DR, DM-A/DM-BR and DM-C/DM-ER) failed to detect an expanded allele (figure 5-1A). Alternatively, an additional lesion such as an insertion within the CTG repeats, a deletion or a rearrangement at the flanking DNA might explain why an expanded allele was not detected by the standard PCR. Taking into account the information given previously, an interruption seemed to be the cause in DM1-UC1, DM1-UC2 and DM1-UC4, whereas a deletion at the 3'-end was assumed to be most likely in DM1-UC11.

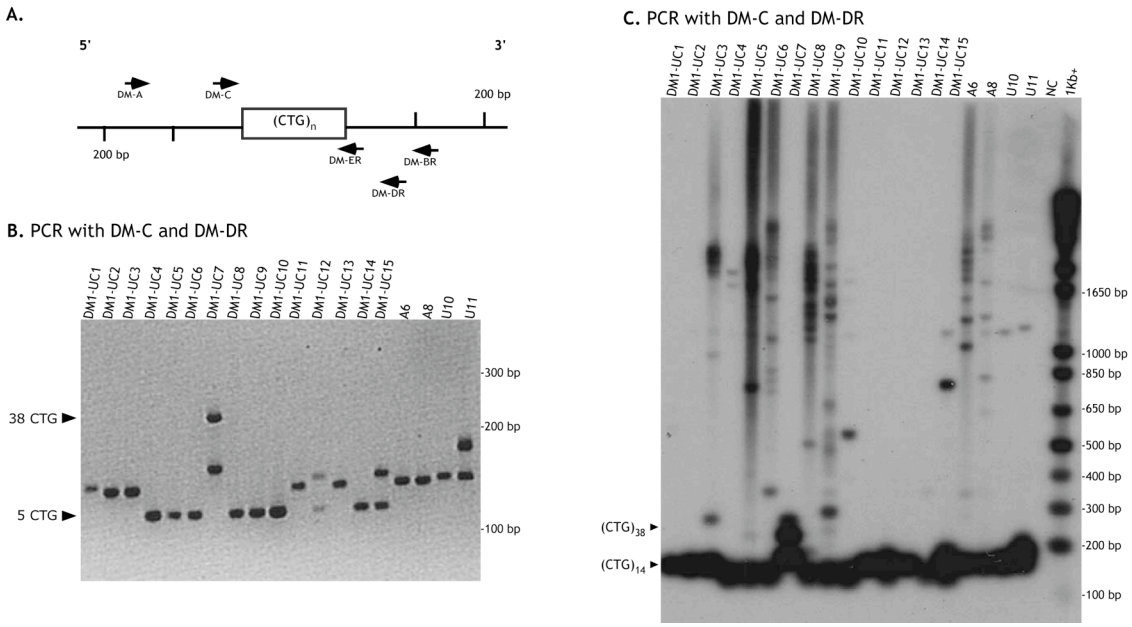


Figure 5-1. Genotype of the *DM1* locus by standard PCR. A. Schematic diagram of the *DM1* locus indicating the position of the primers tested. B. PCR products were resolved on a 2% Nusieve, 1% agarose gel. C. Representative autoradiograph. PCR products were resolved on a 1% agarose gel, transferred by Southern blot hybridisation and detected with the DM1 CTG repeat probe. The sizes of the marker are indicated on the right side and the estimated number of repeats is indicated on the left side. NC (negative control), A6 and A8 (two classic DM1 patients), U10 and U11 (two normal controls). Additional bands with sizes of >1 Kb were detected in U10 and U11. The intensity is not as strong as the normal bands so it was probably contamination. Similarly, the additional bands detected in DM1-UC4, DM1-UC10 and DM1-UC15 could represent contamination.

5.2.1.2 RP-PCR

RP-PCR was performed with the CTG_5 repeat specific primer from each end of the repeat tract both to analyse if an expanded allele could be amplified in DM1-UC1, DM1-UC2, DM1-UC4, DM1-UC10, DM1-UC11, DM1-UC13 and DM1-UC14, and to characterise further the DM1 mutation in the other unusual patients.

In DM1-UC3, DM1-UC5, DM1-UC6, DM1-UC8 and DM1-UC9, an expanded allele was visualised with multiple primers at both ends (figure 5-2). In DM1-UC9, the expanded allele was not detected at the 3'-end when using the primer DM-ER. This result was consistent with the 10 bp deletion at the 3'-end detected by sequencing, which removed part of the DM-ER priming site. In DM1-UC7, DM1-UC12 and DM1-UC15, only ladders corresponding in size to the normal alleles were detected by RP-PCR consistent with the standard PCR. In DM1-UC1, DM1-UC2, DM1-UC4, DM1-U10, DM1-UC11, DM1-UC13 and DM1-UC14, an expanded allele was detected at the 5'-end of the

repeat tract with multiple primers (DM-T, DM-A, DM-H and DM-C), but was negative at the 3'-end of the array even with primers located up to 1 Kb distal to the repeat tract. From this data, it could be postulated that the 5'-end of the CTG repeat tract seemed to be intact in all the unusual cases. However, an additional lesion at the 3'-end must be present in DM1-UC1, DM1-UC2, DM1-UC4, DM1-U10, DM1-UC11, DM1-UC13 and DM1-UC14.

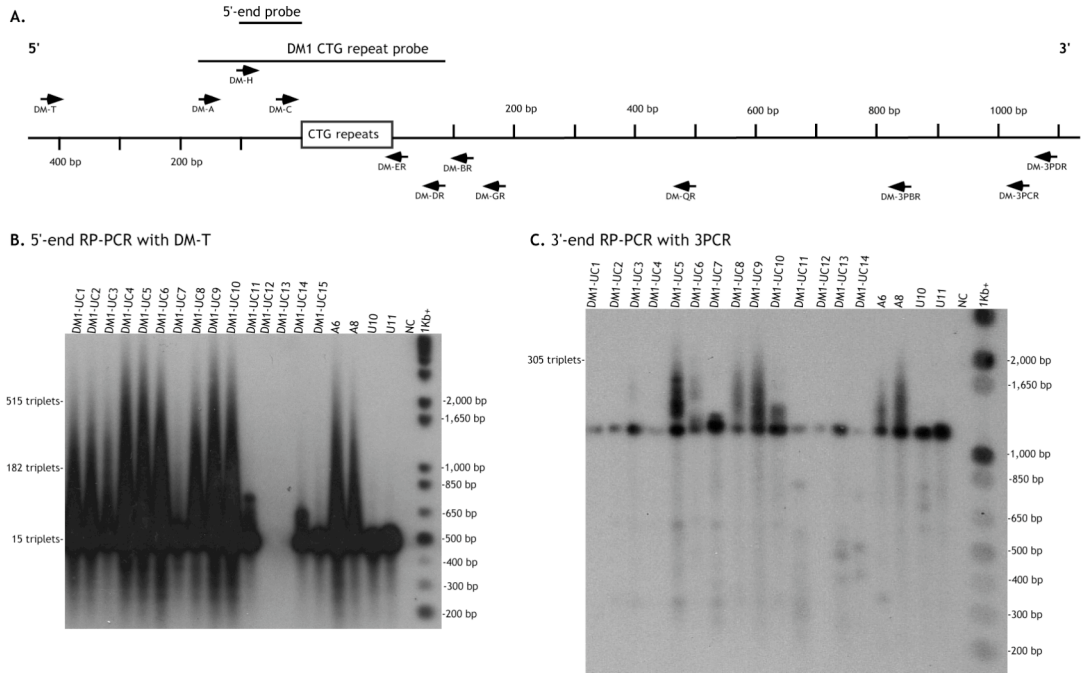


Figure 5-2. RP-PCR studies. A. Diagram of the *DM1* locus indicating the position of the different primers and the two probes used in the analysis. B. 5'-end RP-PCR with primer DM-T. The products were run on a 1.5% agarose gel and detected with the 5'-end probe by Southern blot hybridisation. The DNA from DM1-UC12 and DM1-UC13 did not amplify well; it was probably a pipetting error while setting up the PCR. The samples were repeated and normal alleles were detected in DM1-UC12 and an expanded allele in DM1-UC13. C. 3'-end RP-PCR with primer 3PCR. The products were resolved on a 1.5% agarose gel and detected with the DM1 CTG probe. A faint ladder corresponding to the expanded allele was detected in DM1-UC3. NC (negative control), A6 and A8 (two classic DM1 patients), U10 and U11 (two normal individuals). The size of the marker is indicated on the right and the estimated number of CTG repeats is indicated on the left side of each panel.

The DNA samples from DM1-UC12 and DM1-UC15 were genotyped as both having two normal alleles of 5 and 17 CTG repeats at the *DM1* locus. It was suspected these might be two samples from the same patient, so three highly polymorphic loci *ERDA-1*, *CTG18.1* and *MSH3* were genotyped in both samples. In all loci investigated, both samples were heterozygous presenting alleles of the same size (data not shown). The probability of finding a heterozygous individual carrying the same sizes of repeat alleles at the three loci was calculated by the product of the probabilities of each allele size. It was estimated to be approximately 1:59 $x10^6$ indicating that

both samples almost certainly corresponded to the same patient probably arising from a mistake made when the samples were sent. Consequently, sample DM1-UC15 was not included in further analyses.

5.2.2 Genotyping the *DM1* locus by a modified PCR

A modified PCR was performed to investigate the possibility that an imperfect CTG allele with a high GC content might prevent the detection of the alleles by a standard PCR.

The genomic DNA was amplified with primers DM-C and DM-DR in the presence of 10% DMSO. Expanded alleles were amplified in DM1-UC2, DM1-UC3, DM1-UC4, DM1-UC5, DM1-UC6, DM1-UC8, DM1-UC9, DM1-UC10, DM1-UC11 and DM1-UC13. In DM1-UC13, the expanded allele was still only detected as a faint interrupted smear. This result indicated that the expanded allele in DM1-UC13 did not amplify as effectively as the normal allele and/or the sequence of the expanded allele did not contain pure CTG repeats, consequently the DM1 CTG repeat probe did not hybridise properly. Two alleles within the normal range were amplified with DM1-UC7 and DM1-UC12, verifying the data obtained by the standard technique. Nevertheless, an expanded allele was still not amplified in DM1-UC14 (figure 5-3). The expanded alleles were all smeary, consistent with somatic mosaicism. An expanded allele was also identified in DM1-UC1 (data not shown). The expanded allele of DM1-UC11 (a foetus) was detected as a discrete band. The DM1 expanded allele in foetuses and newborns show practically no repeat length heterogeneity and are detected as single bands on the gels when amplified by PCR. The alleles become more heterogeneous in size with increasing age (Wong, *et al*. 1995; Martorell, *et al*. 1998). Therefore, the absence of somatic mosaicism of the expanded allele of DM1-UC11 was expected.

The "41" repeat allele of DM1-UC7 was cloned and sequenced to determine the structure of the interrupted allele. The sequence revealed that the allele consisted of $(CTG)_6$ $(CCGCTG)_{15}$ $(CTG)_5$. The CTG repeats were interrupted by a series of CCG, in a very similar manner to the structure

sequenced in the "38" repeat allele of III-9, IV-11 and IV-12 (chapter three, section 3.2.3). It is not known yet how many people were screened to identify these unusual cases so the frequency of that allele cannot be estimated. However, it should be pointed out that an interrupted allele of that size is not a unique event, as was previously believed.

The expanded allele in DM1-UC1, DM1-UC2, DM1-UC4, DM1-UC10, DM1-UC11 and DM1-UC13 was not visualised by the standard PCR. Therefore, the most likely explanation is an imperfect CTG repeat allele with a high GC content.

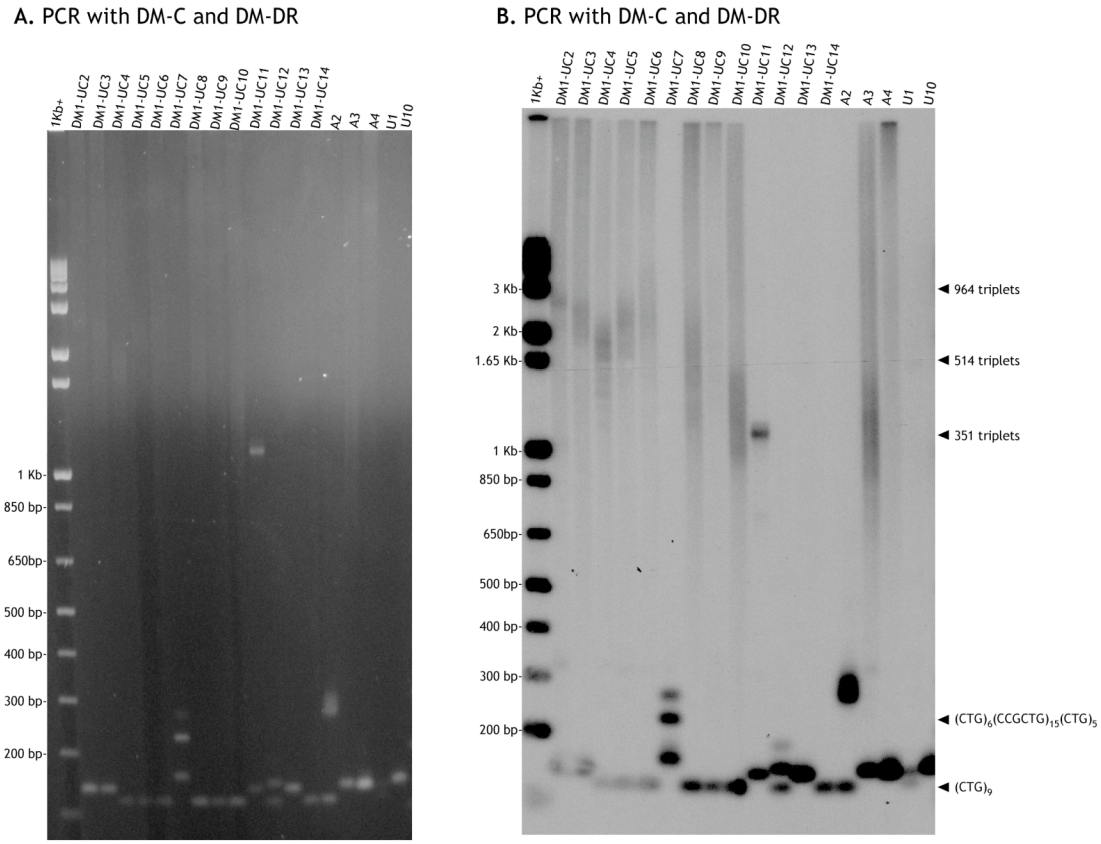


Figure 5-3. Genotyping the *DM1* locus by a modified PCR. Genomic DNA was amplified with primers DM-C and DM-DR in the presence of 10% DMSO. A. The products were resolved on a 1.5% agarose gel and stained with ethidium bromide. B. The gels were detected with the DM1 CTG repeat probe by Southern blot hybridisation. A2, A3 and A4 (classic DM1 patients), U1 and U10 (normal individuals). The top bands of DM1-UC7 and DM1-UC12, which were not as strong as the other two bands probably corresponded to heteroduplex DNA. A faint interrupted smear extending up to the top of the gel was detected in DM1-UC13, although it is not visible here. The size of the marker is indicated to the left of each panel and the estimated sizes of repeats are shown on the right.

5.2.3 Investigating interruptions by RP-PCR assays

Mandel's group have identified CCGCTG interruptions in some unusual cases, so the $CCGCTG_3$ specific primer was used in RP-PCR in order to gain an insight into the number and the location of the interruptions. The patients were investigated with RP-PCR from both ends of the CTG repeat tract using either the $CCGCTG_3$ specific primer or the CTG_5 specific primers. The "41" repeat allele of DM1-UC7 contained 15 CCGCTG repeats and was used as an internal control.

In DM1-UC7, a ladder was detected from both ends when using the $CCGCTG_3$ repeat specific primer. By measuring the top and the bottom of the ladder, 15 copies of the CCGCTG hexamer were estimated. Two small ladders and a gap in the middle were detected with CTG_5 specific primer (figure 5-4). The gap corresponded to the ladder of CCGCTG repeats, and the small ladders corresponded in size to the CTG repeats located at either end of the CCGCTG hexamers. The structure of the "41" repeat allele detected by RP-PCR was consistent with the sequence. In a similar way, the structure of the other unusual cases was postulated. At the 5'-end using the CTG_5 repeat specific primer, a large uninterrupted tract of CTG repeats was observed in DM1-UC2, DM1-UC3, DM1-UC4, DM1-UC5, DM1-UC6, DM1-UC8 and DM1-UC11. In DM1-UC11, the ladder detected with the CTG repeat specific primer goes up to ~450 bp so it could be determined that the allele contained ~105 CTG repeats at the 5'-end. The numbers of CTG repeats in the other unusual cases could not be determined because the ladder extended all the way up to the top of the gel. In addition, a ladder with the $CCGCTG_3$ specific primer was detected at the 5'-end in DM1-UC11, starting at the end of the CTG repeat ladder. The other unusual cases were negative with the $CCGCTG_3$ repeat specific primer at this end.

At the 3'-end, DM1-UC2, DM1-UC3, DM1-UC4, DM1-UC5, DM1-UC6 and DM1-UC8 presented ladders of different sizes with the $CCGCTG_3$ repeat specific primer. The ladder detected with the CTG_5 repeat specific primer presented gaps, some but not all of which corresponded with the CCGCTG

ladders. By sizing the top and bottom part of each ladder, the structure of the repeat tract was estimated for each patient.

The DNA from the top part of the ladders obtained with the $CCGCTG_3$ repeat specific primers from DM1-UC2, DM1-UC3, DM1-UC4, DM1-UC5, DM1-UC6 and DM1-UC8 at the 3'-end was purified and cloned. The inserts were quite stable in *E. coli* so it was possible to obtain a reliable sequence (work performed by Stefanatos Rhoda, an undergraduate project student under my supervision). The sequences verified the majority of the estimated sequences obtained by sizing the ladders obtained by RP-PCR. In DM1-UC2, DM1-UC5, DM1-UC6 and DM1-UC8, a number of CCGs were detected in the sequences.

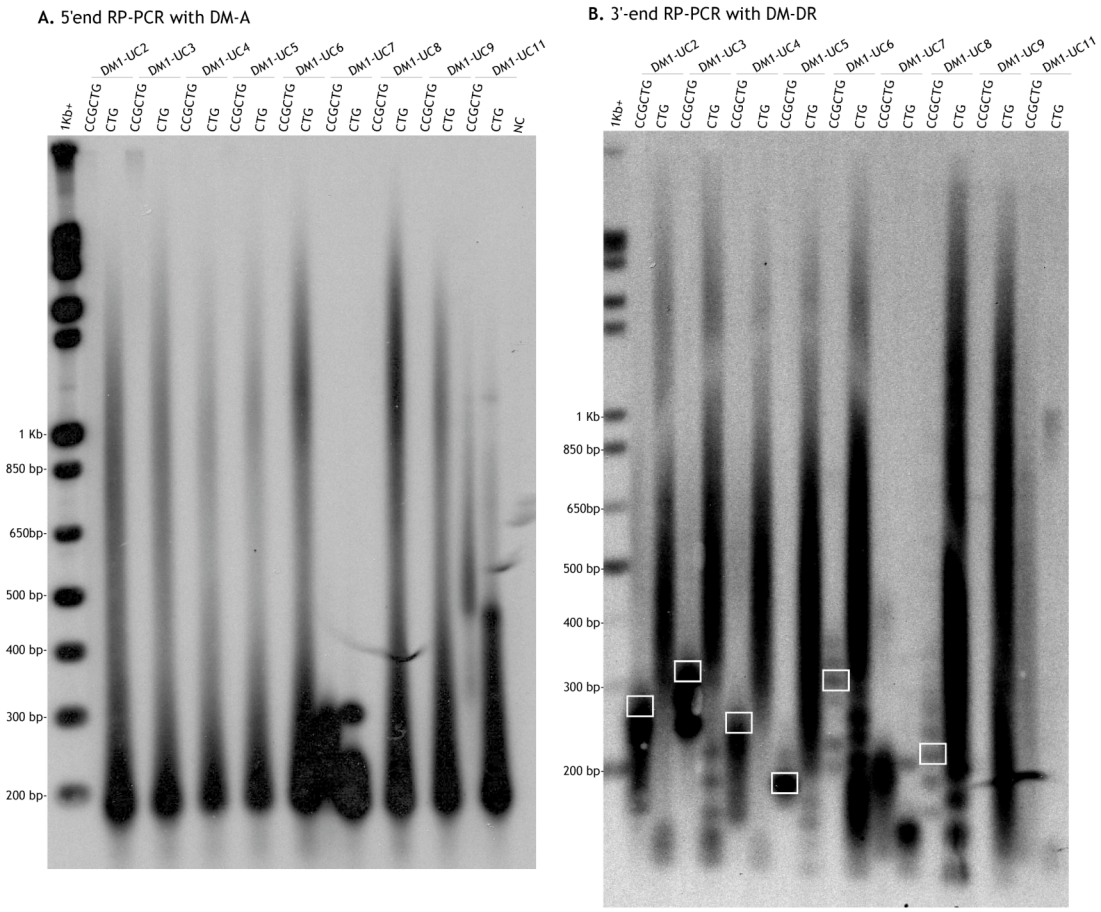


Figure 5-4. RP-PCR in the presence of 10% DMSO. A. 5'-end RP-PCR analysis using primer DM-A. The products were resolved on a 1.5% agarose gel and detected by Southern blot hybridisation with the 5'-end probe. B. 3'-end RP-PCR using primer DM-DR. The products were detected with the DM1 CTG repeat probe. The white boxes show the DNA that were extracted from agarose gel, cloned and sequenced. The size of the DNA molecular weight marker is indicated either on the left of each panel. NC (negative control).

The presence of CCG sequenced in some of the clones was verified by RP-PCR using the CCG_5 repeat specific primer, designed previously to investigate the DM1+CMT++ cases (chapter four). Bands were detected by RP-PCR only from the 3'-end in DM1-UC2, DM1-UC5, DM1-UC6, DM1-UC8 (figure 5-5), and DM1-U13 and DM1-UC14 (figure 5-6). Taking together the information obtained by RP-PCR and sequencing, an estimated structure has been postulated for almost all of the unusual cases investigated (figure 5-7). Most of the unusual cases contained a number of CCGCTG hexamers and CCG repeats. It should be pointed out that in DM1-UC2, DM1-UC3, DM1-UC5 and DM1-UC6, higher order repeats units, containing subrepeats seemed to be present (indicated with black arrows in figure 5-7). For example in DM1-UC2, a higher order repeat unit comprised of 5 CTG repeats and 5 CCG repeats is repeated 3 times. Similarly in DM1-UC3, a higher order repeat containing 2 CTGs and 4 CCGCTGs is repeated two times, and another higher order repeat containing 4 CTGs and 1 CCGCTG is repeated three times.

In DM1-UC9, the ladder detected with the CTG_5 specific primer at both ends was continuous, and no bands were detected either with the $CCGCTG_3$ repeat or with the CCG_5 repeat specific primers (figure 5-5). This data indicated that the CTG repeats in the expanded allele are pure. A deletion of 10 bp was sequenced at the 3'-end, which seemed to be the only unusual feature of this DM1 allele.

The structure of the expanded allele of DM1-UC10 seemed to be interrupted with ~31 CCGCTG hexamers. Nevertheless, the allele inherited by DM1-UC11 (a foetus) was quite different. The number of CTG repeats was reduced, whereas the number of CCGCTG hexamers increased in size. The CCGCTG increased by ~88 repeats, whereas the CTG reduced by ~50 repeats at the 3'-end and by ~23 repeats at the 5'-end. These data revealed that the CCGCTG hexamers and the CTG repeats at both ends were unstable.

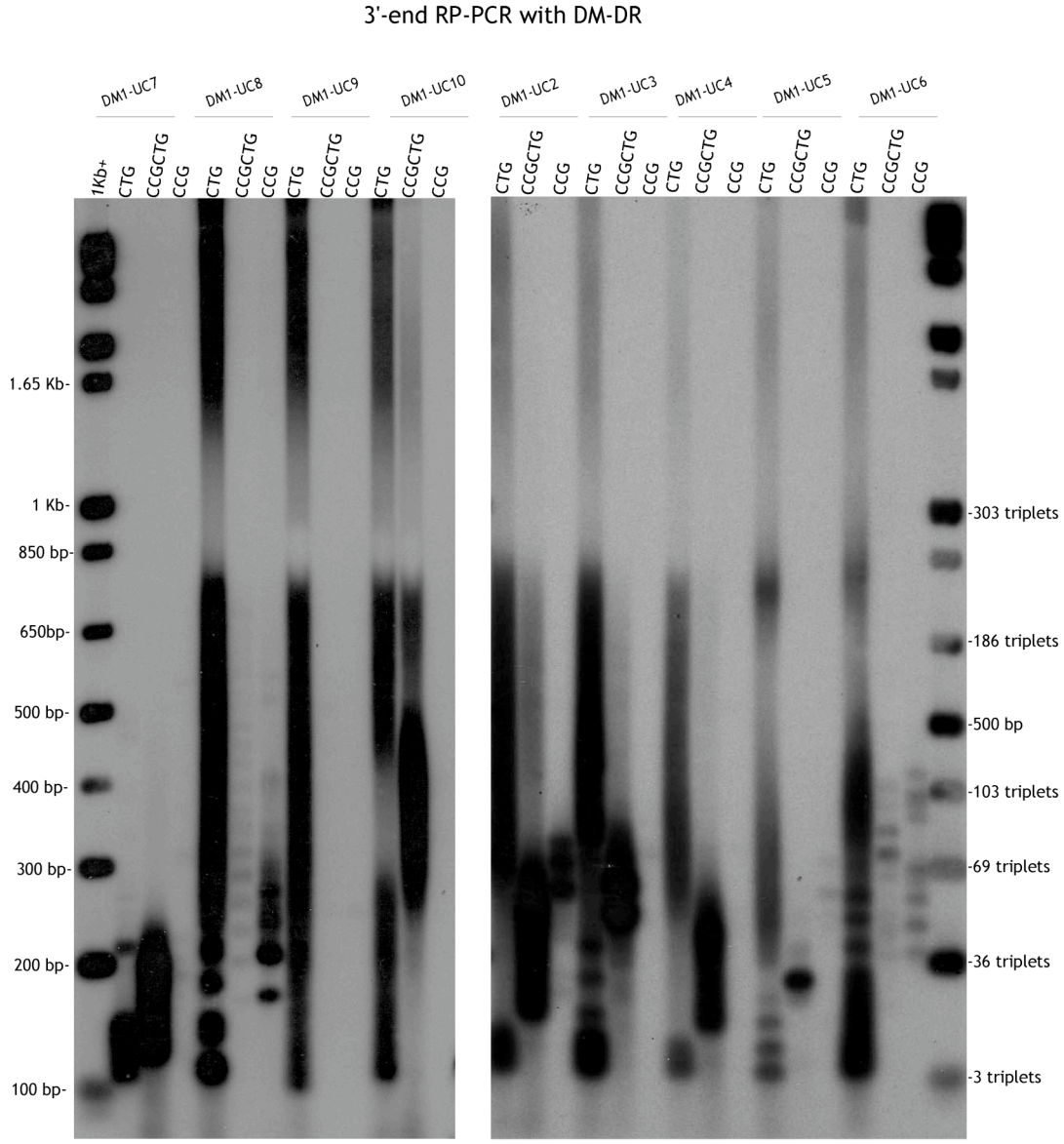


Figure 5-5. Representative 3'-end RP-PCR. Genomic DNA was amplified by RP-PCR using the primer DM-DR. Three different repeat specific primers were used independently: the CTG_5 repeat specific primer, the $CCGCTG_3$ repeat specific or the CCG_5 repeat specific. The products were resolved on a 1.5% agarose gel and detected by Southern blot hybridisation with the DM1 CTG repeat probe. Bands were detected with the CCG_5 repeat specific primers only in DM1-UC5, DM1-UC6 and DM1-UC8. The sizes of the DNA molecular weight marker are indicated on the left and the number of estimated repeats on the right.

The structure of the expanded alleles of DM1-UC13 and DM1-UC14 was not determined completely (figure 5-7 and table 5-1). In DM1-UC13, a long run of CTG repeats was detected at the 5'-end. From the other end, two ladders with a big gap in between were detected with the CTG_5 repeat specific primer. Part but not the entire gap corresponded with the CCG repeat ladder and with two bands detected with CCGCTG. Additionally, the smear detected with the CTG_5 repeat specific primer at the 3'-end was interrupted and faint. These results suggest that other repeat units or

arrangements are present. This result is consistent with the PCR analysis where a faint interrupted smear was detected. The expanded allele of DM1-UC14 did not amplify by PCR, and a short ladder was detected by RP-PCR with the CTG_5 repeat specific at the 5'-end and at the 3'-end with the CCG_5 repeat specific primer. Taking into account all these data, it could be postulated that the expanded allele must contain a large interruption of either a novel repeat unit or arrangement of repeats (figure 5-6).

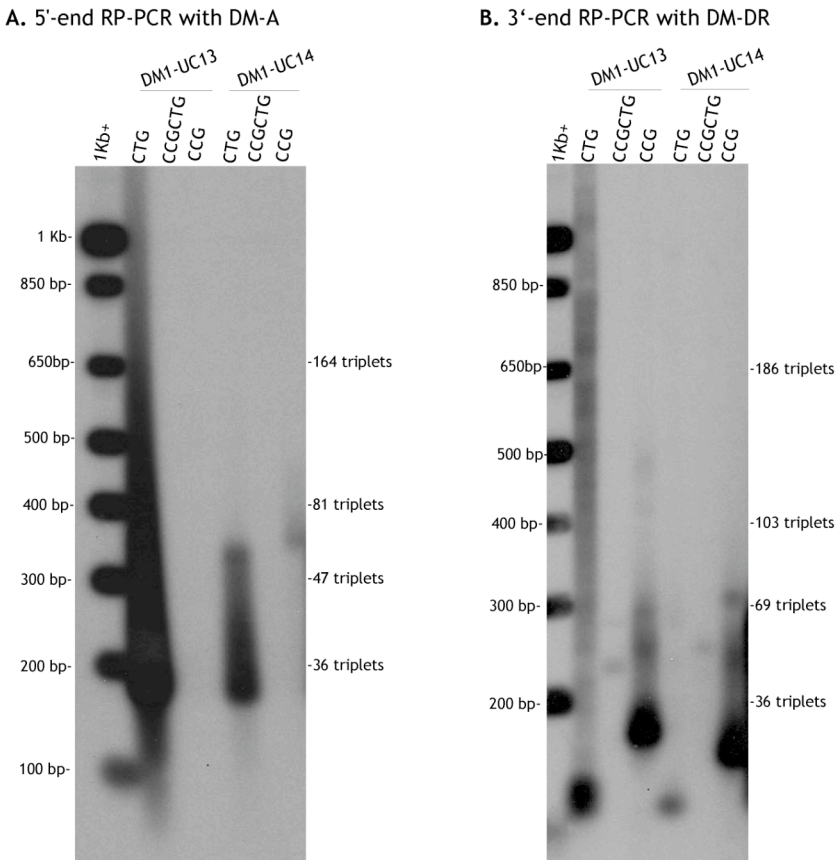


Figure 5-6. RP-PCR analysis in the presence of 10% DMSO. Genomic DNA was amplified by RP-PCR. Three different repeat specific primers were used independently: the CTG_5 repeat specific primer, the $CCGCTG_3$ repeat specific primer and the CCG_5 repeat specific primer. The products were resolved on a 1.5% agarose gel and detected by Southern blot hybridisation with the 5'-end probe (A) or the DM1 CTG repeat probe (B). The position of the molecular weight marker is indicated on the left side of the panels and the estimated number of repeats is indicated on the right side.

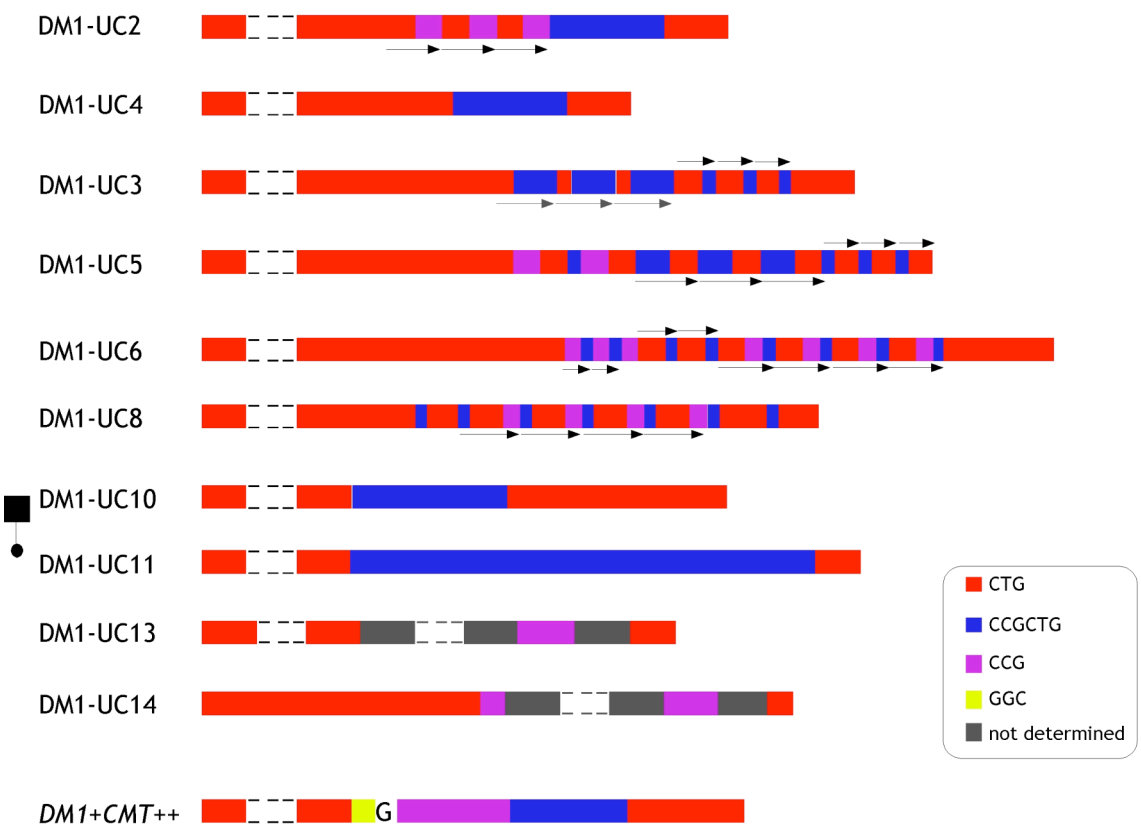


Figure 5-7. Structure of several imperfect CTG repeat alleles detected in DM1 patients with an unusual molecular diagnosis. The structure of each allele was constructed taking into account the data obtained with RP-PCR and sequencing some clones. All unusual cases except DM1-U12 and DM1-UC9 present an imperfect CTG allele at the 3'-end of the repeat tract. The 5'-end seemed to contain long runs of pure CTG repeats. The black arrows indicate higher order repeats containing two different repeat units. The structure of the DM1+CMT++ allele is also indicated. DM1-UC10 and DM1-UC11 are related and the symbols on the left side indicate the relationship (father and foetus).

Table 5-1. Best progenitor allele estimated sequence

Patient	Estimated structure of the *DM1* progenitor allele
DM1-UC1	$(CTG)_{48}$ $(CCGCTG)_{17}$ $(CTG)_{14}$
DM1-UC2	$(CTG)_{222}(CCG)_5((CTG)_5$ $(CCG)_5)_2$ $(CCGCTG)_{22}$ $(CTG)_{14}$
DM1-UC3	$(CTG)_{425}$ $(CCGCTG)_4$ $((CTG)_2(CCGCTG)_4)_2$ $((CTG)_4(CCGCTG))_3$ $(CTG)_{14}$
DM1-UC4	$(CTG)_{318}$ $(CCGCTG)_{19}$ $(CTG)_{13}$
DM1-UC5	$(CTG)_{412}$ $(CCG)_5$ $(CTG)_5$ $(CCGCTG)(CCG)_5(CTG)_5$ $((CCGCTG)_3(CTG)_4)_3$ $((CCGCTG)(CTG)_5)_3$
DM1-UC6	$(CTG)_{516}$ $((CCG)_2$ $(CCGCTG))_2$ $(CCG)_2$ $((CTG)_5(CCGCTG))_2$ $((CTG)_5$ $(CCG)_2$ $(CCGCTG))_4(CTG)_{31}$
DM1-UC7	$(CTG)_5$ $(CCGCTG)_{15}$ $(CTG)_5$
DM1-UC8	$(CTG)_{225}$ $(CCGCTG)$ $(CTG)_7$ $((CCGCTG)$ $(CTG)_7$ $(CCG)_2)_4$ $(CCGCTG)$ $(CTG)_9$ $(CCGCTG)(CTG)_8$
DM1-UC9	$(CTG)_{396}$
DM1-UC10	$(CTG)_{285}$ $(CCGCTG)_{31}$ $(CTG)_{58}$
DM1-UC11	$(CTG)_{128}$ $(CCGCTG)_{119}$ $(CTG)_8$
DM1-UC12	Two normal alleles
DM1-UC13	$(CTG)_x$ $(NNN)_x(CCG)_{10}(NNN)_{11}(CTG)_9$
DM1-UC14	$(CTG)_{50}$ $(CCG)_5(NNN)_x(CCG)_{10}(NNN)_8(CTG)_4$

DM1-UC1 was not assessed by RP-PCR using the CCG_5 repeat specific primer, because there was no DNA left. However, a number of CCG are probably present as in DM1-UC2, suggested by a gap in the CTG_5 repeat specific RP-PCR that is not filled with CCGCTG. In DM1-UC12, two normal alleles with sizes 17 and 5 were genotyped by PCR. Both expanded alleles seemed to amplify well by RP-PCR. It should be investigated why this individual was considered unusual and if this individual presents the DM1 phenotype. The only oddity in DM1-UC9 seemed to be a small 10 bp deletion in the flanking DNA at the 3'-end. DM1-UC13 and DM1-UC14 present other repeat types or arrangements, which remain to be investigated. DM1-UC1 and DM1-UC10 were not investigated by SP-PCR because there was no DNA left; the numbers of repeats were estimated from PCR gels.

5.2.4 Investigating the degree of somatic mosaicism

The structure of the DM1 expanded alleles in the unusual cases has been characterised in all except for DM1-UC13 and DM1-UC14, for which several regions remain to be characterised. The majority of the patients presented imperfect CTG repeat alleles, except for DM1-UC9, which contains a 10 bp deletion at the 3'-end flanking DNA of the CTG repeats. Taking into account the data obtained in other repeat diseases and in the DM1+CMT++ family, it could be postulated that the presence of interruptions in the unusual cases might also stabilise the CTG repeats. However, could the type and/or number of interruptions have a different effect?

To gain an insight into the dynamics of repeats in different imperfect CTG repeat alleles, SP-PCR was performed. More than 100 molecules were amplified and sized for each patient. The length of the progenitor alleles and the degree of somatic mosaicism were determined (figure 5-8 and 5-9). All the different DM1 expanded alleles were somatically unstable and biased toward expansions. DM1-UC13 and DM1-UC14 were investigated by SP-PCR analysis, but the dynamics of the expanded alleles was not quantified. The expanded alleles in both cases did not amplify as well as the normal alleles or the DM1 CTG probe did not hybridise properly. Therefore, it was necessary to optimise the technique, but this optimization was not initiated due to a lack of time.

In an attempt to assess the level of somatic variation of the unusual DM1 expanded alleles, the data was compared with the linear regression analysis between the progenitor allele and the somatic mosaicism in classic DM1 patients (Morales Montero 2006). However, no obvious differences were detected (figure 5-10). It should be pointed out that the age at sampling also has an effect in the degree of somatic mosaicism. Due to the fact that the age at sampling was not available, no further conclusion could be drawn.

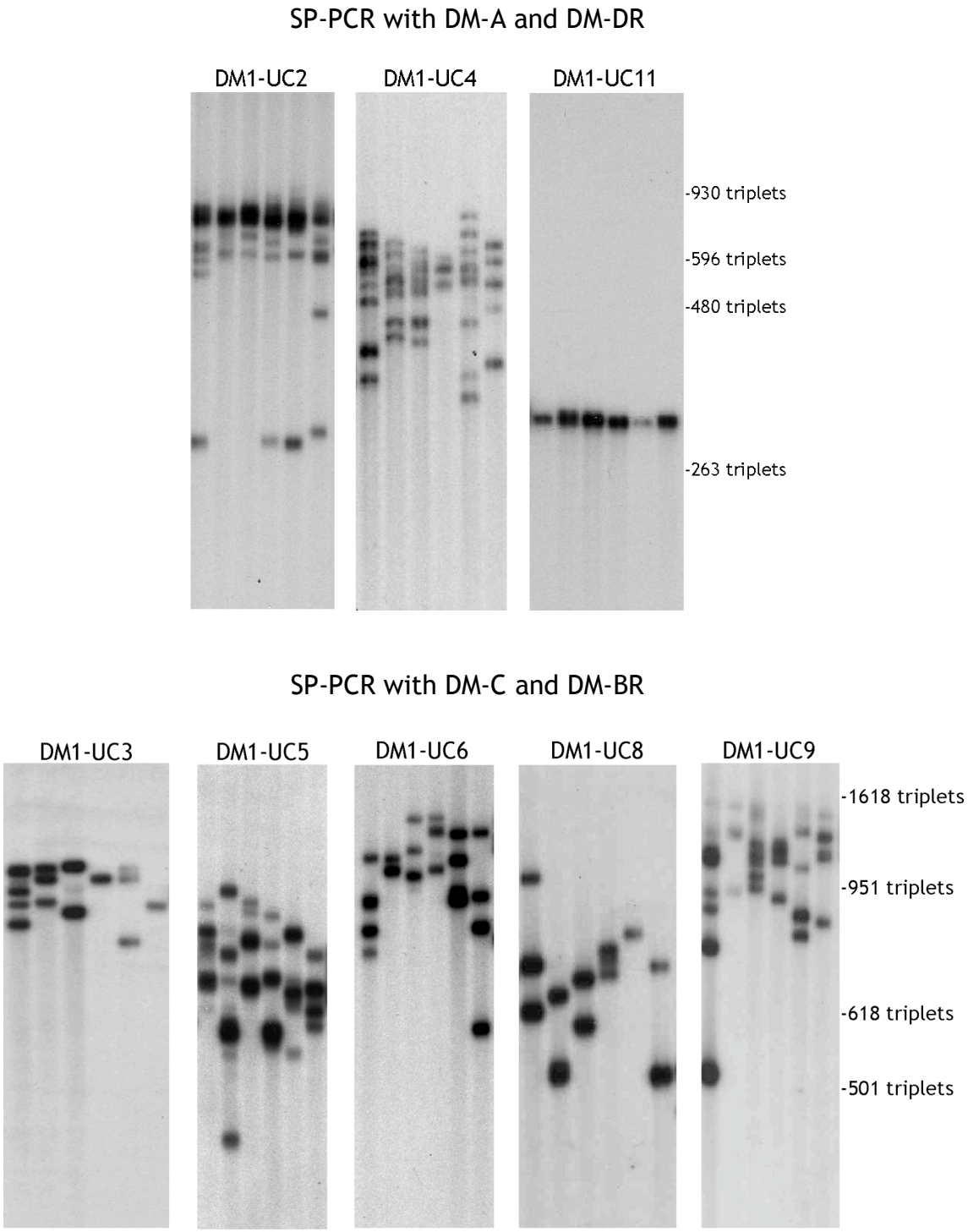


Figure 5-8. Somatic mosaicism of several imperfect CTG repeats alleles. For each sample, 6 representative SP-PCRs each containing ~15 genomic equivalents of DNA are shown. The scale on the right shows the position of the molecular weight marker converted into the number of triplets.

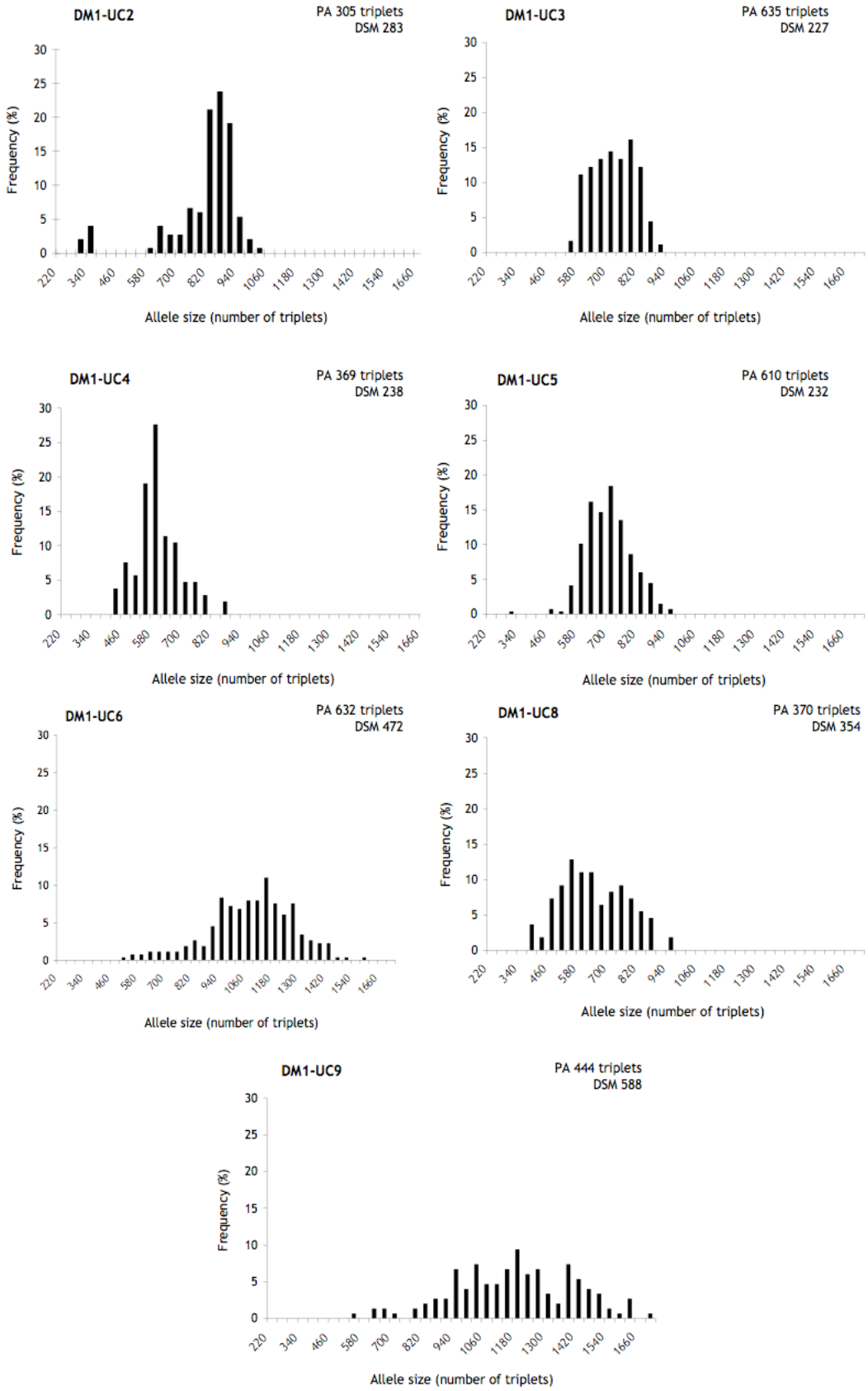


Figure 5-9. Distributions of imperfect CTG repeat alleles in unusual DM1 patients. The progenitor allele (PA) and the degree of somatic mosaicism (DSM) are indicated in each panel. At least 100 expanded molecules were sized for each patient. In DM1-UC2 two different populations of alleles are clearly distinguished. It will be interesting to determine the structure of each allele.

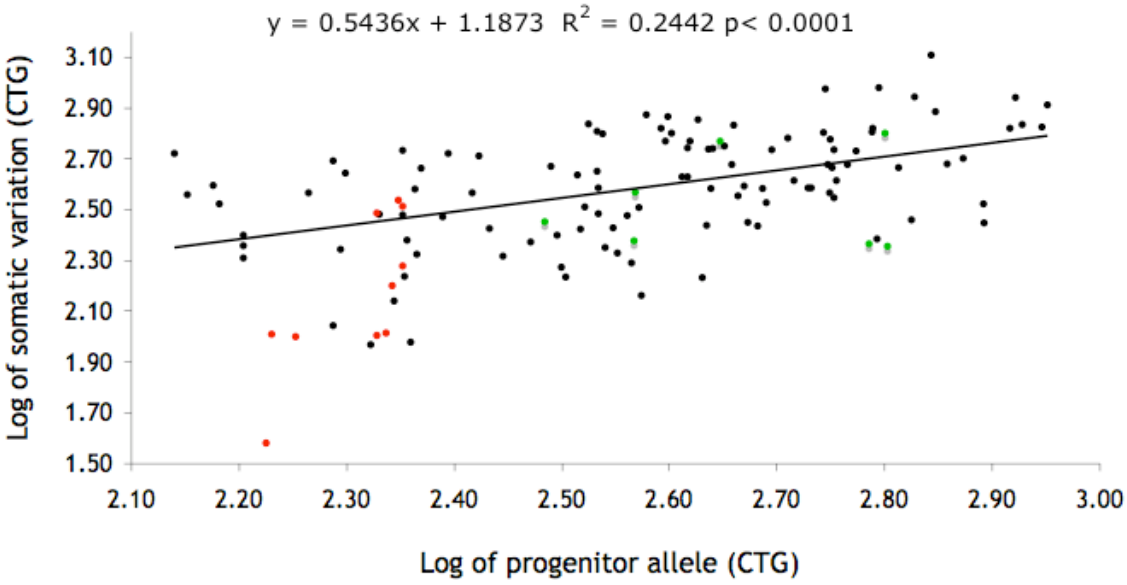


Figure 5-10. Linear regression analysis between the progenitor allele and the somatic variation. Data from classic DM1 patients (Morales 2006) shown in black, data from the DM1+CMT++ cases shown in red and data from the unusual cases shown in green. No obvious differences were observed between the classic DM1 patients and the unusual cases.

5.2.5 Investigating DNA methylation at the *DM1* locus

All the unusual DM1 cases provide new additional CpG methylation sites within the repeat tract, so maybe those sites are methylated. A preliminary attempt to investigate DNA methylation in the flanking DNA at the 3'-end of the repeat tract was performed with the methylation sensitive enzyme *Hha*I. The first *Hha*I site located at the 3'-end of the repeat tract is located at 218 bp distal to the end of the repeat (figure 5-11A). Genomic DNA was digested overnight with *Hha*I enzyme, following PCR amplification with primers DM-C and DM-SR. The *Hha*I site in the normal allele was expected to be unmethylated, so a PCR product should not be detected because the enzyme will be able to digest the site.

As expected, the normal alleles were digested with *Hha*I and an allele was not amplified by PCR. However, it should be pointed out that the normal allele of DM1-UC2, DM1-UC3, DM1-UC6, DM1-UC7 and DM1-UC8 was amplified slightly in the digested sample. The gels were detected by Southern blot hybridisation, which is a very sensitive protocol such that the presence of one molecule would be enough to visualise a band on the autoradiograph. The expanded alleles of DM1-UC2, DM1-UC3, DM1-UC4,

DM1-UC5 and DM1-UC6 seemed to be methylated, since no difference was observed in the intensity of the bands visualised between the digested and undigested control. In contrast, the majority of the expanded alleles of DM1-UC8 and DM1-UC9 seemed to be unmethylated, since the smear present in the undigested control is not reproduced in the digested sample (figure 5-11B). It was expected that the expanded alleles of DM1-UC9 would not be methylated since the size of the CTG repeats is within the adult onset range and the only oddity was a small deletion at the 3'-end. This data was consistent with the study of Steinbach and collaborators, where only the expanded alleles with sizes ≥1,000 repeats were found to be methylated in the flanking DNA (Steinbach*, et al.* 1998). It would be interesting to extend the study and analyse the pattern of DNA methylation both within the repeat tract and at more sites at the 3'-end in the flanking DNA. The expanded alleles of DM1-UC11, DM1-UC13 and DM1-UC14 failed to amplify in the undigested samples. Due to lack of time, these samples were not further investigated.

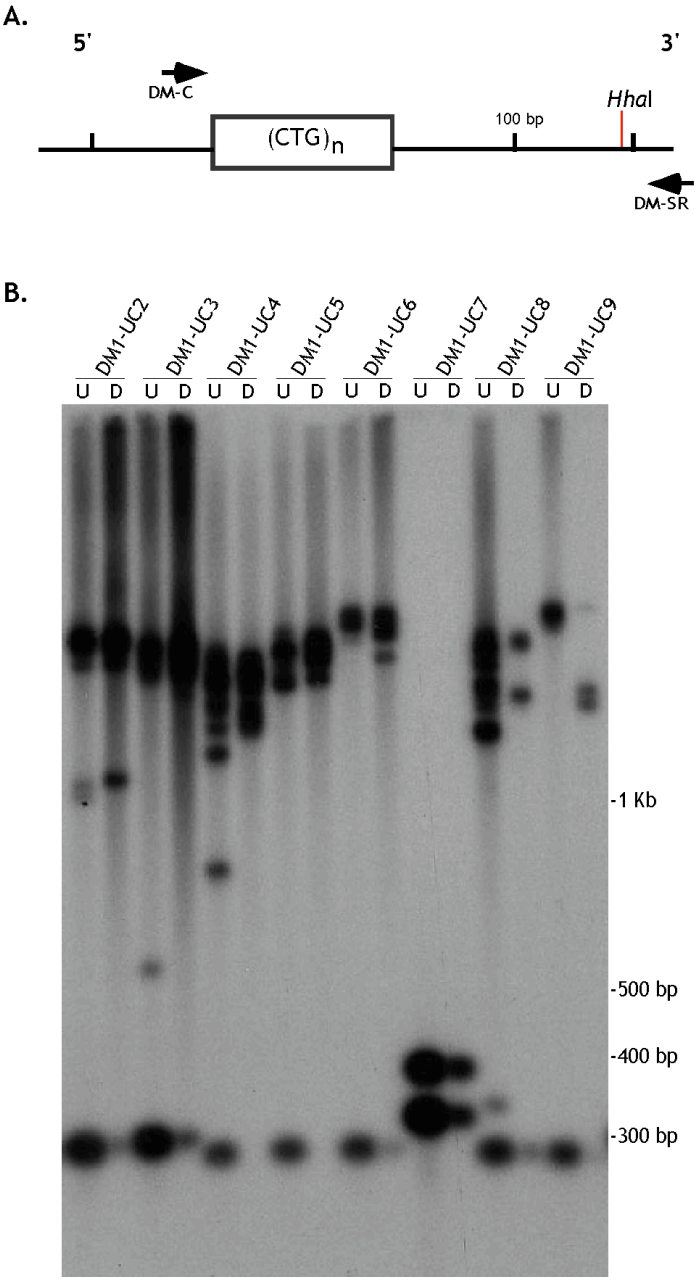


Figure 5-11. Methylation pattern at the 3'-end of the repeat tract. A. Schematic diagram of the *DM1* locus indicating the position of the primers and the *Hha*I site investigated. B. Genomic DNA was digested overnight with *Hha*I. The digested and undigested products were amplified with primers DM-C and DM-SR. The products were resolved on a 1.5% agarose gel and detected with the DM1 CTG repeat probe. The size of the marker is indicated on the right of the panel.

5.3 Discussion

The modified PCR in the presence of 10% DMSO was successful in amplifying the expanded alleles, which previously were not amplified by the standard reaction, except for the expanded allele of DM1-UC14. The combination of RP-PCR with specific primers and sequencing was extremely successful in characterising the structure of the DM1 expanded allele in a number of

unusual DM1 patients. The sequencing data verified the sequence estimated from the RP-PCR gels, indicating that the primers were extremely accurately. The majority of the patients presented with an imperfect CTG repeat allele containing a long tract of CTG repeats at the 5'-end, followed by CCGCTG hexamers and/or CCG repeats, and ending with a small number of CTG repeats at the 3'-end. Additionally, DM1-UC2, DM1-UC3, DM1-UC5 and DM1-UC6 contained two or three higher order repeats units. It would be interesting to investigate the mechanism responsible for generating these higher order repeat units. Two patients, DM1-UC13 and DM1-UC14, seem to present another type of repeat unit and/or arrangement of repeats that still remains to be investigated.

The intergenerational instability in the imperfect CTG repeat allele from DM1-UC1 and DM1-UC10 was investigated. The expanded allele of DM1-UC10 consists of $(CTG)_{128}$ $(CCGCTG)_{31}$ $(CTG)_{58}$. The foetus of DM1-UC10 inherited an increased number of CCGCTG from 31 to 119, but a reduction in the CTG repeats from 58 to 8 at the 3'-end and from 128 to 105 at the 5'-end. It should be noted that the increase in CCGCTG is very dramatic in comparison to what was observed in the DM1+CMT++ allele, where the CCGCTG repeats were stably transmitted. Moreover, it is not comparable to what is observed in pure CTG repeats alleles of the same size. Martorell and colleagues investigated 700 DM1 families and found that alleles up to 25 repeats were always stably transmitted (in 1,043 transmissions), 26-37 repeats gain 1 to 3 repeats (in 84 transmissions), while repeat alleles between 38-54 repeats showed dramatic increases in size in the majority of males transmissions with gains between 45 to 1120 repeats. Small changes were observed in females transmissions (gains of 3 to 7 repeats) and in some male transmissions (gains of 3-15 repeats) (Martorell, *et al*. 2001). The fact that DM1-UC10 is a male and the CCGCTG repeats are surrounded by CTG repeats could explain the instability of that allele. It may be that the threshold is 31 repeats and not 38 repeats as in pure CTG arrays. The structure of the imperfect CTG repeat allele of DM1-UC1 has not been completely determined, however the level of intergenerational instability seemed to be similar to the DM1+CMT++ allele. An increase in the CTG repeat at the 5'-end was detected, while the interrupted 3'-end seemed to

be only slightly modified (gain of 5 CCGCTG repeats). A more in depth interpretation could be performed once the structure for DM1-UC1 is studied in detailed.

The level of somatic mosaicism was investigated in DM1-UC2, DM1-UC3, DM1-UC4, DM1-UC5, DM1-UC6, DM1-UC8 and DM1-UC9. In all cases the alleles were somatically unstable. However, no investigation was undertaken into which part of the repeat tract was unstable. From the RP-PCR, it could be suggested that the interruptions seem to be quite stable since discrete bands are detected, while the CTG repeats at the 5'-end were detected as smears as is consistent with somatic mosaicism. Therefore, the CTG repeats at the 5'-end seemed to be unstable as in the DM1+CMT++ allele. This interpretation could be easily verified by using single molecule separated alleles and RP-PCR analysis.

Investigation of the mutation mechanisms of minisatellites revealed a gain of few repeats units at one end of the tandem array, suggesting a polarity effect. The authors postulated that the polarity effect could be caused by an element outside the array (Jeffreys, *et al*. 1994). It is of interest to consider that the interruptions identified in the DM1 unusual cases and in the DM1+CMT++ allele were always detected at the 3'-end of the repeat tract and the unstable part of the array seem to be at the 5'-end. These results suggest a mutational polarity as observed in minisatellites. Further investigations on the flanking DNA at both ends of the repeat array will help to identify putative *cis*-acting modifiers.

The pattern of methylation at the 3'-end seemed to depend on the number and type of interruptions. Only one single site was investigated at the 3'-end, but differences were observed. The expanded alleles of DM1-UC2, DM1-UC3, DM1-UC4, DM1-UC5 and DM1-UC6 were methylated, whereas the expanded alleles of DM1-UC8 and DM1-UC9 were unmethylated. Further investigations of the methylation pattern at the 3'-end flanking could reveal possible effects on the stability of the repeats.

A correlation with the genotype and the clinical symptoms in the unusual cases would be extremely interesting to determine the effect of interruptions on the pathogenesis. It could be postulated that the interruptions might decrease the severity of the symptoms. From an extended study performed by Morales and colleagues, it was estimated that the average progenitor allele associated with adult-onset patients is ~324 repeats, while progenitor alleles with 485 repeats or more would be associated with child-onset or congenital forms. If the unusual cases had pure CTG repeat tracts, we could predict that DM1-UC3, DM1-UC4, DM1-UC5, DM1-UC6, DM1-UC8 and DM1-UC9 have either the child-onset or congenital onset forms of the disease since all of them presented progenitor alleles with more than 485 repeats. However, the presence of imperfect CTG alleles in these patients might decrease the severity of symptoms and they might have an adult-onset form of the disease instead. Alternatively, it could be postulated that the presence of interruptions might alter the phenotype of the disease. In SCA2, intermediate size interrupted alleles are associated with Parkinsonism, while pure alleles are associated with cerebellar ataxia (Charles, 2007). Similarly, in the DM1+CMT++ family, the imperfect CTG allele seemed to be the cause of CMT and the other symptomatology. Therefore, it could be postulated that a modified phenotype should be expected in the unusual cases with interruptions.

6 Discussion

Myotonic dystrophy type 1 is the most common form of muscular dystrophy in adults and is caused by an expansion of a $(CTG)_n$ repeat located in the 3'-untranslated region (UTR) of the *DMPK* gene (Aslanidis, *et al*. 1992; Brook, *et al*. 1992; Buxton, *et al*. 1992; Harley, *et al*. 1992). Normal individuals contain between 5 and 35 CTG repeats, which are polymorphic and relatively stable. However, if under certain circumstances the number of repeats increases, the repeats become dramatically unstable showing a higher mutation rate. Longer alleles generally always increase in size when transmitted from parent to child, and consequently the child will present anticipation characterised by earlier age of onset and an increase in the severity of the symptoms (Harper 2001). Nevertheless, in a small percentage (< 10%) of transmissions, contractions occur (Ashizawa, *et al*. 1994a) and occasionally these contractions can result in reverse mutations (Shelbourne, *et al*. 1992; Brunner, *et al*. 1993b; Hunter, *et al*. 1993; O'Hoy, *et al*. 1993; Shelbourne, *et al*. 1993). Interestingly, in one of the cases of reverse mutation, a gene conversion event was suggested (O'Hoy, *et al*. 1993). The effects of long CTG repeats on the flanking DNA were investigated in *E. coli* and it was demonstrated that long repeats promote deletions and inversions in the flanking DNA, suggesting that similar mechanisms could also be present in DM1 patients (Wojciechowska, *et al*. 2005).

The DM1+CMT++ family is a very unusual three-generation family in which all the patients co-segregated both DM1 and CMT with the *APOC2* on chromosome 19 with a maximum LOD score of 7.03 and zero recombination. However, both disorders failed to segregate with chromosome 1 or 17. It was suggested that DM1 and CMT might be caused by a single or two closely linked mutations near the *APOC2* marker (Spaans, *et al*. 1986; Brunner, *et al*. 1991). A fragment equivalent to a small CTG expansion (~200-400) at the *DM1* locus was detected in the patients by Southern blot analysis of restriction digested genomic DNA, but not by PCR. Recently, in addition to the DM1 and CMT symptoms a number of patients

developed hearing loss, migraine and recurrent episodes of acute encephalopathy (Spaans*, et al.* 2008). We postulated that in these patients the expanded repeats may have predisposed the repeat tract and the flanking regions to further DNA instability, leading to a secondary deletion, insertion and/or rearrangement. The expression of *DMPK* and/or nearby genes might be modified by these novel mutations explaining the unusual clinical presentation observed in the DM1+CMT++ family. In order to identify the molecular lesion in the DM1+CMT++ family, a variety of molecular approaches were performed.

6.1 The molecular lesion in the DM1+CMT++ family

Previously, it was shown that in the DM1+CMT++ family small expansions were detected by Southern blot analysis of restriction digested genomic DNA, but it was not possible to amplify across the array using the standard approach. This data was confirmed and further investigation of the flanking DNA by RP-PCR showed an expanded allele at the 5'-end, but was negative at the 3'-end. These data suggested that the additional modification must be located at the 3'-end. The most likely explanation was a deletion, which could have disrupted one or more genes causing the complex symptomatology observed in the DM1+CMT++ family. Therefore, the investigation was mainly focused on finding a deletion at the 3'-end. However, genotyping several polymorphisms flanking the CTG repeat revealed that several individuals presented as heterozygous for many SNPs giving evidence against a deletion. Additionally, haplotype analysis revealed that the DM1+CMT++ allele was found on the classic DM1 haplotype.

At this point, either a rearrangement or an insertion were the two remaining possibilities. A breakpoint at the 3'-end was investigated by vectorette PCR. Patients' vectorette libraries generated with a variety of four, five and six-cutter restriction enzymes failed to detect an expanded allele in the DM1+CMT++ family, however the technique was working successfully with classic DM1 patients. An attempt with *Aci*I and *Hha*I, enzymes with GC rich recognition sequences, was performed and finally the

expanded alleles were detected as smears in the DM1+CMT++ family, consistent with somatic mosaicism. The DNA from the smear was cloned and sequenced. The sequence of the "38" repeat allele (non-pathogenic) of III-9 was obtained and consisted of $(CTG)_{5-6}$ $(CCGCTG)_{14}$ $(CTG)_{5-6}$. Further investigations with IV-11 and IV-12 revealed that they both inherited the same allele. Additionally, an incomplete sequence of an expanded allele from III-16 revealed that it consisted of at least one hexamer (CCGCTG), $(GGC)_2$ and a *Hha*I site within the CTG tract. These results suggested that the failure to PCR across the repeat tract might be due to an increase of the GC content within the CTG tract. To test this possibility, a modified PCR in the presence of 10% DMSO was performed. The expanded alleles were finally visualised, confirming that the additional modification was indeed an insertion and not a rearrangement.

6.2 The structure of the DM1+CMT++ allele

The next step in the project was the characterisation of the interruptions in the DM1+CMT++ allele. It was not an easy task, but at the end the combination of cloning and sequencing, restriction mapping and RP-PCR together with the use of single molecule separated alleles helped us reveal the structure of the DM1+CMT++ allele. The allele was comprised of a variable number of CTGs at the 5'-end followed by $(GGC)_3$ G $(CCG)_{20}$ $(CCGCTG)_{14}$ $(CTG)_{35}$. Analysis of single molecule separated alleles revealed that the interrupted 3'-end of the array was stable in blood cells of both the same patients and between the patients investigated. In contrast, the CTG repeats at the 5'-end were unstable. These results suggested the interrupted 3'-end was stable, while the CTG repeats at the 5'-end were unstable in the soma and in the germline.

Imperfect CTG repeat alleles were previously reported in four individuals, but the presence of those alleles were considered an unique event in DM1 (Leeflang and Arnheim 1995; Cockburn, *et al*. 2003). Therefore, finding imperfect CTG repeat allele in the DM1+CMT++ family was not expected. These results led us to suggest that imperfect CTG repeat alleles may not be unique events as previously considered and other DM1 patients may also

contain imperfect alleles. Jean-Louis Mandel and colleagues identified a number of DM1 sporadic patients with an unusual molecular diagnosis, which could also be carriers of imperfect CTG repeats alleles.

6.3 Imperfect CTG repeat alleles present in several DM1 patients with an unusual molecular diagnosis

DNA samples from 14 DM1 patients with an unusual molecular diagnosis were investigated in order to further characterise the mutation. The majority of the patients presented with an imperfect CTG repeat allele containing CCGCTG hexamers and/or CCG repeats (figure 5-7). Two patients, DM1-UC13 and DM1-UC14, seem to present another type and/or arrangement of repeats, which still remains to be investigated. DM1-UC7 presented a "41" repeat allele consisting of $(CTG)_6\,(CCGCTG)_{15}\,(CTG)_5$, a structure very similar to the "38" repeat allele of III-9, IV-11 and IV-12. DM1-UC2, DM1-UC3, DM1-UC5 and DM1-UC6 contained two or three higher order repeats containing between 18 and 30 bp such as $((CTG)_5\ (CCG)_5)$, $((CTG)_2\,(CCGCTG)_4)$ and $((CTG)_5\,(CCG)_2\,(CCGCTG))$ (table 5-1). DM1-UC9 presented a pure CTG repeat array with a small deletion of 10 bp at the 3'-end flanking DNA. These findings further suggest that imperfect CTG repeat alleles might not be a rare event, as was believed. It is not known yet how many people were screened to identify these unusual cases so the frequency of imperfect CTG repeat alleles was not estimated.

6.4 The repeat dynamics of CTG repeat alleles

The repeat dynamics of pure CTG repeats have been extensively studied in humans and in DM1 mouse models. Normal alleles containing up to 37 CTG repeats are generally stable, however alleles with more than 50 CTG repeats are highly unstable. Increases in the length of these unstable alleles are observed when transmitted from parent to child, resulting in anticipation (Harper 2001). As the number of CTG repeats increases, the alleles become even more unstable. The sex of the parent has an effect on

the instability; short expansions are generally more unstable when transmitted by the father, while long expansions are more unstable when transmitted by the mother. The congenital cases are almost always transmitted by adult-onset mothers (Brunner, *et al*. 1993a; Harley, *et al*. 1993; Lavedan, *et al*. 1993; Ashizawa, *et al*. 1994b; Jansen, *et al*. 1994).

Somatic mosaicism has been investigated in different tissues and in repeat samples obtained over one and seven years. Expanded alleles are always longer in skeletal muscle than in lymphocytes (Anvret, *et al*. 1993; Ashizawa, *et al*. 1993; Thornton, *et al*. 1994; Monckton, *et al*. 1995) and always increase in length with the age of the individual (Wong, *et al*. 1995; Martorell, *et al*. 1998). Therefore, it was postulated that the somatic mosaicism in pure CTG repeats is expansion-biased, age-dependant and tissue-dependant.

6.4.1 Germline instability of imperfect CTG repeat alleles

There was no previous knowledge regarding the repeat dynamics of imperfect CTG repeat alleles, so our next aim was to investigate the dynamics of imperfect CTG repeats by SP-PCR.

III-16 and III-17, both adult-onset mothers, transmitted the DM1+CMT++ allele to their sons and daughters, providing an opportunity to investigate the germline instability of imperfect CTG repeat alleles. Adult-onset mothers carrying 100 CTG repeats have a 62% chance of transmiting a large expansion to their child resulting in the congenital form of the disease (Redman, *et al*. 1993). Despite the high risk of transmitting long alleles, both III-16 and III-17 transmitted small increases in the number of CTG repeats (~43 CTG repeats) at the 5'-end on two occasions (IV-19, IV-20 and IV-21 and IV-22, respectively). These results provided the molecular explanation for the slight anticipation observed between the third and fourth generation.

Both individuals (I-6 and I-9) in the first generation were asymptomatic. However, in the second generation, five members presented symptoms of

DM1 and CMT. A decrease in the age of onset and a reduction in the severity of the DM1 symptoms were observed between the second and third generations. Unfortunately, there is no DNA available to investigate the *DM1* locus in the second or first generation. However, it could be postulated that the number of CTG repeats must have increased between the first, second and third generations, respectively. The progenitor alleles of III-9, III-16 and III-17 had different sizes. III-9 inherited 225 repeats, III-16 inherited 170 repeats, and III-17 inherited 179 repeats. These data suggest that the alleles were not stably transmitted. It could be postulated that either I-6 or I-9 must have carried an intermediate size allele. Although the size may have been insufficient to cause a phenotype, the allele could have been unstable in the germline. Assuming that the dynamics of imperfect CTG repeat alleles are similar to the dynamics of pure alleles, probably the father (I-1) carried an interrupted 3'-end allele containing less than 50 CTGs at the 5'-end, since shorter alleles are more unstable when transmitted by males.

A polarity effect has been observed; the CTG repeats at the 5'-end of the repeat array were always unstable, whereas the interrupted 3'-end was stable through transmissions. Similar polarity effects were described for fragile X syndrome (Eichler*, et al*. 1994; Zhong*, et al*. 1995; Kunst*, et al*. 1997), SCA2 (Choudhry*, et al*. 2001) and on minisatellites (Jeffreys*, et al*. 1994). In fragile X syndrome, the polarity is observed at the 3'-end of the repeat tract (relative to transcription orientation), while the instability in SCA2 is observed at the 5'-end. In this project, the polarity effect was observed at the 5'-end as in SCA2. A flanking element was suggested to cause a polarity effect. Two possible mechanisms could mediate the effect. One of these mechanisms postulates that the flanking effect is mediated by an element that causes double strand breaks at a distal site (Jeffreys*, et al*. 1994). The second mechanism postulates that the position of the nearest replication origin is the flanking effect, which causes a difference in the mutation rates between the leading and lagging strand containing long repeat arrays during replication (Richards and Sutherland 1994). In the DM1+CMT++ allele, the presence of interruptions at the 3'-end are probably the flanking element causing a decrease in the instability of the CTG

repeats at the 5'-end. Recently, we obtained a sperm sample from IV-21, so investigations into the repeat dynamics in the sperm DNA will provide a better understanding of the characteristic of germline instability in imperfect CTG repeat alleles.

A primary characterisation of the dynamics of germline instability in two unusual cases was performed. The expanded allele of DM1-UC10 consists of $(CTG)_{128}$ $(CCGCTG)_{31}$ $(CTG)_{58}$, however DM1-UC11 inherited a different allele $(CTG)_{105}$ $(CCGCTG)_{119}$ $(CTG)_{8.}$The CTGs at both ends were reduced and the CCGCTG repeats increased dramatically. This finding is in contrast to the DM1+CMT++ allele, where the interrupted end was stably transmitted and only slight increases in the CTG repeats at the 5'-end was observed. Pure CTG repeat alleles containing 31 repeats are relatively stable, however it could be possible that the threshold for CCGCTG repeats is lower. Alternatively, the presence of CTG repeats surrounding the CCGCTG repeats could have an effect on the instability. DM1-UC1 contains an imperfect CTG repeat allele, which probably consist of $(CTG)_{48}$ $(CCGCTG)_{17}$ $(CTG)_{14}$. The structure of the imperfect CTG repeat allele of DM1-UC1 has not been completely verified because there was no more DNA left. DM1-UC1 is the mother of DM1-UC2 and she transmitted an imperfect allele to her son. The expanded allele of DM1-UC2 consist of $(CTG)_{222}$ $(CCG)_5$ $((CTG)_5$ $(CCG)_5)_2$ $(CCGCTG)_{22}$ $(CTG)_{14}$. It could be suggested that the level of intergenerational instability is similar to the DM1+CMT++ allele. Apparently, the CTG repeat at the 5'-end increases, while the interrupted 3'-end seemed to be only slightly modified (a gain of 5 CCGCTG repeats). The presence of CCG repeats were not investigated in DM1-UC1, but the RP-PCR data with CTG and CCGCTG repeat specific primers suggest that other repeats units are present between the CTG repeats at the 5'-end and the CCGCTG repeats. Future investigations will reveal the complete structure of DM1-UC1 and a better analysis will be performed.

6.4.2 Somatic instability of imperfect CTG repeat alleles

The CTG repeats at the 5'-end in the DM1+CMT++ allele in blood cells were somatically unstable, biased toward expansions and age-dependant. The

analysis of repeat blood samples from IV-21 with a time interval of 8 years revealed a statistically significant increase in the level of somatic variation (figure 4-14). However, no statistically significant differences were observed in repeat samples from III-9 and III-16 over 3 and 4 years, respectively. IV-21 was sampled at 22 and 30 years old; in contrast III-9 and III-16 were sampled after 50 years old. It could be possible that the effects were not as dramatic as IV-21 because of their age. To investigate this further, it would be interesting to analyse the instability of III-9 and III-16 over a longer time period. Nevertheless, it could be postulated that the age effects are probably not as dramatic as in pure CTG repeats. This hypothesis could be further investigated if several repeat samples are obtained.

To determine if the interrupted 3'-end affects the degree of somatic mosaicism of the CTG repeat at the 5'-end, the data from the DM1+CMT++ cases was compared with a previous study. Morales performed a study in more than one hundred DM1 patients and revealed that ~66% of the variation in the somatic mosaicism is caused by complex interactions between the progenitor allele length and the age at sampling (General Linear Model Anova, $r = 0.81$, $p = 0.005$) (Morales 2006). The somatic mosaicism in the DM1+CMT++ was estimated using the model obtained by Morales. The values estimated were always higher than the values observed, except for IV-19 and IV-22 from whom a small decrease was obtained (table 4-3). These results suggested that the presence of interruptions at the 3'-end of the repeat array probably decreases the instability of the CTG repeats at the 5'-end.

The level of somatic mosaicism in blood cells was investigated in DM1-UC2, DM1-UC3, DM1-UC4, DM1-UC5, DM1-UC6, DM1-UC8 and DM1-UC9. In all cases, the alleles were somatically unstable. However, it was not determined which part of the repeat tract was unstable. From the RP-PCR, it could be suggested that the interruptions seem to be quite stable since discrete bands are detected, while the CTG repeats at the 5'-end were detected as smears on bulk DNA consistent with somatic mosaicism. Therefore, the CTG repeats at the 5'-end seemed to be unstable as in the

DM1+CMT++ allele. This interpretation should be verified by using single molecule separated alleles and RP-PCR analysis.

6.5 Possible mechanisms of genetic instability in imperfect CTG alleles

The finding of a number of DM1 patients with imperfect CTG repeats alleles lead us to ask several questions. How do higher order repeats expand? How do pure CTG repeat arrays acquire novel repeat units and how do the novel repeats spread? Can the imperfect CTG repeat alleles be traced to an ancestral mutation, as is observed for the pure CTG repeat alleles, or have the interruptions arisen by multiple *de novo* mutations?

The mechanisms of genetic instability of triplet repeats are not yet completely understood. Initially, it was proposed that the expanded repeats form unusual alternative structures, which may interfere with the normal cell processes such as replication, repair and transcription. The presence of these unusual structures in the DNA has been demonstrated *in vitro*, but there is still no evidence that these structures are formed *in vivo* (Wells 1996; McMurray 1999; Sinden, *et al*. 2002). Replication slippage was one of the favourites mechanism postulated. However the fact that cells that do not undergo replication (i.e. muscle and brain) present sometimes higher levels of instability lead to postulate that probably replication slippage was not the principal mechanism involved (Anvret, *et al*. 1993; Ashizawa, *et al*. 1993; Thornton, *et al*. 1994; Kennedy, *et al*. 2003). Subsequently, the mechanism of mismatch repair becomes the most plausible explanation. Investigations with knock-out mouse models revealed that indeed MSH2, MSH3 and PMS2, proteins involved in the mismatch repair might be directly involved in the generation of instability (Manley, *et al*. 1999; van den Broek, *et al*. 2002; Savouret, *et al*. 2003; Wheeler, *et al*. 2003; Gomes-Pereira, *et al*. 2004b). Mismatch repair is involved in repairing insertion-deletion loops outs up to 12-16 bp (Bellacosa 2001; Schofield and Hsieh 2003) and single base-base mismatches. However, alternatives mechanisms should be involved in generating the instability of

higher order repeats containing between 18 and 30 bp in DM1-UC2, DM1-UC5, DM1-UC6, DM1-UC8 (figure 5-7).

The size of the higher order repeats observed in the patients is comparable to minisatellites so probably we could infer that the mechanisms involved in the instability of minisatellites might cause the instability. Postulated mechanisms involved in genetic instability of minisatellites are: gene conversion and crossover events between sister chromatids (Jeffreys, *et al*. 1994). Studies on S. *cerevisae* have revealed that large loops mismatches are mainly repaired by a mechanism dependant on mismatch repair proteins (MSH2 and PMS1) (Clikeman, *et al*. 2001). In contrast studies on human nuclear extracts revealed that the mechanism is independent of mismatch repair and is directed by a single strand break at the 5'-end of the loop (Littman, *et al*. 1999). It is not clear if the loops are retained, lost or maybe is an equilibrium between the two. A polarity effect has been noticed in the instabilities of imperfect CTG repeat alleles, therefore the repair mechanism independent of mismatch repair directed by the 5'-end strand break seems a plausible mechanism to explain how the higher order repeats expand.

It will be interesting to analyse how pure CTG repeats acquire novel repeats. In fragile X syndrome interruptions are present in normal alleles and then the interruptions are lost when the alleles expand. It was postulated that either a deletion or a transversion event could explain the loss of interruptions (Eichler, *et al*. 1994). It could be postulated that a similar mechanism could explain the presence of novel repeats. A base substitution from T to C could be generated. Three possible mechanisms could cause base substitutions: a polymerase error during replication or during mismatch repair or by chemical damage. Long CTG repeats are probably more prone to errors than other regions because: are difficult regions for the polymerase to process, and can form secondary structures, which can be more sensitive to chemical damage. Another aspect of the instability is how can we explain the substitution spreading to generate several copies of the novel repeat units? Smith postulated that higher order repeats containing a mixture of different repeat units could evolve by

unequal crossover between sister chromatids (Smith 1976). Misalignment could occur between sister chromatids, via double-strand breaks, chromatid invasion, and intraallelic gene conversion. A series of these events could result in sequence homogenization, generating the spreading of the novel repeats detected in the patients (figure 6-1).

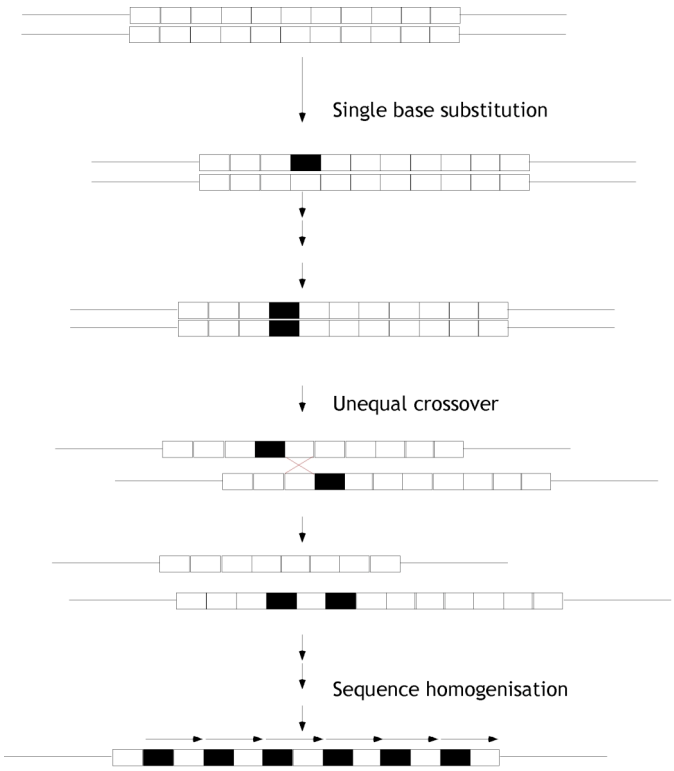


Figure 6-1. Possible mechanism involved in the spreading of novel repeat units. A single base substitution must occur first. Subsequently, unequal crossover between sister chromatids could result with time in sequence homogenisation. A similar mechanism could be involved in generating the array of novel repeat units detected in the patients.

The DM1 mutation is associated with a conserved haplotype. Nine polymorphic sites were investigated to define the haplotypes in III-9, IV-11, IV-12, III-16, IV-19, IV-20, III-17 and IV-21. It was revealed that the DM1+CMT++ allele was present on the DM1 classic conserved haplotype (figure 3-11). Similarly, analysis revealed that the imperfect CTG repeat alleles of DM1-UC1, DM1-UC2, DM1-UC4, DM1-UC11, DM1-UC13 and DM1-UC14 (data not shown) are also associated with the classic DM1 haplotype. There are two possibilities the imperfect CTG alleles and the pure CTG alleles diverged on two different haplotypes or on the same more recently evolved derived haplotype. An analysis of several distal polymorphisms will reveal if there is an extended haplotype or not (figure 6-2).

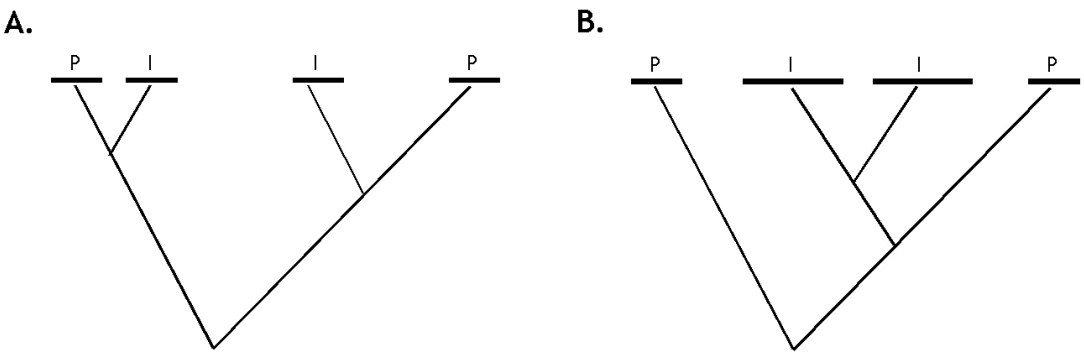


Figure 6-2. Two possible phylogenetic trees. Horizontal lines represent the size of the haplotypes, P (pure CTG repeats alleles), I (interrupted CTG repeat alleles) A. Interruptions are generated by de novo mutations, there are no extended conserved haplotypes. B. Interruptions evolving on a common haplotype, patients with interruptions will present more extended haplotypes than patients with pure CTG repeats.

In this project two size ranges of imperfect CTG repeat alleles were identified: non-pathogenic alleles containing up to "40" triplet repeats and pathogenic alleles with more than "150" triplet repeats. How can we explain the formation of these two range of repeats? It could be postulated that the normal imperfect CTG repeat alleles expand to generate the expanded alleles or alternatively the expanded alleles suffer contractions originating the normal alleles. In pure CTG repeats, normal alleles expand in a process of multiple steps generating expanded alleles. In contrast, in imperfect CTG repeat alleles the most likely explanation seems to be a contraction. The structure of the normal imperfect allele consists of $(CTG)_5$ $(CCGCTG)_{14}$ $(CTG)_5$. The 5 CTG repeats at both ends are extremely stable and to date no expansions have been detected. The 14 CCGCTG repeats seem to be relatively stable, no changes were detected in the patients investigated in this project. Additionally, the stability of the complete structure was demonstrated to be comparable to a "27" pure repeat allele (Leeflang and Arnheim 1995). Taking into account all this, there is no reason to believe that this allele will be unstable. In the patients investigated the pathogenic alleles contain long CTG repeats at the 5'-end, so the most plausible explanation is that long alleles acquire interruptions at the 3'-end and subsequently the 5'-end suffer a contraction, as is has been described for pure CTG repeat alleles (Ashizawa, *et al*. 1994a). Therefore, the non-pathogenic imperfect alleles are most likely secondary products of long interrupted alleles.

6.6 Possible mechanisms of pathogenesis in the DM1+CMT++ family

The molecular lesion causing the complex phenotype observed in the DM1+CMT++ family seems to be an imperfect CTG repeat allele. It could be argued that maybe the CMT is caused by a mutation in *ARCMT2B* locus or *PRX* or *DNM2* genes associated with three forms of CMT also located in the chromosome 19. However, it will be extremely unlikely that 14 affected individuals carry an imperfect CTG allele together with a mutation in one of these loci. Additionally, a two-mutation model would not be able to explain the absence of symptoms in the first generation. Therefore, we are convinced that the only molecular lesion in the family is an imperfect CTG repeat allele. A confirmation of our theory will not be obtained unless a screen for mutations is performed in the whole linked region, but at the moment it will not be cost-effective.

One of the questions that arise is how an imperfect CTG repeat allele could cause a complex symptomatology. It could be assumed that the expanded CTG repeats are responsible for the DM1 symptoms by an RNA gain-of-function mechanism as in classic DM1 patients. However, how the presence of an imperfect CTG repeat allele could also be responsible for the CMT, hearing loss and acute encephalopathy symptoms is not known. Possible mechanisms include a novel RNA gain-of-function, and/or a novel effect on the downstream genes.

The DM1+CMT++ allele contains an interrupted 3'-end containing GGC, CCG and CCGCTG repeats. The presence of these novels repeat units in the *DMPK* transcript probably alters its secondary structure and in consequence novel proteins may be trapped and a novel phenotype could be expressed by analogy to what happens in fragile-X associated tremor/ataxia syndrome (FXTAS). FXTAS is a neurodegenerative disorder caused by premutation alleles (55<CGG<200) in the *FMR1* gene. The phenotype and the pathogenesis are completely different to fragile X syndrome (Hagerman, *et al*. 2001; Berry-Kravis, *et al*. 2003; Jacquemont, *et al*. 2003; Leehey, *et al*.

2003; Hagerman and Hagerman 2004). An RNA-gain of function model similar to the one postulated for DM1 and DM2 was postulated (Willemsen, *et al*. 2003). Investigations in FXTAS patients biopsies revealed the presence of ubiquitin-positive intranuclear inclusions containing *FMR1* mRNA in neurons and astrocytes of patients (Greco, *et al*. 2002; Tassone, *et al*. 2004). Proteins associated with different forms of CMT such as a number of neurofilaments and lamin A/C proteins were identified in the inclusions. In addition MBNL1 one of the RNA binding proteins interacting with the CNG repeat, which is recruited in DM1 and DM2 nuclear foci also co-localised in the intranuclear inclusions (Iwahashi, *et al*. 2006).

In *Drosophila* it was shown that 90 CCG repeats or 90 CGG repeats cause a neurodegenerative phenotype with nuclear inclusions similar to the phenotype caused by CGG repeats in FXTAS (Sofola, *et al*. 2007). It could be postulated that the CCG repeats present in the DM1+CMT++ allele might recruit neurofilaments and lamin A/C proteins explaining the CMT symptoms observed in the DM1+CMT++ family. However, the number of CCG repeats in the DM1+CMT++ allele is less than the threshold (55 CGGs) in FXTAS patients. So how can we explain that 20 CCGs will be sufficient to cause a phenotype? It is not known why at least 55 CGGs are required to have deleterious consequences in the cells, maybe this reflect a threshold for the transcript to be trapped in the nucleus. In DM1 it was demonstrated that the overexpression of the 3'-UTR of the *DMPK* transcript containing 11 CTGs has a negative effect on myogenesis and in the presence of expanded CUG repeats the effect is enhanced (Sabourin, *et al*. 1997; Storbeck, *et al*. 2004). Additionally, it was demonstrated that the overexpression of the 3'UTR containing 5 CTGs repeats in mice was sufficient to reproduce myotonia, cardiac conduction defects and RNA splicing defects observed in DM1 (Mahadevan, *et al*. 2006). In the DM1+CMT++ family the transcripts will be retained in the nucleus for the large number of CTG repeats, so it could be that the presence of just 20 CCGs trapped in the nucleus is sufficient to cause a phenotype.

The localisation and the composition of the nuclear foci in DM1 and FXTAS are different. In DM1 the foci are discrete, small, located at the periphery

of nuclear splicing speckles and containing mainly MBNL proteins and the mutant transcripts (Taneja, *et al*. 1995; Davis, *et al*. 1997; Miller, *et al*. 2000; Fardaei, *et al*. 2001; Holt, *et al*. 2007). In FXTAS the foci are diffuse and contain the mutant transcripts with a number of proteins, at least more than 20 different proteins were identified (Greco, *et al*. 2002; Iwahashi, *et al*. 2006). The presence and characteristics of RNA nuclear foci will be investigated using a lymphoblast cell line from IV-21. Additionally, by imunohistochemistry the presence of neurofilaments and lamin A/C could be tested.

In DM1 the RNA gain-of-function model proposed that the decrease of MBNL and the increase of CELF proteins cause a misregulation of the alternative splicing (Ho, *et al*. 2004). A number of pre-mRNAs were shown to present defects in the alternative splicing including: *TNNT2* (Philips, *et al*. 1998), *IR* (Savkur, *et al*. 2001), *CLCN1* (Charlet, *et al*. 2002; Mankodi, *et al*. 2002), *brain microtubule-associated tau* (Sergeant, *et al*. 2001) and *MTMR1* (Buj-Bello, *et al*. 2002). It will be interesting to investigate the localization of MBNL and CUGBP and investigate if the splicing defects observed with expanded pure CTG repeats are also reproduced in imperfect CTG repeat alleles.

The symptoms of acute encephalopathy and migraine resemble symptoms observed in disorders such as: familial hemiplegic migraine type 1 (FHM1) and 2 (FHM2), cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) or acute confusional migraine (ACM). FHM1 can be caused by mutations in the *calcium channel* gene (*CACNA1A*) (Ophoff, *et al*. 1996), FHM2 by mutations in a gene encoding the *alpha-2 subunit of the sodium/potassium pump* (*ATP1A2*) (De Fusco, *et al*. 2003) and CADASIL by mutations in *NOTCH3* (Joutel, *et al*. 1996). It could be postulated that these symptoms maybe caused by the loss of proteins, which are trapped in the nucleus or by defects in alternative splicing. However neither DM1 nor CMT patients present with migraine and/or acute encephalopathy. Three MBNL proteins have been described to co-localise in the DM1 loci (Fardaei, *et al*. 2002), it could be that the combination of CUG and CGG might have different affinities with three

MBNL proteins, and the levels are affected. Additionally, several CELFs have also been identified (Ladd, *et al*. 2001) and novel effects can also be caused by the combination of CUGs and CCGs trapped in the nucleus. This may have novel effects on the splicing of different premRNAs. Alternatively, the combination of both type of repeat might recruit novel proteins. Further investigations of the characterisations of the nuclear foci could reveal the mechanism. It will be more difficult to explain how the patients developed hearing loss, since mutations associated with hearing loss have been identified in more than 20 different genes (Finsterer and Fellinger 2005).

The second possible mechanism generating pathogenesis in the DM1+CMT++ family could be a novel effect on the downstream genes. It was postulated that the expansion of pure CTG repeats in the *DM1* locus alters the adjacent chromatin structure establishing a region of condensed chromatin in consequence the transcription the *DMPK* gene and/or neighbouring genes might be repressed (Wang, *et al*. 1994; Otten and Tapscott 1995). The CTG repeats in the *DM1* locus are inserted in a CpG island, which controls the activity of the downstream gene *SIX5* (Boucher, *et al*. 1995). In DM1 congenital cases the CpG island is hypermethylated and the expression of *SIX5* is repressed. This repression might be a molecular explanation for the increase severity detected in congenital cases (Steinbach, *et al*. 1998). The interruptions in the DM1+CMT++ allele provide new CpG sites, of which at least some seemed to be methylated (section 4.2.6). It could be postulated that the pattern of methylation could be extended to the CpG island as in congenital DM1 cases; or alternatively, by analogy to the mechanism observed in FXTAS, the CpG island may remain unmethylated and the level of *SIX5* transcription increased (Tassone, *et al*. 2000b). It could be argued that between these two possibilities the most likely explanation will be a increase in the levels of *SIX5* since the DM1 symptoms in the DM1+CMT++ family are not as severe as in congenital cases. To test which mechanism is involved, the pattern of methylation within the interruptions and in the flanking DNA will be further investigated with other methylation sensitive enzymes and by bisulphite sequencing.

In the other unusual cases, the pattern of methylation at the 3'-end seemed to depend on the number and type of interruptions. Only one single site was investigated at the 3'-end, but differences were visualised, the expanded alleles of DM1-UC2, DM1-UC3, DM1-UC4, DM1-UC5 and DM1-UC6 were methylated, whereas the expanded alleles of DM1-UC8 and DM1-UC9 were unmethylated. Further investigations will be necessary to understand the possible effects of methylation on the activities of *DMPK* and *SIX5*.

A correlation with the genotype and the clinical symptoms in the unusual cases would be extremely interesting to determine the effect of interruptions on the pathogenesis. It could be inferred that in the majority of the cases the presence of interruptions decreases the severity of the symptoms if we take into account that the progenitor allele is quite high for DM1 patients with the adult onset form. Additionally, it could be postulated that these unusual cases probably present a modified phenotype since the structure of the imperfect CTG repeat alleles resembles the DM1+CMT++ allele. Further analysis will be performed once our collaborators provide us with more details of the phenotype of the unusual cases.

An increase in the severity of symptoms and a reduced age of onset was observed in the CMT phenotype through the generations in the DM1+CMT++ family. If CMT is caused by the presence of interruptions, the anticipation observed between the third and fourth generation could be explained by the small changes in the number of CTG repeats at the 5'-end. It could be possible that slight changes in the number of CTG repeats at the 5'-end are sufficient to enhance the effects of the CCG. Alternatively, it could be explained by a toxicity effect caused by the high concentration of mutant *DMPK* transcripts in the nucleus. It will be necessary to demonstrate first that the interruptions are definitely the cause of CMT to subsequently test the pathogenesis of the CCG repeats.

6.7 Concluding remarks

In summary, this project has revealed that imperfect CTG repeat alleles are present in DM1 patients. Therefore, this fact should be taken into consideration when performing the molecular diagnosis of DM1, because false negatives obtained by PCRs and/or RP-PCR may be explained by imperfect CTG repeat alleles. A number of DM1 patients present with a disease that is not as severe as expected for the number of repeats, and the presence of imperfect CTG repeat alleles could be a possible explanation for this discrepancy.

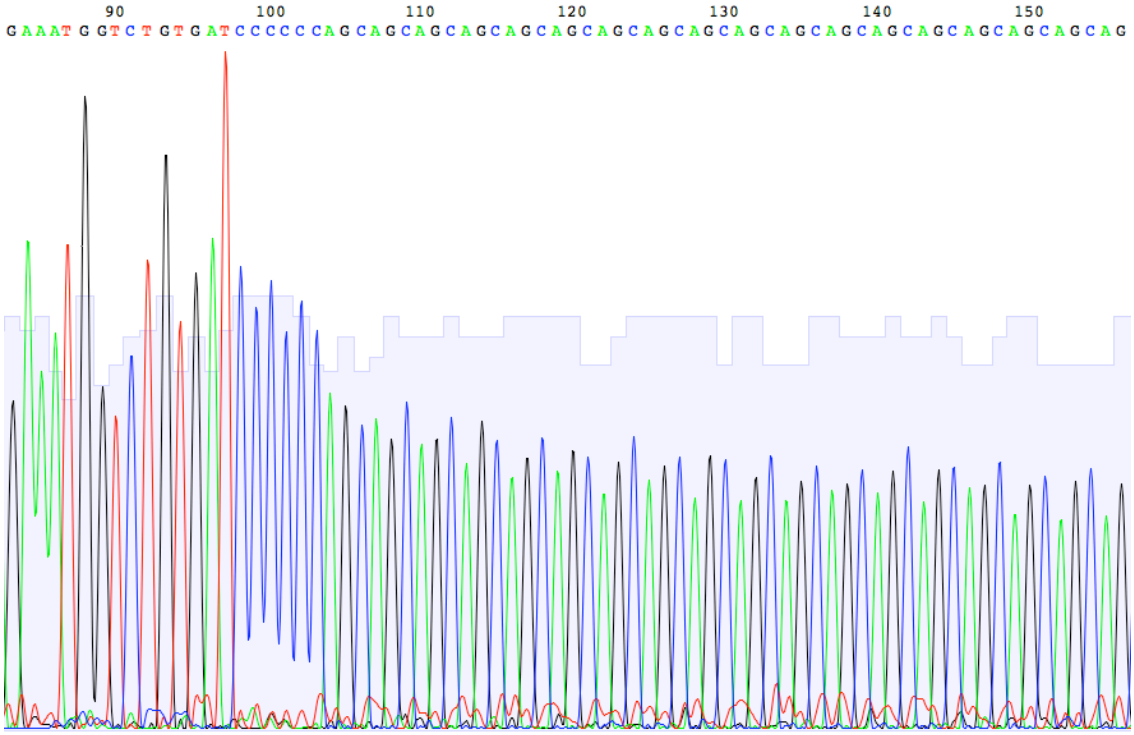
A better characterisation of the mutation in DM1 patients could lead in the future to the provision of more accurate prognosis. In addition to the considerable value of more accurate prognosis to individual families, these data will also facilitate clinical trials once therapies become available. More accurate genotype to phenotype data will reduce un-attributable variation within groups (treated versus placebo) and offer an enhanced ability to detect the effect of the drug. In particular, it will be necessary to investigate the presence of imperfect CTG repeat alleles in the DM1 patients in order to achieve a better characterisation.

Finally, the findings of this project can have more general implications. Imperfect alleles could also be present in other repeats diseases caused by dynamic mutations from which interruptions have not yet been considered. In those diseases such as fragile X syndrome, which are known to present interruptions, the interruptions are presented in the normal alleles. However, in a number of other diseases such as Huntington disease, only a small number of expanded alleles have been sequenced to date and no interruptions have been found, so it has been assumed that the expanded repeats are pure as is the case with DM1. The results from our project suggest that further investigation should be undertaken to test this assumption. Should interruptions be present in diseases such as HD, the presence of interruptions might have a modifier effect on the pathogenesis not only the severity of the symptoms but also in the age of onset, as it was demonstrated in SCA1, SCA2, SCA10, SCA17 and in our project (Chong, *et*

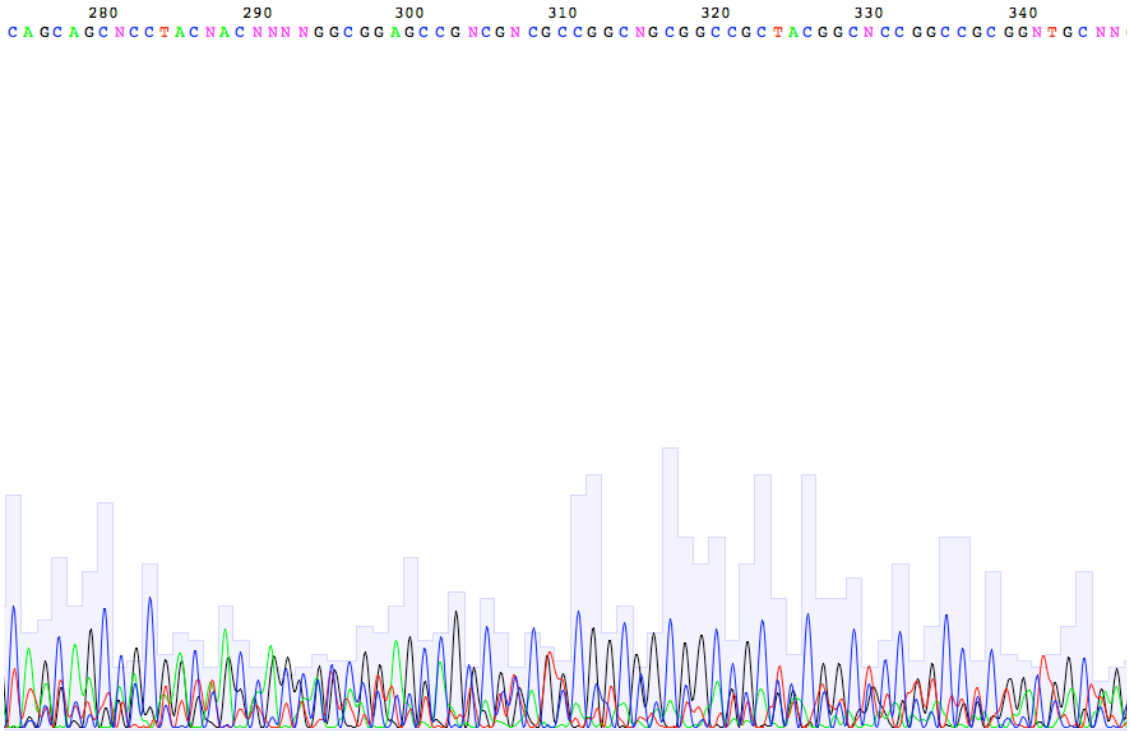
al. 1995; Zuhlke*, et al*. 2002; Matsuura*, et al*. 2006; Charles*, et al*. 2007; Gao*, et al*. 2008). Additionally, it could be possible that families with Charcot-Marie-Tooth disease with an unknown aetiology may be carriers of CCG or GGC expanded repeats as well.

Appendices

Appendix A: sequencing chromatogram showing a region of good sequence with little baseline noise and very strong peaks.



Appendix B: sequencing chromatogram showing a region with a noisy sequence.



List of References

Adlkofer, K.; Martini, R.; Aguzzi, A.; Zielasek, J.; Toyka, K.V.; Suter, U. (1995). Hypermyelination and demyelinating peripheral neuropathy in Pmp22-deficient mice. *Nat Genet* **11**: 274-80.

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

Annesi, G.; Nicoletti, G.; Tarantino, P.; Cutuli, N.; Annesi, F.; Marco, E.V.; Zappia, M.; Morgante, L.; Arabia, G.; Pugliese, P.; Condino, F.; Carrideo, S.; Civitelli, D.; Caracciolo, M.; Romeo, N.; Spadafora, P.; Candiano, I.C.; Quattrone, A. (2004). FRAXE intermediate alleles are associated with Parkinson's disease. *Neurosci Lett* **368**: 21-4.

Antonellis, A.; Ellsworth, R.E.; Sambuughin, N.; Puls, I.; Abel, A.; Lee-Lin, S.Q.; Jordanova, A.; Kremensky, I.; Christodoulou, K.; Middleton, L.T.; Sivakumar, K.; Ionasescu, V.; Funalot, B.; Vance, J.M.; Goldfarb, L.G.; Fischbeck, K.H.; Green, E.D. (2003). Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. *Am J Hum Genet* **72**: 1293-9.

Anvret, M.; Ahlberg, G.; Grandell, U.; Hedberg, B.; Johnson, K.; Edstrom, L. (1993). Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum Mol Genet* **2**: 1397-400.

Arnold, C.; Hodgson, I.J. (1991). Vectorette PCR: a novel approach to genomic walking. *PCR Methods Appl* **1**: 39-42.

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

Ashizawa, T.; Dubel, J.R.; Dunne, P.W.; Dunne, C.J.; Fu, Y.H.; Pizzuti, A.; Caskey, C.T.; Boerwinkle, E.; Perryman, M.B.; Epstein, H.F.; et al. (1992). Anticipation in myotonic dystrophy. II. Complex relationships between clinical findings and structure of the GCT repeat. *Neurology* **42**: 1877-83.

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

Ashizawa, T.; Dunne, P.W.; Ward, P.A.; Seltzer, W.K.; Richards, C.S. (1994b). Effects of the sex of myotonic dystrophy patients on the unstable triplet repeat in their affected offspring. *Neurology* **44**: 120-2.

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

Ashizawa, T.; Harper, P.S. (2006). Myotonic dystrophies: an overview. In: Wells D and Ashizawa T (ed) Genetic Instabilities and Neurological Diseases. Oxford: Academic., pp 21-29

Ashley, C.T., Jr.; Warren, S.T. (1995). Trinucleotide repeat expansion and human disease. *Annu Rev Genet* **29**: 703-28.

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

Barcelo, J.M.; Mahadevan, M.S.; Tsilfidis, C.; MacKenzie, A.E.; Korneluk, R.G. (1993). Intergenerational stability of the myotonic dystrophy protomutation. *Hum Mol Genet* **2**: 705-9.

Barisic, N.; Claeys, K.G.; Sirotkovic-Skerlev, M.; Lofgren, A.; Nelis, E.; De Jonghe, P.; Timmerman, V. (2008). Charcot-Marie-Tooth Disease: A Clinico-genetic Confrontation. *Ann Hum Genet* **72**: 416-41.

Bell, M.V.; Hirst, M.C.; Nakahori, Y.; MacKinnon, R.N.; Roche, A.; Flint, T.J.; Jacobs, P.A.; Tommerup, N.; Tranebjaerg, L.; Froster-Iskenius, U.; et al. (1991). Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. *Cell* **64**: 861-6.

Bellacosa, A. (2001). Functional interactions and signaling properties of mammalian DNA mismatch repair proteins. *Cell Death Differ* **8**: 1076-92.

Benders, A.A.; Groenen, P.J.; Oerlemans, F.T.; Veerkamp, J.H.; Wieringa, B. (1997). Myotonic dystrophy protein kinase is involved in the modulation of the Ca2+ homeostasis in skeletal muscle cells. *J Clin Invest* **100**: 1440-7.

Berry-Kravis, E.; Lewin, F.; Wuu, J.; Leehey, M.; Hagerman, R.; Hagerman, P.; Goetz, C.G. (2003). Tremor and ataxia in fragile X premutation carriers: blinded videotape study. *Ann Neurol* **53**: 616-23.

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

Bird, T.D.; Kraft, G.H.; Lipe, H.P.; Kenney, K.L.; Sumi, S.M. (1997). Clinical and pathological phenotype of the original family with Charcot-Marie-Tooth type 1B: a 20-year study. *Ann Neurol* **41**: 463-9.

Bird, T.D.; Ott, J.; Giblett, E.R. (1982). Evidence for linkage of Charcot-Marie-Tooth neuropathy to the Duffy locus on chromosome 1. *Am J Hum Genet* **34**: 388-94.

Borenstein, S.; Noel, P.; Jacquy, J.; Flamentdurand, J. (1977). Myotonic dystrophy with nerve hypertrophy. Report of a case with electrophysiological and ultrastructural study of the sural nerve. *J Neurol Sci* **34**: 87-99.

Boucher, C.A.; King, S.K.; Carey, N.; Krahe, R.; Winchester, C.L.; Rahman, S.; Creavin, T.; Meghji, P.; Bailey, M.E.; Chartier, F.L.; 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**: 1919-25.

Breschel, T.S.; McInnis, M.G.; Margolis, R.L.; Sirugo, G.; Corneliussen, B.; Simpson, S.G.; McMahon, F.J.; MacKinnon, D.F.; Xu, J.F.; Pleasant, N.; Huo, Y.; Ashworth, R.G.; Grundstrom, C.; Grundstrom, T.; Kidd, K.K.; DePaulo, J.R.; Ross, C.A. (1997). A novel, heritable, expanding CTG repeat in an intron of the SEF2-1 gene on chromosome 18q21.1. *Hum Mol Genet* **6**: 1855-63.

Brock, G.J.; Anderson, N.H.; Monckton, D.G. (1999). Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum Mol Genet* **8**: 1061-7.

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

Brunner, H.G.; Bruggenwirth, H.T.; Nillesen, W.; Jansen, G.; Hamel, B.C.; Hoppe, R.L.; de Die, C.E.; Howeler, C.J.; van Oost, B.A.; Wieringa, B.; et al. (1993a). Influence of sex of the transmitting parent as well as of parental allele size on the CTG expansion in myotonic dystrophy (DM). *Am J Hum Genet* **53**: 1016-23.

Brunner, H.G.; Jansen, G.; Nillesen, W.; Nelen, M.R.; de Die, C.E.; Howeler, C.J.; van Oost, B.A.; Wieringa, B.; Ropers, H.H.; Smeets, H.J. (1993b). Brief report: reverse mutation in myotonic dystrophy. *N Engl J Med* **328**: 476-80.

Brunner, H.G.; Spaans, F.; Smeets, H.J.; Coerwinkel-Driessen, M.; Hulsebos, T.; Wieringa, B.; Ropers, H.H. (1991). Genetic linkage with chromosome 19 but not

chromosome 17 in a family with myotonic dystrophy associated with hereditary motor and sensory neuropathy. *Neurology* **41**: 80-4.

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

Bush, E.W.; Taft, C.S.; Meixell, G.E.; Perryman, M.B. (1996). Overexpression of myotonic dystrophy kinase in BC3H1 cells induces the skeletal muscle phenotype. *J Biol Chem* **271**: 548-52.

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

Caccia, M.R.; Negri, S.; Parvis, V.P. (1972). Myotonic dystrophy with neural involvement. *J Neurol Sci* **16**: 253-69.

Campuzano, V.; Montermini, L.; Molto, M.D.; Pianese, L.; Cossee, M.; Cavalcanti, F.; Monros, E.; Rodius, F.; Duclos, F.; Monticelli, A.; Zara, F.; Canizares, J.; Koutnikova, H.; Bidichandani, S.I.; Gellera, C.; Brice, A.; Trouillas, P.; De Michele, G.; Filla, A.; De Frutos, R.; Palau, F.; Patel, P.I.; Di Donato, S.; Mandel, J.L.; Cocozza, S.; Koenig, M.; Pandolfo, M. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**: 1423-7.

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

Chakrabarti, R.; Schutt, C.E. (2002). Novel sulfoxides facilitate GC-rich template amplification. *Biotechniques* **32**: 866, 68, 70-2, 74.

Charles, P.; Camuzat, A.; Benammar, N.; Sellal, F.; Destee, A.; Bonnet, A.M.; Lesage, S.; Le Ber, I.; Stevanin, G.; Durr, A.; Brice, A. (2007). Are interrupted SCA2 CAG repeat expansions responsible for parkinsonism? *Neurology* **69**: 1970-5.

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

Chen, X.; Mariappan, S.V.; Moyzis, R.K.; Bradbury, E.M.; Gupta, G. (1998). Hairpin induced slippage and hyper-methylation of the fragile X DNA triplets. *J Biomol Struct Dyn* **15**: 745-56.

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

Chong, S.S.; McCall, A.E.; Cota, J.; Subramony, S.H.; Orr, H.T.; Hughes, M.R.; Zoghbi, H.Y. (1995). Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* **10**: 344-50.

Choudhry, S.; Mukerji, M.; Srivastava, A.K.; Jain, S.; Brahmachari, S.K. (2001). CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. *Hum Mol Genet* **10**: 2437-46.

Chung, M.Y.; Ranum, L.P.; Duvick, L.A.; Servadio, A.; Zoghbi, H.Y.; Orr, H.T. (1993). Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat Genet* **5**: 254-8.

Cleary, J.D.; Pearson, C.E. (2003). The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet Genome Res* **100**: 25-55.

Clikeman, J.A.; Wheeler, S.L.; Nickoloff, J.A. (2001). Efficient incorporation of large (>2 kb) heterologies into heteroduplex DNA: Pms1/Msh2-dependent and -independent large loop mismatch repair in Saccharomyces cerevisiae. *Genetics* **157**: 1481-91.

Cockburn, D.J.; Mavrogiannis, L.A.; Flintoff, K.J.; Taylor, G.R. (2003). Triplet and quadruplet repeat-primed PCR in DM1 and DM2 diagnosis: detection of unusual mutations. 4th International Myotonic Dystrophy Consortium Meeting, Glasgow University, Glasgow, Scotland

Cossee, M.; Schmitt, M.; Campuzano, V.; Reutenauer, L.; Moutou, C.; Mandel, J.L.; Koenig, M. (1997). Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. *Proc Natl Acad Sci U S A* **94**: 7452-7.

Cros, D.; Harnden, P.; Pouget, J.; Pellissier, J.F.; Gastaut, J.L.; Serratrice, G. (1988). Peripheral neuropathy in myotonic dystrophy: a nerve biopsy study. *Ann Neurol* **23**: 470-6.

Cummings, C.J.; Zoghbi, H.Y. (2000a). Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum Mol Genet* **9**: 909-16.

Cummings, C.J.; Zoghbi, H.Y. (2000b). Trinucleotide repeats: mechanisms and pathophysiology. *Annu Rev Genomics Hum Genet* **1**: 281-328.

David, G.; Abbas, N.; Stevanin, G.; Durr, A.; Yvert, G.; Cancel, G.; Weber, C.; Imbert, G.; Saudou, F.; Antoniou, E.; Drabkin, H.; Gemmill, R.; Giunti, P.; Benomar, A.; Wood, N.; Ruberg, M.; Agid, Y.; Mandel, J.L.; Brice, A. (1997). Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* **17**: 65-70.

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

Davis, C.J.; Bradley, W.G.; Madrid, R. (1978). The peroneal muscular atrophy syndrome: clinical, genetic, electrophysiological and nerve biopsy studies. I. Clinical, genetic and electrophysiological findings and classification. *J Genet Hum* **26**: 311-49.

Day, J.W.; Ranum, L.P. (2005). RNA pathogenesis of the myotonic dystrophies. *Neuromuscul Disord* **15**: 5-16.

Day, J.W.; Ricker, K.; Jacobsen, J.F.; Rasmussen, L.J.; Dick, K.A.; Kress, W.; Schneider, C.; Koch, M.C.; Beilman, G.J.; Harrison, A.R.; Dalton, J.C.; Ranum, L.P. (2003). Myotonic dystrophy type 2: molecular, diagnostic and clinical spectrum. *Neurology* **60**: 657-64.

De Fusco, M.; Marconi, R.; Silvestri, L.; Atorino, L.; Rampoldi, L.; Morgante, L.; Ballabio, A.; Aridon, P.; Casari, G. (2003). Haploinsufficiency of ATP1A2 encoding the Na+/K+ pump alpha2 subunit associated with familial hemiplegic migraine type 2. *Nat Genet* **33**: 192-6.

de Graaff, E.; Rouillard, P.; Willems, P.J.; Smits, A.P.; Rousseau, F.; Oostra, B.A. (1995). Hotspot for deletions in the CGG repeat region of FMR1 in fragile X patients. *Hum Mol Genet* **4**: 45-9.

Delague, V.; Bareil, C.; Tuffery, S.; Bouvagnet, P.; Chouery, E.; Koussa, S.; Maisonobe, T.; Loiselet, J.; Megarbane, A.; Claustres, M. (2000). Mapping of a new locus for autosomal recessive demyelinating Charcot-Marie-Tooth disease to 19q13.1-13.3 in a large consanguineous Lebanese family: exclusion of MAG as a candidate gene. *Am J Hum Genet* **67**: 236-43.

Dunne, P.W.; Ma, L.; Casey, D.L.; Epstein, H.F. (1996). Myotonic protein kinase expression in human and bovine lenses. *Biochem Biophys Res Commun* **225**: 281-8.

Dupre, N.; Bouchard, J.P.; Cossette, L.; Brunet, D.; Vanasse, M.; Lemieux, B.; Mathon, G.; Puymirat, J. (1999). Clinical and electrophysiological study in French-Canadian population with Charcot-Marie-tooth disease type 1A associated with 17p11.2 duplication. *Can J Neurol Sci* **26**: 196-200.

Eichler, E.E.; Holden, J.J.; Popovich, B.W.; Reiss, A.L.; Snow, K.; Thibodeau, S.N.; Richards, C.S.; Ward, P.A.; Nelson, D.L. (1994). Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat Genet* **8**: 88-94.

Eriksson, M.; Ansved, T.; Edstrom, L.; Anvret, M.; Carey, N. (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**: 1053-60.

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

Fardaei, M.; Rogers, M.T.; Thorpe, H.M.; Larkin, K.; Hamshere, M.G.; Harper, P.S.; Brook, J.D. (2002). Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum Mol Genet* **11**: 805-14.

Feng, Y.; Lakkis, L.; Devys, D.; Warren, S.T. (1995). Quantitative comparison of FMR1 gene expression in normal and premutation alleles. *Am J Hum Genet* **56**: 106-13.

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

Finsterer, J.; Fellinger, J. (2005). Nuclear and mitochondrial genes mutated in nonsyndromic impaired hearing. *Int J Pediatr Otorhinolaryngol* **69**: 621-47.

Fleischer, B. (1918). Über myotonische Dystrophiemit Katarakt. *Albrecht von Graefes Arch Klin Ophthalmol* **96**:91-133. In: Harper, P.S. (2001). Myotonic Dystrophy. 3er edn. WB Saunders Co., London, pp 311

Flynn, G.A.; Hirst, M.C.; Knight, S.J.; Macpherson, J.N.; Barber, J.C.; Flannery, A.V.; Davies, K.E.; Buckle, V.J. (1993). Identification of the FRAXE fragile site in two families ascertained for X linked mental retardation. *J Med Genet* **30**: 97-100.

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

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

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

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

Fujigasaki, H.; Martin, J.J.; De Deyn, P.P.; Camuzat, A.; Deffond, D.; Stevanin, G.; Dermaut, B.; Van Broeckhoven, C.; Durr, A.; Brice, A. (2001). CAG repeat expansion in the TATA box-binding protein gene causes autosomal dominant cerebellar ataxia. *Brain* **124**: 1939-47.

Galvao, R.; Mendes-Soares, L.; Camara, J.; Jaco, I.; Carmo-Fonseca, M. (2001a). Triplet repeats, RNA secondary structure and toxic gain-of-function models for pathogenesis. *Brain Res Bull* **56**: 191-201.

Galvao, R.; Mendes-Soares, L.; Camara, J.; Jaco, I.; Carmo-Fonseca, M.; si (2001b). Triplet repeats, RNA secondary structure and toxic gain-of-function models for pathogenesis. *Brain Res Bull* **56**: 191-201.

Gao, R.; Matsuura, T.; Coolbaugh, M.; Zuhlke, C.; Nakamura, K.; Rasmussen, A.; Siciliano, M.J.; Ashizawa, T.; Lin, X. (2008). Instability of expanded CAG/CAA repeats in spinocerebellar ataxia type 17. *Eur J Hum Genet* **16**: 215-22.

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

Gecz, J. (2000) The FMR2 gene, FRAXE and non-specific X-linked mental retardation: clinical and molecular aspects. *Ann Hum Genet* **64**: 95-106.

Gecz, J.; Gedeon, A.K.; Sutherland, G.R.; Mulley, J.C. (1996). Identification of the gene FMR2, associated with FRAXE mental retardation. *Nat Genet* **13**: 105-8.

Gomes-Pereira, M.; Bidichandani, S.I.; Monckton, D.G. (2004a). Analysis of unstable triplet repeats using small-pool polymerase chain reaction. *Methods Mol Biol* **277**: 61-76.

Gomes-Pereira, M.; Foiry, L.; Gourdon, G. (2006). Transgenic mouse models of unstable trinucleotide repeats: toward an understanding of disease-associated repeat size mutation. In: Wells D and Ashizawa T (ed) Genetic Instabilities and Neurological Diseases. Oxford: Academic., pp 563-83

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

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

Gomes-Pereira, M.; Fortune, M.T.; Monckton, D.G. (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**: 845-54.

Good, P.J.; Chen, Q.; Warner, S.J.; Herring, D.C. (2000). A family of human RNA-binding proteins related to the Drosophila Bruno translational regulator. *J Biol Chem* **275**: 28583-92.

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

Greco, C.M.; Hagerman, R.J.; Tassone, F.; Chudley, A.E.; Del Bigio, M.R.; Jacquemont, S.; Leehey, M.; Hagerman, P.J. (2002). Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain* **125**: 1760-71.

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

Gu, Y.; Nelson, D.L. (2003). FMR2 function: insight from a mouse knockout model. *Cytogenet Genome Res* **100**: 129-39.

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

Hagerman, P.J.; Hagerman, R.J. (2004). The fragile-X premutation: a maturing perspective. *Am J Hum Genet* **74**: 805-16.

Hagerman, R.J.; Leehey, M.; Heinrichs, W.; Tassone, F.; Wilson, R.; Hills, J.; Grigsby, J.; Gage, B.; Hagerman, P.J. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* **57**: 127-30.

Hamiel, O.P.; Raas-Rothschild, A.; Upadhyaya, M.; Frydman, M.; Sarova-Pinhas, I.; Brand, N.; Passwell, J.H. (1993). Hereditary motor-sensory neuropathy (Charcot-Marie-Tooth disease) with nerve deafness: a new variant. *J Pediatr* **123**: 431-4.

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

Harding, A.E.; Thomas, P.K. (1980). The clinical features of hereditary motor and sensory neuropathy types I and II. *Brain* **103**: 259-80.

Harley, H.G.; Brook, J.D.; Floyd, J.; Rundle, S.A.; Crow, S.; Walsh, K.V.; Thibault, M.C.; Harper, P.S.; Shaw, D.J. (1991). Detection of linkage disequilibrium between the myotonic dystrophy locus and a new polymorphic DNA marker. *Am J Hum Genet* **49**: 68-75.

Harley, H.G.; Brook, J.D.; Rundle, S.A.; Crow, S.; Reardon, W.; Buckler, A.J.; Harper, P.S.; Housman, D.E.; Shaw, D.J. (1992). Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* **355**: 545-6.

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

Harper, P.S. (2001) Myotonic dystrophy. WB Saunders Co., London

Harper, P.S.; Harley, H.G.; Reardon, W.; Shaw, D.J. (1992). Anticipation in myotonic dystrophy: new light on an old problem. *Am J Hum Genet* **51**: 10-6.

(HDCRG), The Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**: 971-83.

Hegde, M.R.; Chong, B.; Fawkner, M.; Lambiris, N.; Peters, H.; Kenneson, A.; Warren, S.T.; Love, D.R.; McGaughran, J. (2001). Microdeletion in the FMR-1 gene: an apparent null allele using routine clinical PCR amplification. *J Med Genet* **38**: 624-9.

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

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

Hofmann-Radvanyi, H.; Lavedan, C.; Rabes, J.P.; Savoy, D.; Duros, C.; Johnson, K.; Junien, C. (1993). Myotonic dystrophy: absence of CTG enlarged transcript in congenital forms, and low expression of the normal allele. *Hum Mol Genet* **2**: 1263-6.

Holmes, S.E.; Hearn, E.O.; Ross, C.A.; Margolis, R.L. (2001a). SCA12: an unusual mutation leads to an unusual spinocerebellar ataxia. *Brain Res Bull* **56**: 397-403.

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

Holmes, S.E.; O'Hearn, E.E.; McInnis, M.G.; Gorelick-Feldman, D.A.; Kleiderlein, J.J.; Callahan, C.; Kwak, N.G.; Ingersoll-Ashworth, R.G.; Sherr, M.; Sumner, A.J.;

Sharp, A.H.; Ananth, U.; Seltzer, W.K.; Boss, M.A.; Vieria-Saecker, A.M.; Epplen, J.T.; Riess, O.; Ross, C.A.; Margolis, R.L. (1999). Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nat Genet* **23**: 391-2.

Holt, I.; Mittal, S.; Furling, D.; Butler-Browne, G.S.; Brook, J.D.; Morris, G.E. (2007). Defective mRNA in myotonic dystrophy accumulates at the periphery of nuclear splicing speckles. *Genes Cells* **12**: 1035-48.

Howeler, C.J.; Busch, H.F.; Geraedts, J.P.; Niermeijer, M.F.; Staal, A. (1989). Anticipation in myotonic dystrophy: fact or fiction? *Brain* **112**: 779-97.

Hunter, A.G.; Jacob, P.; O'Hoy, K.; MacDonald, I.; Mettler, G.; Tsilfidis, C.; Korneluk, R.G. (1993). Decrease in the size of the myotonic dystrophy CTG repeat during transmission from parent to child: implications for genetic counselling and genetic anticipation. *Am J Med Genet* **45**: 401-7.

(IDMC), The International Myotonic Dystrophy Consortium. (2000). New nomenclature and DNA testing guidelines for myotonic dystrophy type 1 (DM1). . *Neurology* **54**: 1218-21.

Ikeuchi, T.; Sanpei, K.; Takano, H.; Sasaki, H.; Tashiro, K.; Cancel, G.; Brice, A.; Bird, T.D.; Schellenberg, G.D.; Pericak-Vance, M.A.; Welsh-Bohmer, K.A.; Clark, L.N.; Wilhelmsen, K.; Tsuji, S. (1998). A novel long and unstable CAG/CTG trinucleotide repeat on chromosome 17q. *Genomics* **49**: 321-6.

Imbert, G.; Kretz, C.; Johnson, K.; Mandel, J.L. (1993). Origin of the expansion mutation in myotonic dystrophy. *Nat Genet* **4**: 72-6.

Imbert, G.; Saudou, F.; Yvert, G.; Devys, D.; Trottier, Y.; Garnier, J.M.; Weber, C.; Mandel, J.L.; Cancel, G.; Abbas, N.; Durr, A.; Didierjean, O.; Stevanin, G.; Agid, Y.; Brice, A. (1996). Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat Genet* **14**: 285-91.

Iwahashi, C.K.; Yasui, D.H.; An, H.J.; Greco, C.M.; Tassone, F.; Nannen, K.; Babineau, B.; Lebrilla, C.B.; Hagerman, R.J.; Hagerman, P.J. (2006). Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**: 256-71.

Jacquemont, S.; Hagerman, R.J.; Leehey, M.; Grigsby, J.; Zhang, L.; Brunberg, J.A.; Greco, C.; Des Portes, V.; Jardini, T.; Levine, R.; Berry-Kravis, E.; Brown, W.T.; Schaeffer, S.; Kissel, J.; Tassone, F.; Hagerman, P.J. (2003). Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am J Hum Genet* **72**: 869-78.

Jansen, G.; Bachner, D.; Coerwinkel, M.; Wormskamp, N.; Hameister, H.; Wieringa, B. (1995). Structural organization and developmental expression pattern of the mouse WD-repeat gene DMR-N9 immediately upstream of the myotonic dystrophy locus. *Hum Mol Genet* **4**: 843-52.

Jansen, G.; Groenen, P.J.; Bachner, D.; Jap, P.H.; Coerwinkel, M.; Oerlemans, F.; van den Broek, W.; Gohlsch, B.; Pette, D.; Plomp, J.J.; Molenaar, P.C.; Nederhoff, M.G.; van Echteld, C.J.; Dekker, M.; Berns, A.; Hameister, H.; Wieringa, B. (1996). Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. *Nat Genet* **13**: 316-24.

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

Jansen, G.; Willems, P.; Coerwinkel, M.; Nillesen, W.; Smeets, H.; Vits, L.; Howeler, C.; Brunner, H.; Wieringa, B. (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**: 575-85.

Jeffreys, A.J.; MacLeod, A.; Tamaki, K.; Neil, D.L.; Monckton, D.G. (1991). Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354**: 204-9.

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

Jin, P.; Zarnescu, D.C.; Zhang, F.; Pearson, C.E.; Lucchesi, J.C.; Moses, K.; Warren, S.T. (2003). RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Neuron* **39**: 739-47.

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

Jones, C.; Mullenbach, R.; Grossfeld, P.; Auer, R.; Favier, R.; Chien, K.; James, M.; Tunnacliffe, A.; Cotter, F. (2000). Co-localisation of CCG repeats and chromosome deletion breakpoints in Jacobsen syndrome: evidence for a common mechanism of chromosome breakage. *Hum Mol Genet* **9**: 1201-8.

Jones, C.; Penny, L.; Mattina, T.; Yu, S.; Baker, E.; Voullaire, L.; Langdon, W.Y.; Sutherland, G.R.; Richards, R.I.; Tunnacliffe, A. (1995). Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene CBL2. *Nature* **376**: 145-9.

Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat, H.; Mouton, P.; Alamowitch, S.; Domenga, V.; Cecillion, M.; Marechal, E.; Maciazek, J.; Vayssiere, C.; Cruaud, C.; Cabanis, E.A.; Ruchoux, M.M.; Weissenbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserve, E. (1996). Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature* **383**: 707-10.

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

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

Kang, S.; Jaworski, A.; Ohshima, K.; Wells, R.D. (1995). Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. *Nat Genet* **10**: 213-8.

Kang, S.; Ohshima, K.; Jaworski, A.; Wells, R.D. (1996). CTG triplet repeats from the myotonic dystrophy gene are expanded in Escherichia coli distal to the replication origin as a single large event. *J Mol Biol* **258**: 543-7.

Kawaguchi, Y.; Okamoto, T.; Taniwaki, M.; Aizawa, M.; Inoue, M.; Katayama, S.; Kawakami, H.; Nakamura, S.; Nishimura, M.; Akiguchi, I.; et al. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* **8**: 221-8.

Kennedy, L.; Evans, E.; Chen, C.M.; Craven, L.; Detloff, P.J.; Ennis, M.; Shelbourne, P.F. (2003). Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum Mol Genet* **12**: 3359-67.

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

Kirby, R.J.; Hamilton, G.M.; Finnegan, D.J.; Johnson, K.J.; Jarman, A.P. (2001). Drosophila homolog of the myotonic dystrophy-associated gene, SIX5, is required for muscle and gonad development. *Curr Biol* **11**: 1044-9.

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

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

Knight, S.J.; Flannery, A.V.; Hirst, M.C.; Campbell, L.; Christodoulou, Z.; Phelps, S.R.; Pointon, J.; Middleton-Price, H.R.; Barnicoat, A.; Pembrey, M.E.; et al. (1993). Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74**: 127-34.

Koide, R.; Ikeuchi, T.; Onodera, O.; Tanaka, H.; Igarashi, S.; Endo, K.; Takahashi, H.; Kondo, R.; Ishikawa, A.; Hayashi, T.; et al. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nat Genet* **6**: 9-13.

Koide, R.; Kobayashi, S.; Shimohata, T.; Ikeuchi, T.; Maruyama, M.; Saito, M.; Yamada, M.; Takahashi, H.; Tsuji, S. (1999). A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? *Hum Mol Genet* **8**: 2047-53.

Koob, M.D.; Moseley, M.L.; Schut, L.J.; Benzow, K.A.; Bird, T.D.; Day, J.W.; Ranum, L.P. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat Genet* **21**: 379-84.

Kovach, M.J.; Campbell, K.C.; Herman, K.; Waggoner, B.; Gelber, D.; Hughes, L.F.; Kimonis, V.E. (2002). Anticipation in a unique family with Charcot-Marie-Tooth syndrome and deafness: delineation of the clinical features and review of the literature. *Am J Med Genet* **108**: 295-303.

Kovach, M.J.; Lin, J.P.; Boyadjiev, S.; Campbell, K.; Mazzeo, L.; Herman, K.; Rimer, L.A.; Frank, W.; Llewellyn, B.; Jabs, E.W.; Gelber, D.; Kimonis, V.E. (1999). A unique point mutation in the PMP22 gene is associated with Charcot-Marie-Tooth disease and deafness. *Am J Hum Genet* **64**: 1580-93.

Krahe, R.; Eckhart, M.; Ogunniyi, A.O.; Osuntokun, B.O.; Siciliano, M.J.; Ashizawa, T. (1995). De novo myotonic dystrophy mutation in a Nigerian kindred. *Am J Hum Genet* **56**: 1067-74.

Kremer, B.; Almqvist, E.; Theilmann, J.; Spence, N.; Telenius, H.; Goldberg, Y.P.; Hayden, M.R. (1995). Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. *Am J Hum Genet* **57**: 343-50.

Kunst, C.B.; Leeflang, E.P.; Iber, J.C.; Arnheim, N.; Warren, S.T. (1997). The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing. *J Med Genet* **34**: 627-31.

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

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

Lam, L.T.; Pham, Y.C.; Nguyen, T.M.; Morris, G.E. (2000). Characterization of a monoclonal antibody panel shows that the myotonic dystrophy protein kinase, DMPK, is expressed almost exclusively in muscle and heart. *Hum Mol Genet* **9**: 2167-73.

Lammerding, J.; Schulze, P.C.; Takahashi, T.; Kozlov, S.; Sullivan, T.; Kamm, R.D.; Stewart, C.L.; Lee, R.T. (2004). Lamin A/C deficiency causes defective nuclear mechanics and mechanotransduction. *J Clin Invest* **113**: 370-8.

Laurent, A.; Costa, J.M.; Assouline, B.; Voyer, M.; Vidaud, M. (1997). Myotonic dystrophy protein kinase gene expression in skeletal muscle from congenitally affected infants. *Ann Genet* **40**: 169-74.

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

Le Ber, I.; Martinez, M.; Campion, D.; Laquerriere, A.; Betard, C.; Bassez, G.; Girard, C.; Saugier-Veber, P.; Raux, G.; Sergeant, N.; Magnier, P.; Maisonobe, T.; Eymard, B.; Duyckaerts, C.; Delacourte, A.; Frebourg, T.; Hannequin, D. (2004). A non-DM1, non-DM2 multisystem myotonic disorder with frontotemporal dementia: phenotype and suggestive mapping of the DM3 locus to chromosome 15q21-24. *Brain* **127**: 1979-92.

Leal, A.; Morera, B.; Del Valle, G.; Heuss, D.; Kayser, C.; Berghoff, M.; Villegas, R.; Hernandez, E.; Mendez, M.; Hennies, H.C.; Neundorfer, B.; Barrantes, R.; Reis, A.; Rautenstrauss, B. (2001). A second locus for an axonal form of autosomal recessive Charcot-Marie-Tooth disease maps to chromosome 19q13.3. *Am J Hum Genet* **68**: 269-74.

Leeflang, E.P.; Arnheim, N. (1995). A novel repeat structure at the myotonic dystrophy locus in a 37 repeat allele with unexpectedly high stability. *Hum Mol Genet* **4**: 135-6.

Leehey, M.A.; Munhoz, R.P.; Lang, A.E.; Brunberg, J.A.; Grigsby, J.; Greco, C.; Jacquemont, S.; Tassone, F.; Lozano, A.M.; Hagerman, P.J.; Hagerman, R.J. (2003). The fragile X premutation presenting as essential tremor. *Arch Neurol* **60**: 117-21.

Li, L.B.; Yu, Z.; Teng, X.; Bonini, N.M. (2008). RNA toxicity is a component of ataxin-3 degeneration in Drosophila. *Nature* **453**: 1107-11.

Lia, A.S.; Seznec, H.; Hofmann-Radvanyi, H.; Radvanyi, F.; Duros, C.; Saquet, C.; Blanche, M.; Junien, C.; Gourdon, G. (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**: 1285-91.

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

Littman, S.J.; Fang, W.H.; Modrich, P. (1999). Repair of large insertion/deletion heterologies in human nuclear extracts is directed by a 5' single-strand break and is independent of the mismatch repair system. *J Biol Chem* **274**: 7474-81.

Lu, X.; Timchenko, N.A.; Timchenko, L.T. (1999). Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy. *Hum Mol Genet* **8**: 53-60.

Mahadevan, M.; Tsilfidis, C.; Sabourin, L.; Shutler, G.; Amemiya, C.; Jansen, G.; Neville, C.; Narang, M.; Barcelo, J.; O'Hoy, K.; et al. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**: 1253-5.

Mahadevan, M.S.; Amemiya, C.; Jansen, G.; Sabourin, L.; Baird, S.; Neville, C.E.; Wormskamp, N.; Segers, B.; Batzer, M.; Lamerdin, J.; et al. (1993a). Structure and genomic sequence of the myotonic dystrophy (DM kinase) gene. *Hum Mol Genet* **2**: 299-304.

Mahadevan, M.S.; Foitzik, M.A.; Surh, L.C.; Korneluk, R.G. (1993b). Characterization and polymerase chain reaction (PCR) detection of an Alu deletion polymorphism in total linkage disequilibrium with myotonic dystrophy. *Genomics* **15**: 446-8.

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

Mankodi, A.; Logigian, E.; Callahan, L.; McClain, C.; White, R.; Henderson, D.; Krym, M.; Thornton, C.A. (2000). Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* **289**: 1769-73.

Mankodi, A.; Takahashi, M.P.; Jiang, H.; Beck, C.L.; Bowers, W.J.; Moxley, R.T.; Cannon, S.C.; Thornton, C.A. (2002). Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol Cell* **10**: 35-44.

Manley, K.; Shirley, T.L.; Flaherty, L.; Messer, A. (1999). Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet* **23**: 471-3.

Margolis, J.M.; Schoser, B.G.; Moseley, M.L.; Day, J.W.; Ranum, L.P. (2006). DM2 intronic expansions: evidence for CCUG accumulation without flanking sequence or effects on ZNF9 mRNA processing or protein expression. *Hum Mol Genet* **15**: 1808-15.

Marrosu, M.G.; Vaccargiu, S.; Marrosu, G.; Vannelli, A.; Cianchetti, C.; Muntoni, F. (1997). A novel point mutation in the peripheral myelin protein 22 (PMP22) gene associated with Charcot-Marie-Tooth disease type 1A. *Neurology* **48**: 489-93.

Martini, R.; Schachner, M. (1997). Molecular bases of myelin formation as revealed by investigations on mice deficient in glial cell surface molecules. *Glia* **19**: 298-310.

Martorell, L.; Martinez, J.M.; Carey, N.; Johnson, K.; Baiget, M. (1995). Comparison of CTG repeat length expansion and clinical progression of myotonic dystrophy over a five year period. *J Med Genet* **32**: 593-6.

Martorell, L.; Monckton, D.G.; Gamez, J.; Johnson, K.J.; Gich, I.; de Munain, A.L.; Baiget, M. (1998). Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Hum Mol Genet* **7**: 307-12.

Martorell, L.; Monckton, D.G.; Sanchez, A.; Lopez De Munain, A.; Baiget, M. (2001). Frequency and stability of the myotonic dystrophy type 1 premutation. *Neurology* **56**: 328-35.

Mathieu, J.; De Braekeleer, M.; Prevost, C. (1990). Genealogical reconstruction of myotonic dystrophy in the Saguenay-Lac-Saint-Jean area (Quebec, Canada). *Neurology* **40**: 839-42.

Matsunami, N.; Smith, B.; Ballard, L.; Lensch, M.W.; Robertson, M.; Albertsen, H.; Hanemann, C.O.; Muller, H.W.; Bird, T.D.; White, R.; et al. (1992). Peripheral myelin protein-22 gene maps in the duplication in chromosome 17p11.2 associated with Charcot-Marie-Tooth 1A. *Nat Genet* **1**: 176-9.

Matsuura, T.; Fang, P.; Pearson, C.E.; Jayakar, P.; Ashizawa, T.; Roa, B.B.; Nelson, D.L. (2006). Interruptions in the expanded ATTCT repeat of spinocerebellar ataxia type 10: repeat purity as a disease modifier? *Am J Hum Genet* **78**: 125-9.

Matsuura, T.; Yamagata, T.; Burgess, D.L.; Rasmussen, A.; Grewal, R.P.; Watase, K.; Khajavi, M.; McCall, A.E.; Davis, C.F.; Zu, L.; Achari, M.; Pulst, S.M.; Alonso, E.; Noebels, J.L.; Nelson, D.L.; Zoghbi, H.Y.; Ashizawa, T. (2000). Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat Genet* **26**: 191-4.

McMurray, C.T. (1999). DNA secondary structure: a common and causative factor for expansion in human disease. *Proc Natl Acad Sci U S A* **96**: 1823-5.

McNiven, M.A. (1998). Dynamin: a molecular motor with pinchase action. *Cell* **94**: 151-4.

Meijer, H.; de Graaff, E.; Merckx, D.M.; Jongbloed, R.J.; de Die-Smulders, C.E.; Engelen, J.J.; Fryns, J.P.; Curfs, P.M.; Oostra, B.A. (1994). A deletion of 1.6 kb proximal to the CGG repeat of the FMR1 gene causes the clinical phenotype of the fragile X syndrome. *Hum Mol Genet* **3**: 615-20.

Michalowski, S.; Miller, J.W.; Urbinati, C.R.; Paliouras, M.; Swanson, M.S.; Griffith, J. (1999). Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucleic Acids Res* **27**: 3534-42.

Miller, J.W.; Urbinati, C.R.; Teng-Umnuay, P.; Stenberg, M.G.; Byrne, B.J.; Thornton, C.A.; Swanson, M.S. (2000). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *Embo J* **19**: 4439-48.

Monckton, D.G.; Coolbaugh, M.I.; Ashizawa, K.T.; Siciliano, M.J.; Caskey, C.T. (1997). Hypermutable myotonic dystrophy CTG repeats in transgenic mice. *Nat Genet* **15**: 193-6.

Monckton, D.G.; Wong, L.J.; Ashizawa, T.; Caskey, C.T. (1995). Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum Mol Genet* **4**: 1-8.

Montermini, L.; Richter, A.; Morgan, K.; Justice, C.M.; Julien, D.; Castellotti, B.; Mercier, J.; Poirier, J.; Capozzoli, F.; Bouchard, J.P.; Lemieux, B.; Mathieu, J.; Vanasse, M.; Seni, M.H.; Graham, G.; Andermann, F.; Andermann, E.; Melancon, S.B.; Keats, B.J.; Di Donato, S.; Pandolfo, M. (1997). Phenotypic variability in Friedreich ataxia: role of the associated GAA triplet repeat expansion. *Ann Neurol* **41**: 675-82.

Morales Montero, F. (2006). Somatic mosaicism and genotype-phenotype correlations in myotonic dystrophy type 1. Degree of Doctor of Philosophy, University of Glasgow.

Morton, J.E.; Bundey, S.; Webb, T.P.; MacDonald, F.; Rindl, P.M.; Bullock, S. (1997). Fragile X syndrome is less common than previously estimated. *J Med Genet* **34**: 1-5.

Moseley, M.L.; Schut, L.J.; Bird, T.D.; Koob, M.D.; Day, J.W.; Ranum, L.P. (2000). SCA8 CTG repeat: en masse contractions in sperm and intergenerational sequence changes may play a role in reduced penetrance. *Hum Mol Genet* **9**: 2125-30.

Mounsey, J.P.; Mistry, D.J.; Ai, C.W.; Reddy, S.; Moorman, J.R. (2000). Skeletal muscle sodium channel gating in mice deficient in myotonic dystrophy protein kinase. *Hum Mol Genet* **9**: 2313-20.

Murray, A. (2000). Premature ovarian failure and the FMR1 gene. *Semin Reprod Med* **18**: 59-66.

Nagafuchi, S.; Yanagisawa, H.; Sato, K.; Shirayama, T.; Ohsaki, E.; Bundo, M.; Takeda, T.; Tadokoro, K.; Kondo, I.; Murayama, N.; et al. (1994). Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat Genet* **6**: 14-8.

Nakamoto, M.; Takebayashi, H.; Kawaguchi, Y.; Narumiya, S.; Taniwaki, M.; Nakamura, Y.; Ishikawa, Y.; Akiguchi, I.; Kimura, J.; Kakizuka, A. (1997). A CAG/CTG expansion in the normal population. *Nat Genet* **17**: 385-6.

Nakamura, K.; Jeong, S.Y.; Uchihara, T.; Anno, M.; Nagashima, K.; Nagashima, T.; Ikeda, S.; Tsuji, S.; Kanazawa, I. (2001). SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet* **10**: 1441-8.

Napieraa, M.; Krzyosiak, W.J. (1997). CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J Biol Chem* **272**: 31079-85.

Neville, C.E.; Mahadevan, M.S.; Barcelo, J.M.; Korneluk, R.G. (1994). High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. *Hum Mol Genet* **3**: 45-51.

Nolin, S.L.; Lewis, F.A., 3rd; Ye, L.L.; Houck, G.E., Jr.; Glicksman, A.E.; Limprasert, P.; Li, S.Y.; Zhong, N.; Ashley, A.E.; Feingold, E.; Sherman, S.L.; Brown, W.T. (1996). Familial transmission of the FMR1 CGG repeat. *Am J Hum Genet* **59**: 1252-61.

Novelli, G.; Gennarelli, M.; Menegazzo, E.; Mostacciuolo, M.L.; Pizzuti, A.; Fattorini, C.; Tessarolo, D.; Tomelleri, G.; Giacanelli, M.; Danieli, G.A.; et al. (1993a). (CTG)n triplet mutation and phenotype manifestations in myotonic dystrophy patients. *Biochem Med Metab Biol* **50**: 85-92.

Novelli, G.; Gennarelli, M.; Zelano, G.; Pizzuti, A.; Fattorini, C.; Caskey, C.T.; Dallapiccola, B. (1993b). Failure in detecting mRNA transcripts from the mutated allele in myotonic dystrophy muscle. *Biochem Mol Biol Int* **29**: 291-7.

O'Cochlain, D.F.; Perez-Terzic, C.; Reyes, S.; Kane, G.C.; Behfar, A.; Hodgson, D.M.; Strommen, J.A.; Liu, X.K.; Van Den Broek, W.; Wansink, D.G.; Wieringa, B.; Terzic, A. (2004). Transgenic overexpression of human DMPK accumulates into hypertrophic cardiomyopathy, myotonic myopathy and hypotension traits of myotonic dystrophy. *Hum Mol Genet* **13**: 2505-18.

O'Hoy, K.L.; Tsilfidis, C.; Mahadevan, M.S.; Neville, C.E.; Barcelo, J.; Hunter, A.G.; Korneluk, R.G. (1993). Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. *Science* **259**: 809-12.

Ophoff, R.A.; Terwindt, G.M.; Vergouwe, M.N.; van Eijk, R.; Oefner, P.J.; Hoffman, S.M.; Lamerdin, J.E.; Mohrenweiser, H.W.; Bulman, D.E.; Ferrari, M.; Haan, J.; Lindhout, D.; van Ommen, G.J.; Hofker, M.H.; Ferrari, M.D.; Frants, R.R. (1996). Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca2+ channel gene CACNL1A4. *Cell* **87**: 543-52.

Orr, H.T.; Chung, M.Y.; Banfi, S.; Kwiatkowski, T.J., Jr.; Servadio, A.; Beaudet, A.L.; McCall, A.E.; Duvick, L.A.; Ranum, L.P.; Zoghbi, H.Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* **4**: 221-6.

Otten, A.D.; Tapscott, S.J. (1995). Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc Natl Acad Sci U S A* **92**: 5465-9.

Panayiotopoulos, C.P.; Scarpalezos, S. (1976). Dystrophia myotonica. Peripheral nerve involvement and pathogenetic implications. *J Neurol Sci* **27**: 1-16.

Pareyson, D. (1999). Charcot-marie-tooth disease and related neuropathies: molecular basis for distinction and diagnosis. *Muscle Nerve* **22**: 1498-509.

Pareyson, D. (2003) Diagnosis of hereditary neuropathies in adult patients. *J Neurol* **250**: 148-60.

Pareyson, D. (2004). Differential diagnosis of Charcot-Marie-Tooth disease and related neuropathies. *Neurol Sci* **25**: 72-82.

Pareyson, D.; Scaioli, V.; Berta, E.; Sghirlanzoni, A. (1995). Acoustic nerve in peripheral neuropathy: a BAEP study. Brainstem Auditory Evoked Potentials. *Electromyogr Clin Neurophysiol* **35**: 359-64.

Pearson, C.E.; Nichol Edamura, K.; Cleary, J.D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet* **6**: 729-42.

Pearson, C.E.; Sinden, R.R. (1998). Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. *Curr Opin Struct Biol* **8**: 321-30.

Penrose, L.S. (1948). The problem of anticipation in pedigrees of dystrophia myotonica. *Ann. Eugen.* **14**: 125-32.

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

Pieretti, M.; Zhang, F.P.; Fu, Y.H.; Warren, S.T.; Oostra, B.A.; Caskey, C.T.; Nelson, D.L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**: 817-22.

Pulst, S.M.; Nechiporuk, A.; Nechiporuk, T.; Gispert, S.; Chen, X.N.; Lopes-Cendes, I.; Pearlman, S.; Starkman, S.; Orozco-Diaz, G.; Lunkes, A.; DeJong, P.; Rouleau, G.A.; Auburger, G.; Korenberg, J.R.; Figueroa, C.; Sahba, S. (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet* **14**: 269-76.

Quan, F.; Grompe, M.; Jakobs, P.; Popovich, B.W. (1995). Spontaneous deletion in the FMR1 gene in a patient with fragile X syndrome and cherubism. *Hum Mol Genet* **4**: 1681-4.

Raeymaekers, P.; Timmerman, V.; De Jonghe, P.; Swerts, L.; Gheuens, J.; Martin, J.J.; Muylle, L.; De Winter, G.; Vandenberghe, A.; Van Broeckhoven, C. (1989). Localization of the mutation in an extended family with Charcot-Marie-Tooth neuropathy (HMSN I). *Am J Hum Genet* **45**: 953-8.

Ranum, L.P.; Cooper, T.A. (2006). RNA-Mediated Neuromuscular Disorders. *Annu Rev Neurosci* **29**: 259-77.

Ranum, L.P.; Rasmussen, P.F.; Benzow, K.A.; Koob, M.D.; Day, J.W. (1998) Genetic mapping of a second myotonic dystrophy locus. *Nat Genet* **19**: 196-8.

Reddy, S.; Smith, D.B.; Rich, M.M.; Leferovich, J.M.; Reilly, P.; Davis, B.M.; Tran, K.; Rayburn, H.; Bronson, R.; Cros, D.; Balice-Gordon, R.J.; Housman, D. (1996). Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. *Nat Genet* **13**: 325-35.

Redman, J.B.; Fenwick, R.G., Jr.; Fu, Y.H.; Pizzuti, A.; Caskey, C.T. (1993). Relationship between parental trinucleotide GCT repeat length and severity of myotonic dystrophy in offspring. *JAMA* **269**: 1960-5.

Richards, R.I. (2001). Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. *Hum Mol Genet* **10**: 2187-94.

Richards, R.I.; Sutherland, G.R. (1992). Heritable unstable DNA sequences. *Nat Genet* **1**: 7-9.

Richards, R.I.; Sutherland, G.R. (1994). Simple repeat DNA is not replicated simply. *Nat Genet* **6**: 114-6.

Riley, J.; Butler, R.; Ogilvie, D.; Finniear, R.; Jenner, D.; Powell, S.; Anand, R.; Smith, J.C.; Markham, A.F. (1990). A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res* **18**: 2887-90.

Roberts, R.; Timchenko, N.A.; Miller, J.W.; Reddy, S.; Caskey, C.T.; Swanson, M.S.; Timchenko, L.T. (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**: 13221-6.

Rogoff, J.; Ziegler, D.K. (1956). Rare variant of myotonia atrophica, clinical and electro-myographic study of a family. *Brain* **79**: 349-57.

Ross, C.A.; Poirier, M.A. (2005). Opinion: What is the role of protein aggregation in neurodegeneration? *Nat Rev Mol Cell Biol* **6**: 891-8.

Rubinsztein, D.C.; Leggo, J.; Amos, W.; Barton, D.E.; Ferguson-Smith, M.A. (1994). Myotonic dystrophy CTG repeats and the associated insertion/deletion polymorphism in human and primate populations. *Hum Mol Genet* **3**: 2031-5.

Sabouri, L.A.; Mahadevan, M.S.; Narang, M.; Lee, D.S.; Surh, L.C.; Korneluk, R.G. (1993). Effect of the myotonic dystrophy (DM) mutation on mRNA levels of the DM gene. *Nat Genet* **4**: 233-8.

Sabourin, L.A.; Tamai, K.; Narang, M.A.; Korneluk, R.G. (1997). Overexpression of 3'-untranslated region of the myotonic dystrophy kinase cDNA inhibits myoblast differentiation in vitro. *J Biol Chem* **272**: 29626-35.

Sanpei, K.; Takano, H.; Igarashi, S.; Sato, T.; Oyake, M.; Sasaki, H.; Wakisaka, A.; Tashiro, K.; Ishida, Y.; Ikeuchi, T.; Koide, R.; Saito, M.; Sato, A.; Tanaka, T.; Hanyu, S.; Takiyama, Y.; Nishizawa, M.; Shimizu, N.; Nomura, Y.; Segawa, M.; Iwabuchi, K.; Eguchi, I.; Tanaka, H.; Takahashi, H.; Tsuji, S. (1996). Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet* **14**: 277-84.

Sarkar, P.S.; Appukuttan, B.; Han, J.; Ito, Y.; Ai, C.; Tsai, W.; Chai, Y.; Stout, J.T.; Reddy, S. (2000). Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts. *Nat Genet* **25**: 110-4.

Sarkar, P.S.; Paul, S.; Han, J.; Reddy, S. (2004). Six5 is required for spermatogenic cell survival and spermiogenesis. *Hum Mol Genet* **13**: 1421-31.

Savkur, R.S.; Philips, A.V.; Cooper, T.A. (2001). Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat Genet* **29**: 40-7.

Savouret, C.; Brisson, E.; Essers, J.; Kanaar, R.; Pastink, A.; te Riele, H.; Junien, C.; Gourdon, G. (2003). CTG repeat instability and size variation timing in DNA repair-deficient mice. *Embo J* **22**: 2264-73.

Scherzinger, E.; Lurz, R.; Turmaine, M.; Mangiarini, L.; Hollenbach, B.; Hasenbank, R.; Bates, G.P.; Davies, S.W.; Lehrach, H.; Wanker, E.E. (1997). Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell* **90**: 549-58.

Schofield, M.J.; Hsieh, P. (2003). DNA mismatch repair: molecular mechanisms and biological function. *Annu Rev Microbiol* **57**: 579-608.

Sergeant, N.; Sablonniere, B.; Schraen-Maschke, S.; Ghestem, A.; Maurage, C.A.; Wattez, A.; Vermersch, P.; Delacourte, A. (2001). Dysregulation of human brain microtubule-associated tau mRNA maturation in myotonic dystrophy type 1. *Hum Mol Genet* **10**: 2143-55.

Seznec, H.; Agbulut, O.; Sergeant, N.; Savouret, C.; Ghestem, A.; Tabti, N.; Willer, J.C.; Ourth, L.; Duros, C.; Brisson, E.; Fouquet, C.; Butler-Browne, G.; Delacourte, A.; Junien, C.; Gourdon, G. (2001). Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. *Hum Mol Genet* **10**: 2717-26.

Shelbourne, P.; Davies, J.; Buxton, J.; Anvret, M.; Blennow, E.; Bonduelle, M.; Schmedding, E.; Glass, I.; Lindenbaum, R.; Lane, R.; et al. (1993) Direct diagnosis of myotonic dystrophy with a disease-specific DNA marker. *N Engl J Med* **328**: 471-5.

Shelbourne, P.; Winqvist, R.; Kunert, E.; Davies, J.; Leisti, J.; Thiele, H.; Bachmann, H.; Buxton, J.; Williamson, B.; Johnson, K. (1992). Unstable DNA may be responsible for the incomplete penetrance of the myotonic dystrophy phenotype. *Hum Mol Genet* **1**: 467-73.

Shelbourne, P.F.; Monckton, D.G. (2006). Somatic mosaicism of expanded CAGCTG repeats in humans and mice: dynamics, mechanisms,and consequences. In: Wells D and Ashizawa T (ed) Genetic Instabilities and Neurological Diseases. Oxford: Academic, pp 537-61

Sherman, S.L.; Jacobs, P.A.; Morton, N.E.; Froster-Iskenius, U.; Howard-Peebles, P.N.; Nielsen, K.B.; Partington, M.W.; Sutherland, G.R.; Turner, G.; Watson, M. (1985). Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum Genet* **69**: 289-99.

Sherman, S.L.; Morton, N.E.; Jacobs, P.A.; Turner, G. (1984). The marker (X) syndrome: a cytogenetic and genetic analysis. *Ann Hum Genet* **48**: 21-37.

Shy, M.E. (2004). Charcot-Marie-Tooth disease: an update. *Curr Opin Neurol* **17**: 579-85.

Silveira, I.; Alonso, I.; Guimaraes, L.; Mendonca, P.; Santos, C.; Maciel, P.; Fidalgo De Matos, J.M.; Costa, M.; Barbot, C.; Tuna, A.; Barros, J.; Jardim, L.; Coutinho, P.; Sequeiros, J. (2000). High germinal instability of the (CTG)n at the SCA8 locus of both expanded and normal alleles. *Am J Hum Genet* **66**: 830-40.

Sinden, R.R.; Potaman, V.N.; Oussatcheva, E.A.; Pearson, C.E.; Lyubchenko, Y.L.; Shlyakhtenko, L.S. (2002). Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J Biosci* **27**: 53-65.

Skre, H. (1974). Genetic and clinical aspects of Charcot-Marie-Tooth's disease. *Clin Genet* **6**: 98-118.

Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* **191**: 528-35.

Sofola, O.A.; Jin, P.; Botas, J.; Nelson, D.L. (2007). Argonaute-2-dependent rescue of a Drosophila model of FXTAS by FRAXE premutation repeat. *Hum Mol Genet* **16**: 2326-32.

Spaans, F.; Faber, C.G.; Smeets, H.J.M.; Hofman, P.A.M.; Braida, C.; Monckton, D.G.; de Die-Smulders, C.E.M. (2008). Encephalopathic attacks in a family co-segregating myotonic dystrophy type 1 and a CMT neuropathy. *Manuscript submitted.*

Spaans, F.; Jennekens, F.G.; Mirandolle, J.F.; Bijlsma, J.B.; de Gast, G.C. (1986). Myotonic dystrophy associated with hereditary motor and sensory neuropathy. *Brain* **109**: 1149-68.

Steinbach, P.; Glaser, D.; Vogel, W.; Wolf, M.; Schwemmle, S. (1998). The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. *Am J Hum Genet* **62**: 278-85.

Stevanin, G.; Camuzat, A.; Holmes, S.E.; Julien, C.; Sahloul, R.; Dode, C.; Hahn-Barma, V.; Ross, C.A.; Margolis, R.L.; Durr, A.; Brice, A. (2002). CAG/CTG repeat expansions at the Huntington's disease-like 2 locus are rare in Huntington's disease patients. *Neurology* **58**: 965-7.

Storbeck, C.J.; Drmanic, S.; Daniel, K.; Waring, J.D.; Jirik, F.R.; Parry, D.J.; Ahmed, N.; Sabourin, L.A.; Ikeda, J.E.; Korneluk, R.G. (2004). Inhibition of myogenesis in transgenic mice expressing the human DMPK 3' UTR. *Hum Mol Genet* **15**: 589-600.

Suter, U.; Scherer, S.S. (2003). Disease mechanisms in inherited neuropathies. *Nat Rev Neurosci* **4**: 714-26.

Taneja, K.L.; McCurrach, M.; Schalling, M.; Housman, D.; Singer, R.H. (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J Cell Biol* **128**: 995-1002.

Tassone, F.; Hagerman, R.J.; Garcia-Arocena, D.; Khandjian, E.W.; Greco, C.M.; Hagerman, P.J. (2004) .Intranuclear inclusions in neural cells with premutation alleles in fragile X associated tremor/ataxia syndrome. *J Med Genet* **41**: e43.

Tassone, F.; Hagerman, R.J.; Loesch, D.Z.; Lachiewicz, A.; Taylor, A.K.; Hagerman, P.J. (2000a). Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. *Am J Med Genet* **94**: 232-6.

Tassone, F.; Hagerman, R.J.; Taylor, A.K.; Gane, L.W.; Godfrey, T.E.; Hagerman, P.J. (2000b). Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am J Hum Genet* **66**: 6-15.

Thornton, C.A.; Johnson, K.; Moxley, R.T., 3rd (1994). Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann Neurol* **35**: 104-7.

Thornton, C.A.; Wymer, J.P.; Simmons, Z.; McClain, C.; Moxley, R.T., 3rd (1997). Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat Genet* **16**: 407-9.

Timchenko, L.T.; Miller, J.W.; Timchenko, N.A.; DeVore, D.R.; Datar, K.V.; Lin, L.; Roberts, R.; Caskey, C.T.; Swanson, M.S. (1996a). Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res* **24**: 4407-14.

Timchenko, L.T.; Timchenko, N.A.; Caskey, C.T.; Roberts, R. (1996b). Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. *Hum Mol Genet* **5**: 115-21.

Timchenko, N.A.; Cai, Z.J.; Welm, A.L.; Reddy, S.; Ashizawa, T.; Timchenko, L.T. (2001). RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *J Biol Chem* **276**: 7820-6.

Timmerman, V.; Nelis, E.; Van Hul, W.; Nieuwenhuijsen, B.W.; Chen, K.L.; Wang, S.; Ben Othman, K.; Cullen, B.; Leach, R.J.; Hanemann, C.O.; et al. (1992). The peripheral myelin protein gene PMP-22 is contained within the Charcot-Marie-Tooth disease type 1A duplication. *Nat Genet* **1**: 171-5.

Tiscornia, G.; Mahadevan, M.S. (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**: 959-67.

Trinh, T., Jessee J, Bloom F. Hirsch, V (1994). STBL2™: An Escherichia coli Strain for the Stable Propagation of Retroviral Clones and Direct Repeat Sequences. *FOCUS* **16**: 78-80.

Valentijn, L.J.; Bolhuis, P.A.; Zorn, I.; Hoogendijk, J.E.; van den Bosch, N.; Hensels, G.W.; Stanton, V.P., Jr.; Housman, D.E.; Fischbeck, K.H.; Ross, D.A.; et al. (1992). The peripheral myelin gene PMP-22/GAS-3 is duplicated in Charcot-Marie-Tooth disease type 1A. *Nat Genet* **1**: 166-70.

van den Broek, W.J.; Nelen, M.R.; Wansink, D.G.; Coerwinkel, M.M.; te Riele, H.; Groenen, P.J.; Wieringa, B. (2002). Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum Mol Genet* **11**: 191-8.

Vance, J.M.; Nicholson, G.A.; Yamaoka, L.H.; Stajich, J.; Stewart, C.S.; Speer, M.C.; Hung, W.Y.; Roses, A.D.; Barker, D.; Pericak-Vance, M.A. (1989). Linkage of Charcot-Marie-Tooth neuropathy type 1a to chromosome 17. *Exp Neurol* **104**: 186-9.

Verkerk, A.J.; Pieretti, M.; Sutcliffe, J.S.; Fu, Y.H.; Kuhl, D.P.; Pizzuti, A.; Reiner, O.; Richards, S.; Victoria, M.F.; Zhang, F.P.; et al. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**: 905-14.

von Giesen, H.J.; Stoll, G.; Koch, M.C.; Beneck, R. (1994). Mixed axonal-demyelinating polyneuropathy as predominant manifestation of myotonic dystrophy. *Muscle Nerve* **17**: 701-3.

Wakimoto, H.; Maguire, C.T.; Sherwood, M.C.; Vargas, M.M.; Sarkar, P.S.; Han, J.; Reddy, S.; Berul, C.I. (2002). Characterization of cardiac conduction system abnormalities in mice with targeted disruption of Six5 gene. *J Interv Card Electrophysiol* **7**: 127-35.

Wang, J.; Pegoraro, E.; Menegazzo, E.; Gennarelli, M.; Hoop, R.C.; Angelini, C.; Hoffman, E.P. (1995). Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation. *Hum Mol Genet* **4**: 599-606.

Wang, Y.H.; Amirhaeri, S.; Kang, S.; Wells, R.D.; Griffith, J.D. (1994). Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science* **265**: 669-71.

Wang, Y.H.; Griffith, J. (1995). Expanded CTG triplet blocks from the myotonic dystrophy gene create the strongest known natural nucleosome positioning elements. *Genomics* **25**: 570-3.

Wang, Y.H.; Griffith, J. (1996). Methylation of expanded CCG triplet repeat DNA from fragile X syndrome patients enhances nucleosome exclusion. *J Biol Chem* **271**: 22937-40.

Warner, J.P.; Barron, L.H.; Goudie, D.; Kelly, K.; Dow, D.; Fitzpatrick, D.R.; Brock, D.J. (1996). A general method for the detection of large CAG repeat expansions by fluorescent PCR. *J Med Genet* **33**: 1022-6.

Wells, R.D. (1996). Molecular basis of genetic instability of triplet repeats. *J Biol Chem* **271**: 2875-8.

Wheeler, V.C.; Lebel, L.A.; Vrbanac, V.; Teed, A.; te Riele, H.; MacDonald, M.E. (2003). Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. *Hum Mol Genet* **12**: 273-81.

Willemsen, R.; Hoogeveen-Westerveld, M.; Reis, S.; Holstege, J.; Severijnen, L.A.; Nieuwenhuizen, I.M.; Schrier, M.; van Unen, L.; Tassone, F.; Hoogeveen, A.T.; Hagerman, P.J.; Mientjes, E.J.; Oostra, B.A. (2003). The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome. *Hum Mol Genet* **12**: 949-59.

Winchester, C.L.; Ferrier, R.K.; Sermoni, A.; Clark, B.J.; Johnson, K.J. (1999). Characterization of the expression of DMPK and SIX5 in the human eye and implications for pathogenesis in myotonic dystrophy. *Hum Mol Genet* **8**: 481-92.

Wojciechowska, M.; Bacolla, A.; Larson, J.E.; Wells, R.D. (2005). The myotonic dystrophy type 1 triplet repeat sequence induces gross deletions and inversions. *J Biol Chem* **280**: 941-52.

Wong, L.J.; Ashizawa, T.; Monckton, D.G.; Caskey, C.T.; Richards, C.S. (1995). Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am J Hum Genet* **56**: 114-22.

Yamagata, H.; Miki, T.; Ogihara, T.; Nakagawa, M.; Higuchi, I.; Osame, M.; Shelbourne, P.; Davies, J.; Johnson, K. (1992). Expansion of unstable DNA region in Japanese myotonic dystrophy patients. *Lancet* **339**: 692.

Yamagata, H.; Nakagawa, M.; Johnson, K.; Miki, T. (1998). Further evidence for a major ancient mutation underlying myotonic dystrophy from linkage disequilibrium studies in the Japanese population. *J Hum Genet* **43**: 246-9.

Zhong, N.; Yang, W.; Dobkin, C.; Brown, W.T. (1995). Fragile X gene instability: anchoring AGGs and linked microsatellites. *Am J Hum Genet* **57**: 351-61.

Zoghbi, H.Y.; Orr, H.T. (2000). Glutamine repeats and neurodegeneration. *Annu Rev Neurosci* **23**: 217-47.

Zuchner, S.; Noureddine, M.; Kennerson, M.; Verhoeven, K.; Claeys, K.; De Jonghe, P.; Merory, J.; Oliveira, S.A.; Speer, M.C.; Stenger, J.E.; Walizada, G.; Zhu, D.; Pericak-Vance, M.A.; Nicholson, G.; Timmerman, V.; Vance, J.M. (2005). Mutations in the Pleckstrin Homology Domain of Dynamin 2 Cause Dominant Intermediate Charcot-Marie-Tooth Disease: LBS.002. *Neurology* **64**: 1826.

Zuhlke, C.; Dalski, A.; Hellenbroich, Y.; Bubel, S.; Schwinger, E.; Burk, K. (2002). Spinocerebellar ataxia type 1 (SCA1): phenotype-genotype correlation studies in intermediate alleles. *Eur J Hum Genet* **10**: 204-9.