Am. J. Hum. Genet. 52:1164–1174, 1993

Size of the Unstable CTG Repeat Sequence in Relation to Phenotype and Parental Transmission in Myotonic Dystrophy

Helen G. Harley, Shelley A. Rundle, John C. MacMillan, Jenny Myring, J. David Brook,[1] Stephen Crow, William Reardon,[2] Iain Fenton, Duncan J. Shaw, and Peter S. Harper

Institute of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff

Summary

A clinical and molecular analysis of 439 individuals affected with myotonic dystrophy, from 101 kindreds, has shown that the size of the unstable CTG repeat detected in nearly all cases of myotonic dystrophy is related both to age at onset of the disorder and to the severity of the phenotype. The largest repeat sizes (1.5–6.0 kb) are seen in patients with congenital myotonic dystrophy, while the minimally affected patients have repeat sizes of <0.5 kb. Comparison of parent-child pairs has shown that most offspring have an earlier age at onset and a larger repeat size than their parents, with only 4 of 182 showing a definite decrease in repeat size, accompanied by a later age at onset or less severe phenotype. Increase in repeat size from parent to child is similar for both paternal and maternal transmissions when the increase is expressed as a proportion of the parental repeat size. Analysis of congenitally affected cases shows not only that they have, on average, the largest repeat sizes but also that their mothers have larger mean repeat sizes, supporting previous suggestions that a maternal effect is involved in the pathogenesis of this form of the disorder.

Introduction

The mutation causing myotonic dystrophy (DM) is an unstable CTG repeat sequence (Aslanidis et al. 1992; Buxton et al. 1992; Harley et al. 1992*a*) in the 3′ untranslated region (3′-UTR) of a gene whose sequence predicts the protein product to be a member of the protein kinase family (Brook et al. 1992; Fu et al. 1992; Mahadevan et al. 1992). Normal individuals have ≤30 copies of this CTG repeat, but in DM the number varies from 50 to >2,000 (Harley et al. 1992*b*). It has already been shown that those individuals with the smallest increase in repeat number are minimally affected, some showing cataract as the sole symptom while others are entirely normal clinically (Brook et al. 1992; Fu et al. 1992; Harley et al. 1992*b*; Mahadevan et al. 1992; Reardon et al. 1992*b*). Progressive expansion of the sequence occurs on transmission, with increasing severity and earlier onset of disease, in successive generations of a family, thus explaining the long-debated phenomenon of "anticipation" in DM (Harper et al. 1992).

We have reported preliminary data relating the size (or copy number) of this unstable repeat to approximate category of severity (Harley et al. 1992*b*) and have indicated the clinical implications of the molecular findings. In the present paper we describe a more comprehensive analysis of the relationship of phenotype to molecular abnormality in a series of 439 DM patients from 101 families. We also examine the relationship of sex and size of repeat in the transmitting parent to the size of repeat in the offspring.

Methods

Clinical Aspects

Families with DM came from the extensive series investigated by our institute over a period of 20 years (Harper 1989). They do not represent an unselected

Received November 6, 1992; revision received January 25, 1993.

Address for correspondence and reprints: Helen G. Harley, Institute of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, United Kingdom.

1. Present address: Department of Medical Genetics, University of Nottingham City Hospital, Nottingham.

2. Present address: Institute of Child Health, University of London, London.

0002-9297/93/5206-0018$02.00

sample or completely ascertained population of DM patients, a factor which requires consideration in interpreting the data. The three principal sources were as follows:

1. Families from South Wales examined personally by at least one of us and probably representing almost all DM families in this region.
2. Families from other parts of Britain, either seen by request for diagnosis and genetic counseling or forming part of a nationwide study of congenital DM. These families were also examined clinically by at least one of us but are biased toward ascertainment through a congenitally affected case.
3. Families referred primarily for molecular diagnosis from all parts of Britain and from other (mostly European) countries. Although many were examined personally, clinical data were often dependent on the referring center and thus are limited; data were not included unless considered to be reliable.

Age at onset was recorded as the age when the first symptoms clearly attributable to DM appeared, including cataract as well as myotonia and muscle weakness. Where onset could not be determined precisely, a 10-year period was recorded, and the midpoint was used for analysis. Four graded clinical categories were used:

a. Minimal—Cataract was the principal clinical feature. Neuromuscular abnormalities were absent or mild, with onset at >50 years.
b. Classical—Myotonia and progressive muscle weakness, generally presenting in adult life.
c. Congenital—Symptoms were present from birth or in utero. These included respiratory insufficiency, hypotonia, and developmental delay in survivors, a strikingly different clinical picture from cases with the "classical" form.
d. Early childhood—This relatively small group showed some features comparable to those in cases with congenital onset, notably developmental delay, but had no documented abnormality at the time of birth.

Individuals were scored as affected if there was clear evidence of myotonia or progressive muscle weakness, or cataracts in the case of known gene carriers. In order to account for the possibility of ascertainment bias, analyses of the data were done both with and without propositi. Unless indicated otherwise, there were no statistically significant differences between pairs of analyses.

DNA Analysis

Molecular analyses were performed according to methods described elsewhere (Brook et al. 1992; Harley et al. 1992*a*, 1992*b*; Reardon et al. 1992*b*). Increase in size of DNA fragments seen on *Eco*RI and *Pst*I digestion of DNA isolated from white blood cells was measured to the nearest 0.5 kb, while all affected individuals with repeat sizes <0.5 kb were also analyzed by PCR. In those samples where smears could be detected, indicative of somatic expansion of the repeat giving a range of sizes, the presence of the smear was noted, and both its lower and upper limits were sized. When smears were present, the lower limit of the smear was used in any statistical analyses. All molecular results are expressed in terms of kilobases of additional DNA.

Results

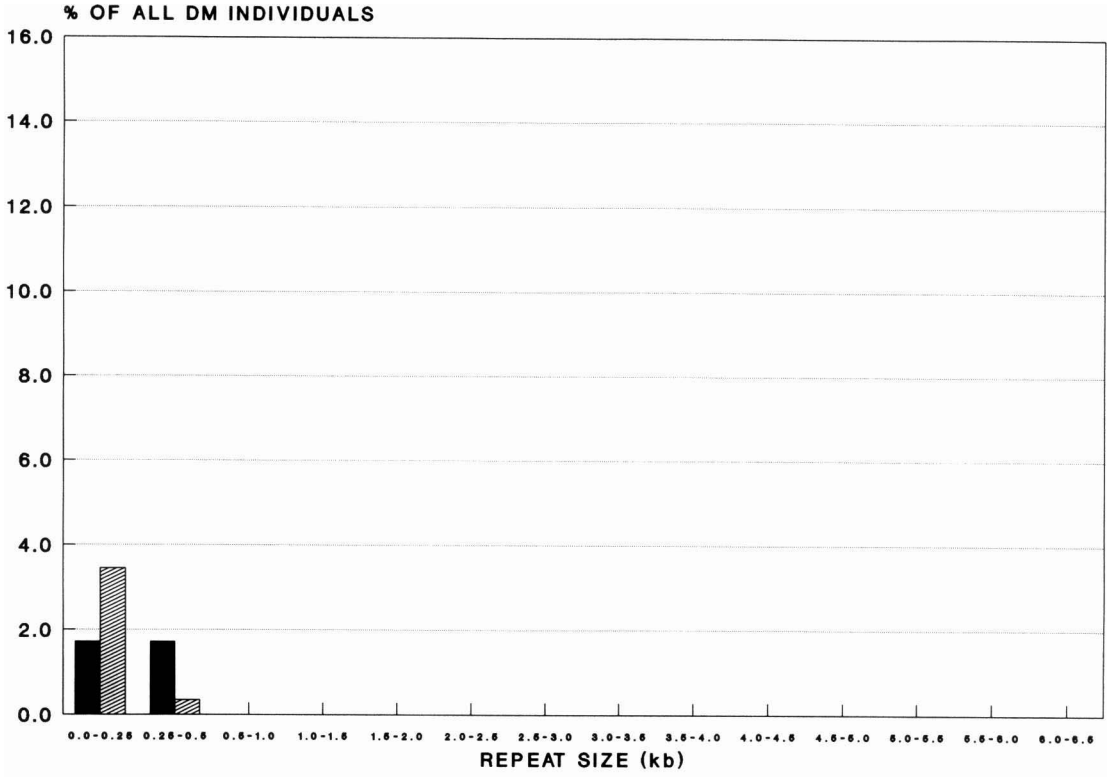
Relationship of Size of Repeat to Phenotype in Individual Patients

Clinical and molecular data were available on a total of 439 affected individuals, from 101 apparently unrelated kindreds. Figure 1 shows the range of repeat size seen in each clinical category, separated by sex of individual. In general, the more severe the phenotype, the larger the repeat size. The only significant difference found between the sexes was in the minimally affected group, with an excess of males having the smallest repeat size (<0.25 kb). Figure 2 shows the relationship of repeat size to apparent age at onset, which shows strong correlation when repeat size is plotted on a logarithmic scale ($r = -.816$, $P < .001$). No significant difference was seen when the data were divided by sex.

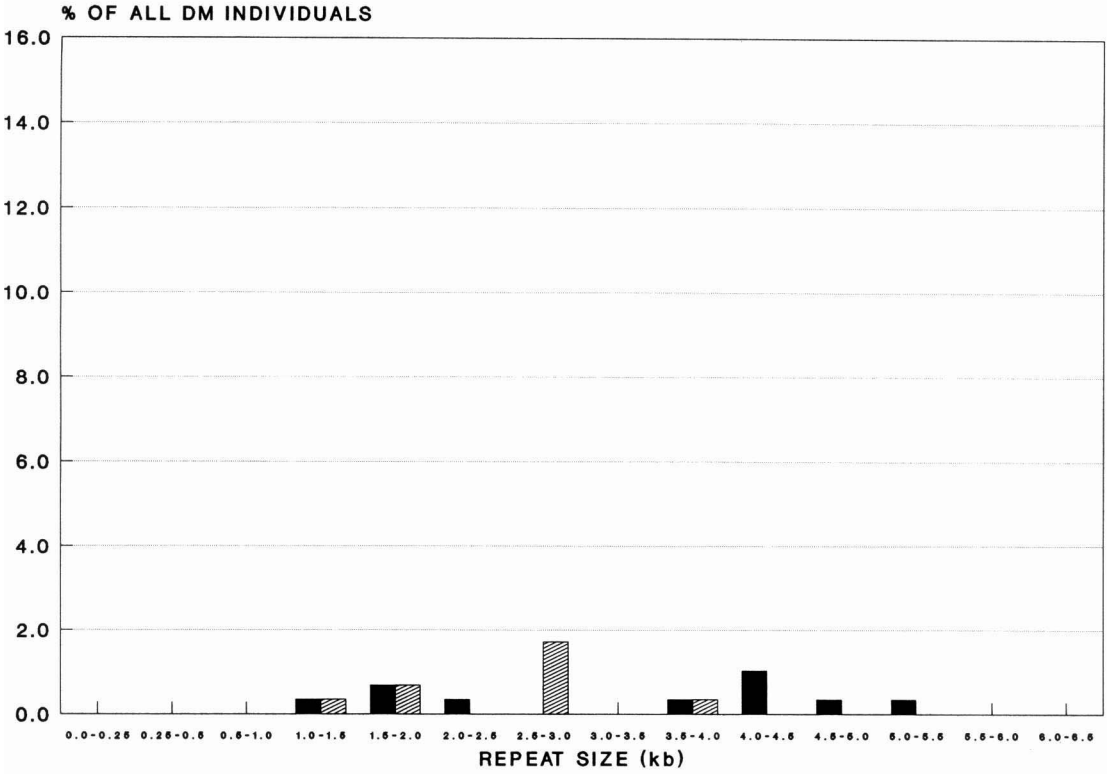
Intergenerational Differences in Phenotype and Repeat Size

For parent and child pairs, age at onset is shown in the top graph of figure 3, and repeat size is shown in the bottom graph of figure 3. Both analyses show a weak correlation between parent and child ($r = .723$, $P < .001$; and $r = .504$, $P < .001$, respectively), but the striking feature is that, in most cases, age at onset is earlier in child than in parent, while repeat size is correspondingly greater in the children. There are 3 cases with onset apparently later in the child than in the parent, and there are 14 cases where the repeat size has decreased. The exceptions in age at onset may be explained by the method of ascertainment, which was often based on a patient's recollection of his or her first awareness of muscular problems. It was more common

(a)


% OF ALL DM INDIVIDUALS
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0.0-0.25
0.25-0.5
0.5-1.0
1.0-1.5
1.5-2.0
2.0-2.5
2.5-3.0
3.0-3.5
3.5-4.0
4.0-4.5
4.5-5.0
5.0-5.5
5.5-6.0
6.0-6.5
REPEAT SIZE (kb)

(c)


% OF ALL DM INDIVIDUALS
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0.0-0.25
0.25-0.5
0.5-1.0
1.0-1.5
1.5-2.0
2.0-2.5
2.5-3.0
3.0-3.5
3.5-4.0
4.0-4.5
4.5-5.0
5.0-5.5
5.5-6.0
6.0-6.5
REPEAT SIZE (kb)

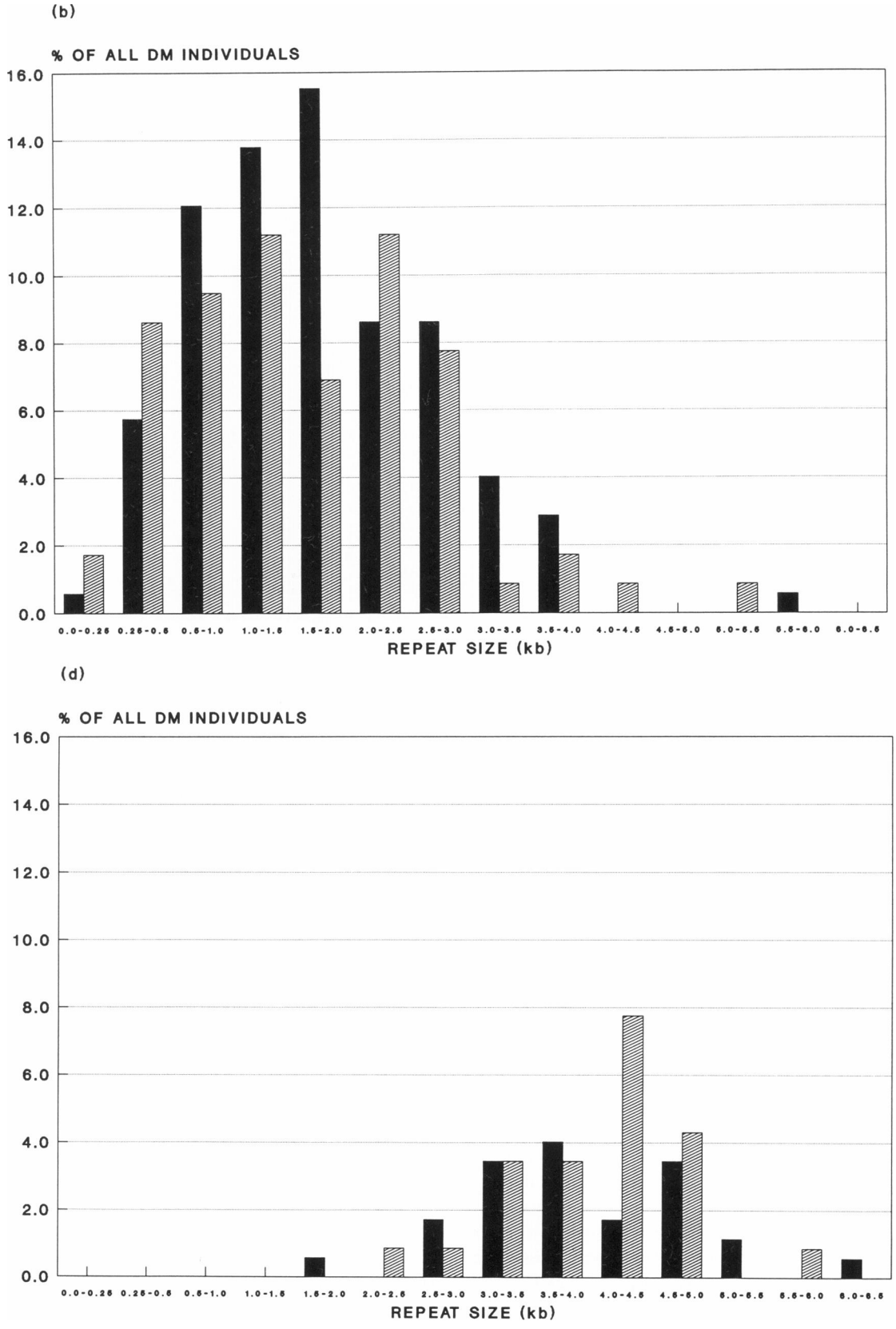


Figure 1 CTG repeat sequence sizes (in kb) for DM patients, in minimal (*a*), classical (*b*), early childhood (*c*), and congenital (*d*) categories. The results are expressed as the percentage of the total number of male (hatched bars) or female (black bars) patients with repeat sequences of the size indicated on the *X*-axis. The total number of patients was 301.

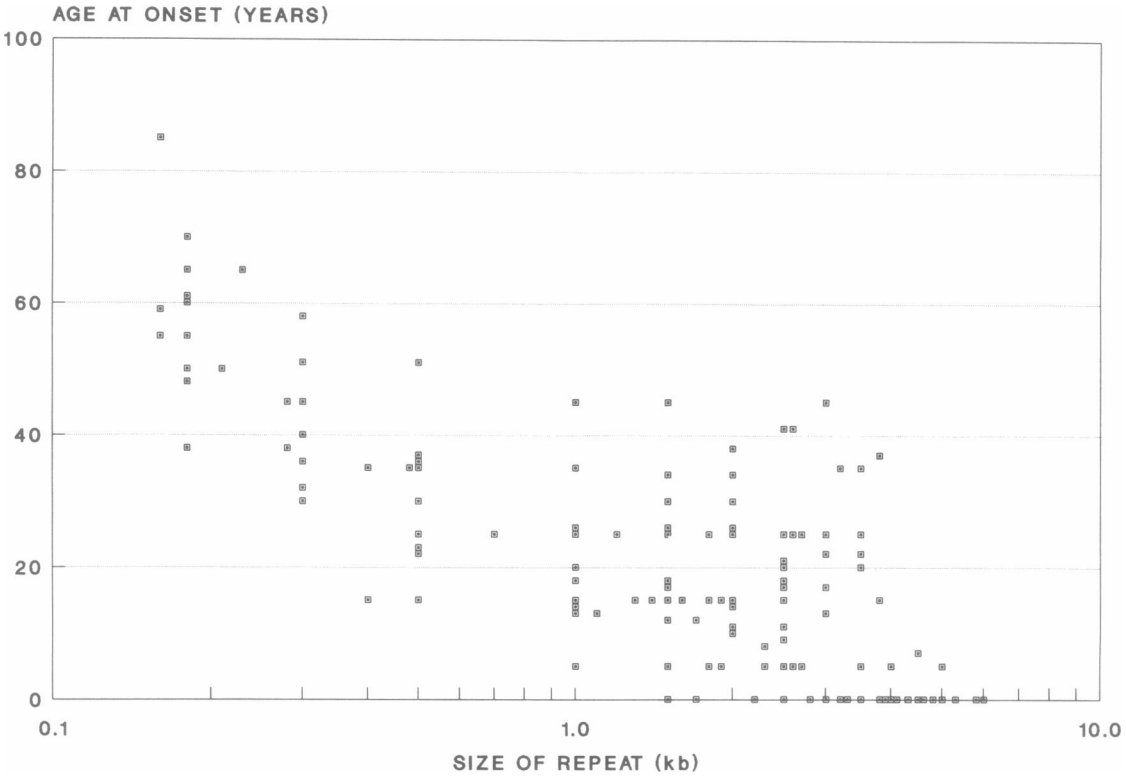


Figure 2 Age at onset for 218 DM patients, plotted against CTG repeat length (logarithmic scale).

for a child's age at onset to be remembered more accurately than that of the parent. Of the 14 cases where there has been an apparent decrease in size of repeat, 4 were maternally transmitted, and 10 were paternally transmitted. However, only 4 are considered as true decreases, the other 10 parent-child pairs having some overlap in the range of repeat sizes detected. Although the lower limit of a smear detected in an individual was used in the statistical analyses and graphical representations, this could lead to an overestimate of the number of true decreases if overlap of smears between a parent and child is not taken into account; however, we are reporting all 14 cases, for completeness. Also to be considered is the fact that blood samples are usually taken from families at the same time, and we cannot normalize the data to take account of age at time of sampling, which may well have an effect on the amount of detectable mosaicism of repeat size. A general observation was that smears were more likely to be seen in older individuals and in those with larger repeat sizes. The four cases where repeat size was clearly smaller in the child than in the parent are (i) a decrease from 1.0 kb to 0.38 kb, with a change in phenotype from classical (age at onset in the 3d decade) to asymptomatic, at age 24 years; (ii) a decrease from 0.50 kb to 0.48 kb, with a change in phenotype from classical to minimal; (iii) a decrease from 0.50 kb to 0.45 kb, with a change in phenotype from classical to asymptomatic; and (iv) a decrease from 0.40 kb to 0.23 kb, with a change in phenotype from classical (age at onset in the 4th decade) to asymptomatic, at age 32 years. In all four cases the decrease was transmitted by an affected father.

Analysis of sib pairs for age at onset revealed a correlation ($r = .718$, $P < .001$) which was not significantly different from that seen between parent and child. For repeat size, there was stronger correlation for sib pairs ($r = .651$, $P < .001$) than was seen between parent and child. No effect of birth order was seen in either analysis.

Sex of Transmitting Parent and Repeat Size

The effect of the sex of the transmitting parent on the increase in repeat size on transmission from parent to child is shown in figure 4. The absolute increase in

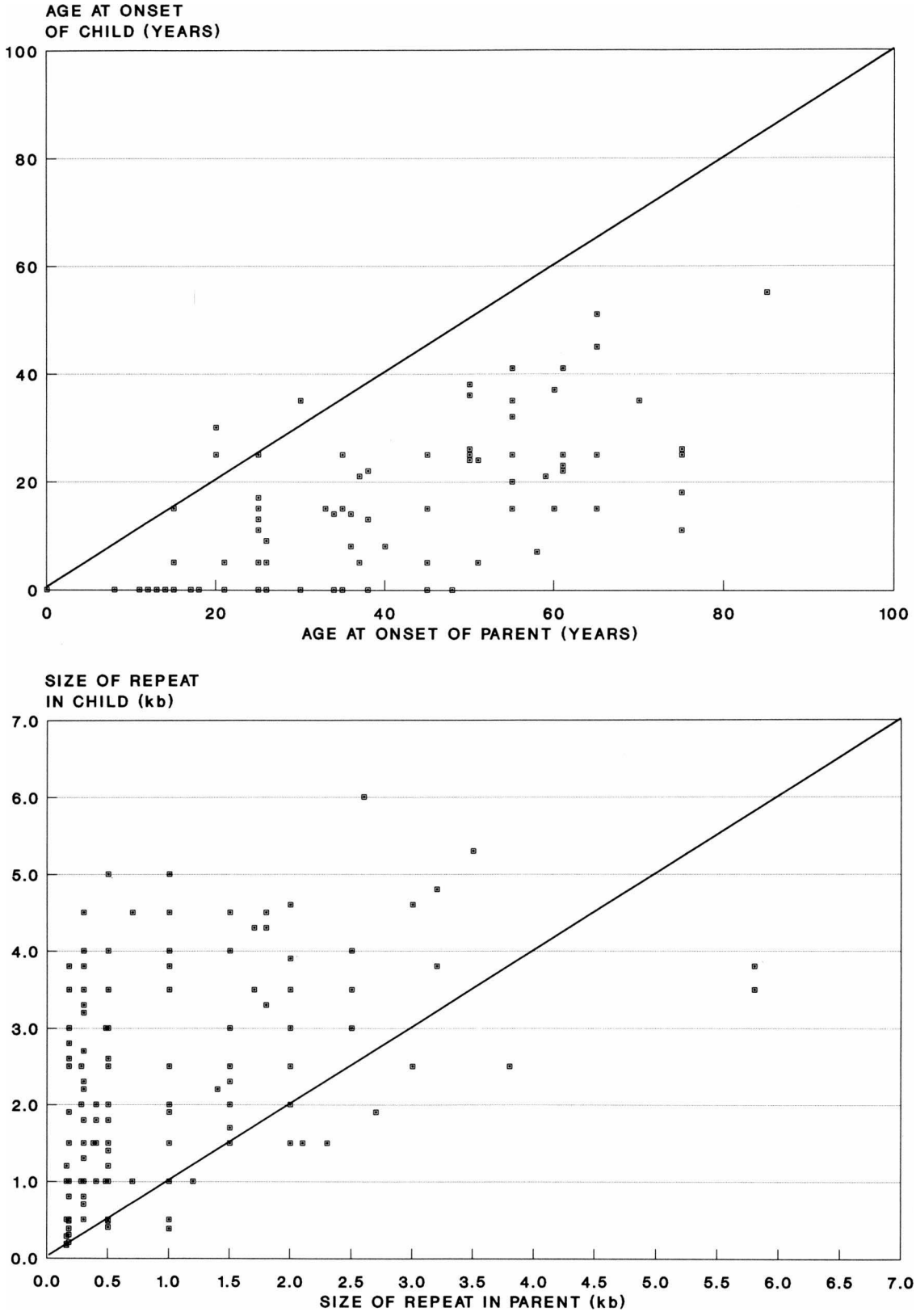


Figure 3 *Top*, Ages at onset for 149 parent-child pairs. Points below the diagonal indicate onset earlier in child than in parent. *Bottom*, CTG repeat sizes for 181 parent-child pairs. Points above the diagonal indicate repeat size greater in child than in parent.

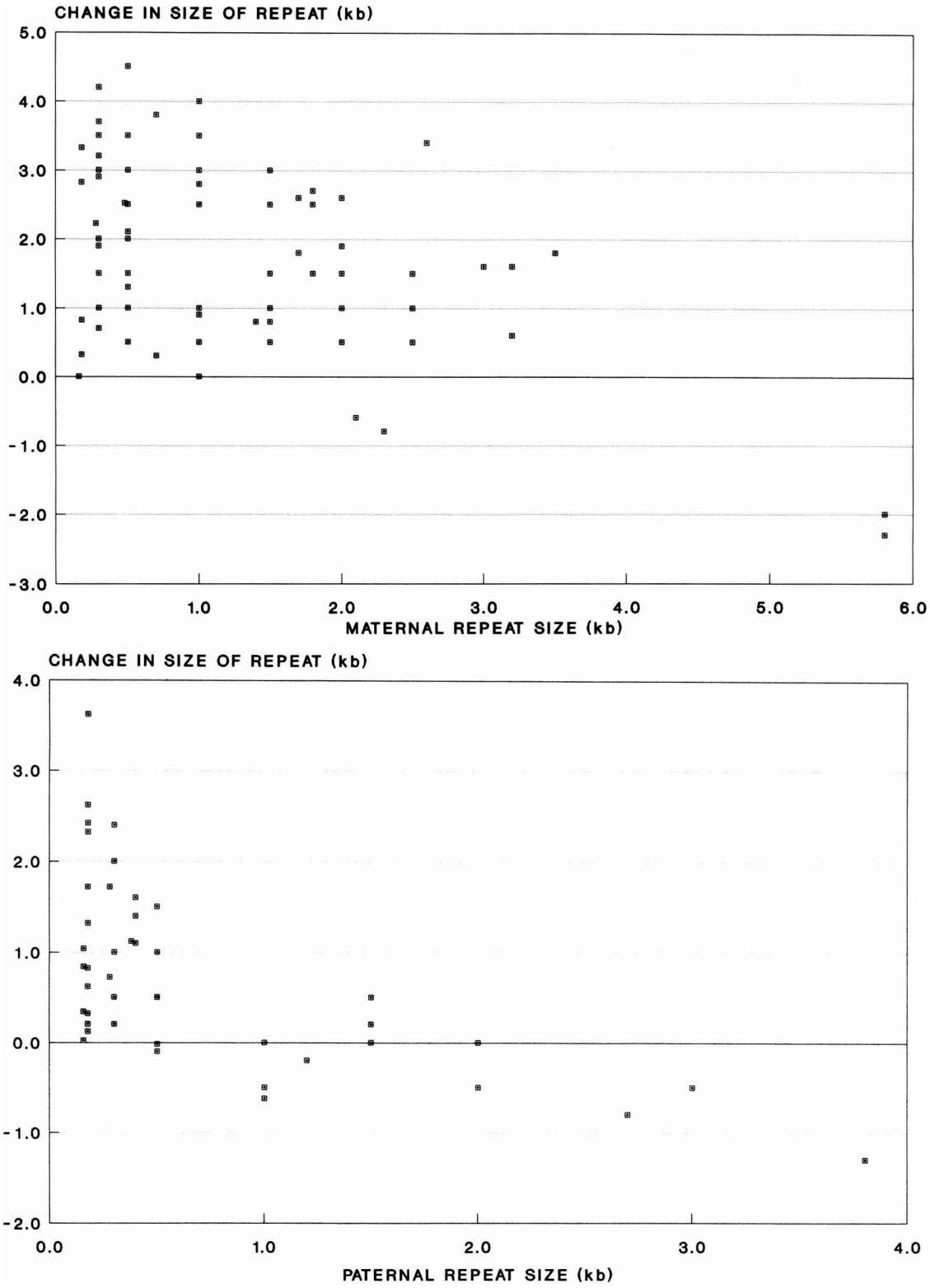


Figure 4 Change in repeat size (in kb) on transmission from mothers to 98 offspring (*top*) or on transmission from fathers to 83 offspring (*bottom*), plotted against repeat size of parent.

size of repeat is greater on transmission from females (mean = 1.48 kb, SD = 1.41) than on transmission from males (mean = 0.83, SD = 0.84). However, when the increase is expressed as a proportion of the size in the parent, this sex difference is no longer seen. The apparent difference is the result of a repeat size which, on average, is larger in mothers than in fathers (mean = 1.13, SD = 0.75; and mean = 0.59, SD = 0.73, respectively; $P < .01$). Although there is a correlation between the absolute size of repeat in the parent and the change in size when it is transmitted to the offspring, the confidence limits are wide ($r = -.3167$, $P = .001$, 95% confidence limits = $-.127$ to $-.484$). No father with a repeat size >2.0 kb (5/83 transmissions) has passed a larger-sized repeat to his offspring, whereas this was not the case for mothers with repeat size >2.0 kb (22/93 transmissions).

The Molecular Basis of Congenital DM

All cases of congenital DM were transmitted by affected mothers. However, in those families where the affected grandparent could be identified, there was an excess of affected grandfathers (28 of 43 cases, significant at $P < .005$).

The largest repeat sizes are seen in individuals who are congenitally affected, although there is considerable overlap between this group and other clinical groups. Figure 5 shows repeat size in offspring, plotted against that of their parents, subdivided into paternal transmission (all offspring), maternal transmission (minimal or classical offspring), and maternal transmission (congenital offspring). The maternal transmissions giving rise to congenital offspring form a separate group, in terms of their distribution on the graph, from those giving rise to noncongenital offspring. The mothers of congenital offspring have repeat sizes significantly greater than those of mothers of noncongenital offspring (mean = 1.8 kb and 0.65 kb, respectively; $P < .01$). Fathers have repeat sizes (mean = 0.59 kb) similar to those of mothers of noncongenital offspring.

Discussion

The recognition that expansion of an unstable CTG repeat is the mutational basis underlying DM has begun to resolve many of the hitherto puzzling clinical and genetic aspects of the disorder. A specific biological mechanism for explaining the marked variability in severity and age at onset, particularly between different generations of the same family, is now available, while the phenomenon of anticipation, long disputed but recently validated in terms of pedigree data, can also be explained.

The data presented here provide further evidence that the size of the repeat is related to the phenotype of DM. This can be seen both when patients are categorized by clinical severity and when they are categorized by age at onset. Minimally affected patients usually show a small expansion, <300 bp; when these are analyzed using PCR, they show 50–100 CTG repeats. Twenty-eight of 43 grandparents in families with a congenitally affected child are grandfathers. The male excess in this group is of interest and has also been seen in previous family studies (Bell 1947). Since minimally affected individuals are principally ascertained through an affected child or grandchild and would probably remain undetected in the absence of such a relative, the sex difference may reflect a greater tendency to initial instability in male meiosis, rather than an absolute excess of minimally affected males in the population.

The other group of patients showing a marked relationship to size of repeat is the severe group with congenital onset, whose range of expansion varies from 2 to >6 kb. While this range overlaps that of other childhood-onset cases and, to a lesser extent, that of the classical group, mostly with adult onset, there is an absolute distinction, in our series, between it and the "minimal" group.

The correlations shown here for age at onset between parent and child and between sibs correspond closely with those found in much earlier studies; Bell (1947) and Penrose (1948) commented on the low parent-child correlation. The similarity of our own data to these older family studies suggests that the biases in ascertainment in our study, already mentioned above, are not likely to have masked the fundamental genetic patterns in transmission. When size of repeat is substituted for age at onset, a strikingly similar pattern is seen, reflecting the close correlation betweeen these two variables. The tendency to "anticipation" is strikingly demonstrated in figure 3, with only a few instances of later age at onset or reduction in repeat size in the offspring noted. Howeler et al. (1989) showed a later age at onset for the child in only 1 of 61 parent-child pairs. That such instances do occur is illustrated by the cases described, while the undoubted bias of ascertainment, toward those families where instability has already occurred, makes it possible that they will prove to be more frequent in the population overall. Undetected "minimal" cases in the younger generation

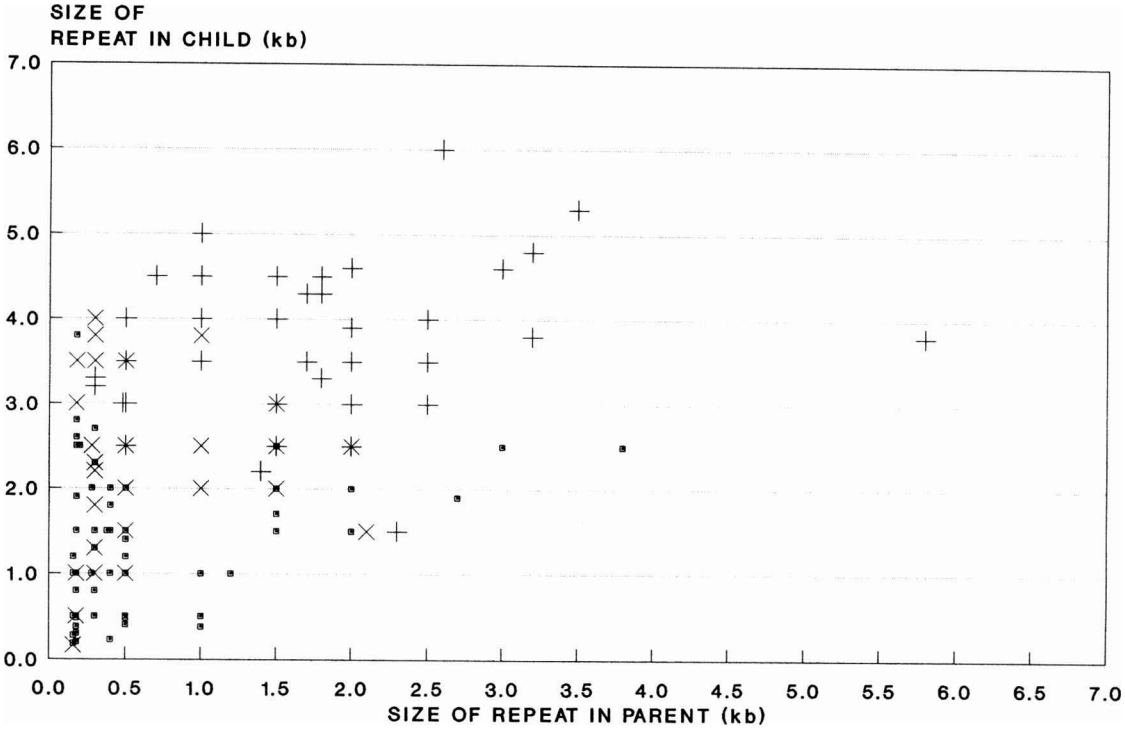


Figure 5 Repeat size in parent-child pairs, separated into maternally transmitted congenital onset (+), maternally transmitted noncongenital onset (×), and paternally transmitted (■). In these three categories, the number of individuals was 46, 33, and 83, respectively.

may exist, as originally postulated by Penrose (1948). Two separate studies of presymptomatic detection using linked DNA markers have suggested that around 10% of clinically normal sibs of affected individuals may carry the mutation (Brunner et al. 1991; Reardon et al. 1992*a*).

Division of data according to sex of the transmitting parent is of particular relevance to congenital DM, which is invariably female transmitted. Our data and those of others confirm this pattern of transmission. However, it is possible that the lack of congenital offspring in male transmission may reflect a detrimental effect of larger repeat size on male fertility. In our sample we have only four fathers with a repeat size >2.0 kb. The largest repeat sizes are seen in congenital cases (also see Harley et al. 1992*b*; Tsilfidis et al. 1992), and the mothers of these cases also have larger repeat sizes than do the parents of noncongenital cases (also see Tsilfidis et al. 1992). The overlap in the range of repeat sizes, between congenital cases and others, indicates that size of the expanded sequence is not the only feature determining congenital onset of the disease. The considerable separation of the congenital group from other cases shown in figure 5 emphasizes the relevance of the maternal DNA expansion (and probably phenotype) in the production of this form of DM, perhaps via a direct intrauterine effect related to maternal severity, as suggested previously by Koch et al. (1991). In the present study, DNA expansion was measured only in lymphocytes, and it is possible that different degrees of expansion may cause varying phenotypic effects in other, more relevant tissues.

Analysis of the influence of repeat size on instability suggests that larger repeats are more prone to undergo further expansion. The true effect may be underestimated, since families where a small initial expansion has remained unchanged would be unlikely to be ascertained. Systematic population studies of isolated populations with a high frequency of DM derived from a single common ancestor, such as the population of northern Quebec (Mathieu 1990), may give valuable information on the factors relating to the initiation of instability.

The processes involved in instability of the CTG repeat in DM may be comparable to those occurring in the fragile X syndrome, where a similar underlying se-

quence, a CGG repeat, is responsible (Oberle et al. 1991; Yu et al. 1991) and where an asymptomatic male with a small expansion of the repeat sequence is also found to be the earliest detectable member of most kindreds. Comparison between these two disorders and other conditions will be valuable in elucidating the mechanisms involved in producing disease phenotypes. Strong linkage disequilibrium in both DM (Harley et al. 1991, 1992*a*; Yamagata et al. 1992) and fragile X syndrome (Richards et al. 1992) indicates an origin of the disorder from a very small number of original premutations, possibly a unique event in the case of DM, although the existence of a haplotype predisposing to new mutations could also explain these observations.

It is not possible to relate the clinical or genetic features of DM to the properties of the protein involved, predicted by sequence to be a member of the protein kinase family (Brook et al. 1992; Fu et al. 1992; Mahadevan et al. 1992). The unstable repeat is located in the 3′-UTR of the gene, making it possible that the effects of expansion of the sequence are quantitatively related either to regulation of transcription or translation or to messenger RNA stability, or even making it possible that the functioning of neighboring genes may be affected. Mutations of a different nature elsewhere in the DM gene may produce either clinical features of DM or a different abnormal phenotype. Study of >500 unrelated DM families in a number of different centers has shown that almost all cases result from the same mutational mechanism. The recognition of the unstable repeat sequence and its mode of transmission in families has already begun to clarify many of the unusual clinical and genetic problems in DM.

Acknowledgments

This work was supported by the Muscular Dystrophy Group of Great Britain, The Wellcome Trust, and the Muscular Dystrophy Group of America/Piton Foundation. We would also like to thank Dr. Lodewijk Sandkuijl for help with analysis of data.

References

Aslanidis C, Jansen G, Amemiya C, Shutler G, Mahedevan M, Tsilfidis C, Chen C, et al (1992) Cloning of the essential myotonic dystrophy region and mapping of the putative defect. Nature 255:548–551

Bell J (1947) Dystrophia myotonica and allied diseases. In: Penrose LS (ed) Treasury of human inheritance. Vol 4, part 5. Cambridge University Press, Cambridge, pp 343–410

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter KI, et al (1992) Molecular basis of the 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 HG, Smeets HJM, Nillesen W, Van Oost BA, Van Den Biezenbos JBM, Joosten EMJ, Pinckers AJLG, et al (1991) Myotonic dystrophy: predictive value of normal results on clinical examination. Brain 114:2303–2311

Buxton J, Shelbourne P, Davies J, Jones C, Van Tongeren T, Aslanidis C, de Jong, et al (1992) Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. Nature 335:547–548

Fu Y-H, Pizzuti A, Fenwick RG Jr, King J, Rajnarayan S, Dunne PW, Dubel J, et al (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science 255:1256–1258

Harley HG, Brook JD, Floyd J, Rundle SA, Crow S, Walsh KV, Thibault M-C, et al (1991) Detection of linkage disequilibrium between the myotonic dystrophy locus and a new polymorphic DNA marker. Am J Hum Genet 49:68–75

Harley HG, Brook JD, Rundle SA, Crow S, Reardon W, Buckler AJ, Harper PS, et al (1992*a*) Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. Nature 355:545–546

Harley HG, Rundle SA, Reardon W, Myring J, Crow S, Brook JD, Harper PS, et al (1992*b*) Unstable DNA sequence in myotonic dystrophy. Lancet 339:1125–1128

Harper PS (1989) Myotonic dystrophy. WB Saunders, London

Harper PS, Harley HG, Reardon W, Shaw DJ (1992) Anticipation in myotonic dystrophy: new light on an old problem. Am J Hum Genet 51:10–16

Howeler CJ, Busch HFM, Geraedts JPM, Niermeijer MF, Staal A (1989) Anticipation in myotonic dystrophy: fact or fiction. Brain 112:779–797

Koch MC, Grimm T, Harley HG, Harper PS (1991) Genetic risks for children of women with myotonic dystrophy. Am J Hum Genet 48:1084–1091

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, et al (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. Science 255:1253–1255

Mathieu J (1990) Genealogical reconstruction of myotonic dystrophy in the Saguenay-Lac-Saint-Jean area (Quebec, Canada). Neurology 40:839–842

Oberle I, Rousseau P, Heitz D, Kretz C, Devys D, Hanauer A, Boué J, et al (1991) Instability of a 550 base pair DNA segment and abnormal methylation in fragile X syndrome. Science 252:1097–1102

Penrose LS (1948) The problem of anticipation in pedigrees of dystrophia myotonica. Ann Eugenics 14:125–132

Reardon W, Floyd JL, Myring J, Lazarou LP, Meredith AL, Harper PS (1992*a*) Five years experience of predictive testing for myotonic dystrophy using linked DNA markers. Am J Med Genet 43:1006–1011

Reardon W, Harley HG, Brook JD, Rundle SA, Crow S, Harper PS, Shaw DJ (1992*b*) Minimal expression of myotonic dystrophy: a clinical and molecular analysis. J Med Genet 29:770–773

Richards RA, Holman K, Friend K, Kremer E, Hillen D, Staples A, Brown WT, et al (1992) Evidence of founder chromosomes in fragile X syndrome. Nature Genet 1:257–260

Tsilfidis C, MacKenzie AE, Mettler G, Barcelo J, Korneluk RG (1992) Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. Nature Genet 1:192–195

Yamagata H, Miki T, Ogihara T, Nakagawa M, Higuchi I, Osame M, Shelbourne P, et al (1992) Expansion of unstable DNA region in Japanese myotonic dystrophy patients. Lancet 339:692

Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, Baker E, Holman K, et al (1991) Fragile X genotype characterized by an unstable region of DNA. Science 252:1179–1181