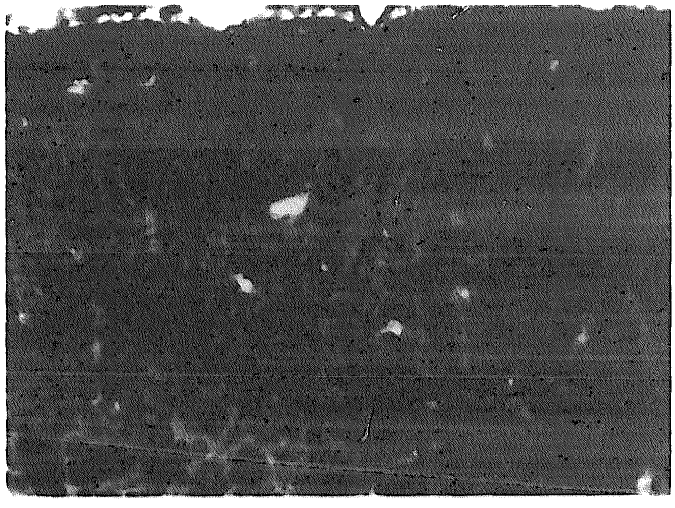
Triplet repeats and human disease

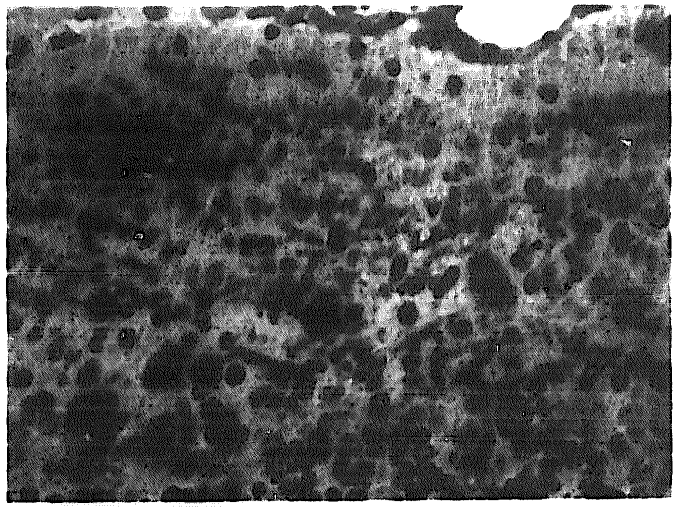
Robert H. Singer

In recent years a new mechanism of genetic disease has been discovered in which triplet DNA sequences (trinucleotides) expand either to interrupt or to compromise a gene. So far, ten genetic loci have been identified in which this event takes place. The presence of these expansions, sometimes containing thousands of repeated trinucleotides, provides a clue as to how processing of DNA and RNA may sometimes go awry. How this expansion in a single allele may exert a dominant effect and how the extent of the expansion increases the severity of the disease remain a mystery.



Purkinje cells of normal (top) and human SCA1 transgenic (bottom) mice. Reproduced, with permission, from Ref. 36

Robert H. Singer
Department of Cell Biology,
University of Massachusetts
Medical School, 55 Lake Avenue North, Worcester,
MA 01655, USA.
Tel: +1 508 856 2429
Fax: +1 508 856 5612
e-mail: rhsinger@insitu.ummed.edu



REPEATS of the nucleosides adenosine (A), cytidine (C), guanosine (G) and thymidine (T) occur throughout the DNA of the human genome. Simple tandem repeats of up to 1000 units of 2–4 nucleotides are known as microsatellites[1]. The triplet repeat, a type of microsatellite, has been found in every animal species studied, including humans, in which stretches of five or more triplet repeats have been detected in at least 50 genes. Some of these three-base-pair repeated DNA sequences are unstable and amplify to a higher than normal number of copies. In the past four years, these expanding triplet repeats have been shown to be a new type of mutation responsible for at least eight hereditary disorders[2–7], including fragile X syndrome, spinobulbar muscular atrophy, myotonic dystrophy, Huntington's disease, dentatorubral pallidolusyian atrophy, Machado–Joseph disease and spinocerebellar ataxias (see Box 1).

These anomalous mutations do not conform to the classic mendelian rules of genetic inheritance. After transmission from one generation to the next, the triplet repeat diseases tend to be unpredictable, but in at least one case, myotonic dystrophy, it can have earlier onset and be more severe, a phenomenon referred to as 'anticipation'. While expanded repeats are sometimes shorter in the offspring of affected individuals, the trend is for these repeats to amplify from one generation to the next. Remarkably, the severity of the disease is directly proportional to the repeat length. This can result in a severe disorder appearing in a family that had been disease free, or nearly so, for generations until the number of repeats reached a threshold for a 'disease phenotype' to be expressed in the patient inheriting the affected chromosome. Incomplete penetrance may also occur (i.e. not everyone who inherits the disease gene is symptomatic). The wide range of symptoms in those with triplet repeat diseases may be explained by the variable numbers of repeats in an individual[2–7].

Understanding why these nucleotide sequences in humans are unstable and inclined to expand, or occasionally contract, would explain not only how these hereditary disorders arise, but also how other sequences in the genome are stably transmitted to future generations. Perhaps the presence of triplet repeats suggests a function in the genome. As several of the triplet repeat diseases are neurodegenerative, further studies should shed light on the mechanisms involved in neuronal cell degeneration.

Mechanisms of triplet expansion

One model for triplet expansion is that the repeats distort the structure of the DNA, forming 'hairpin' structures[21]. These may cause deletions in the lagging strand synthesis or expansions in the leading

©1996, Elsevier Science Ltd 1357 - 4310/96/$15.00

strand synthesis (Fig. 1)[22] during replication. The DNA polymerase then begins to 'slip' or 'stutter' on the repeat[2-7], causing continual expansions. This may occur either during the meiotic divisions to produce eggs or sperm, or in the rapidly dividing somatic cells of the early embryo.

The normal range of trinucleotide repeats is usually less than 50. In normal individuals, the trinucleotide repeat number in a specific gene may vary within this normal range[2-7] and the number of repeats is stably transmitted to future generations. In individuals affected by, or carriers of, triplet repeat diseases, the repeat numbers are unstable and expand beyond this normal range.

For instance, in fragile X syndrome[2-7], an overlap between the high end of the range of repeats in normal individuals (6–52 repeats) and the low end of the range of repeats (50–200 repeats) in carriers of pre-mutation alleles exists. Affected individuals carry at least 600 repeats. Therefore, the length of the repeat section alone correlates with the repeat-instability in this syndrome, but by itself is probably not sufficient to cause disease. Two AGG trinucleotides interrupt the CGG repeats in most *FMR1* alleles in fragile X. When one or both of these AGGs is lost, instability characteristic of the pre-mutation generally occurs. This is usually in the 3' tract relative to transcription, resulting in 39 or more uninterrupted CGGs. When there are 33 or fewer uninterrupted CGGs, the repeats are stably inherited[14].

A model has been proposed to explain the generation[23] of unstable *FMR1* CGG repeat alleles by the loss of AGG triplets. Preferential AGG loss in the 3' region may result from the differential potential for the leading and lagging strands to mutate during replication. Okazaki fragments are more likely to form slipped structures and therefore to mutate. Pure CGG repeats averaging 70 copies in pre-mutation alleles (range, 56–74 copies) are approximately the same length as a typical Okazaki fragment (150–200 base pairs). Okazaki fragment synthesis during DNA replication would lead to a fragment beginning and ending with pure CGG repeats, having no single point of reference on the DNA, and thus being susceptible to formation of slipped structures. Failure of mismatch repair enzymes to repair the resultant structures could lead to the exponential increases in repeat lengths that characterize the fragile X full mutation (Fig. 1)[22].

Recently, it has been proposed that two mechanisms may account for the triplet repeat expansion: (1) a small slippage that gradually leads to longer expansions followed by (2) hairpin formation and a rapid increase to large-scale expansions[22,24].

Box 1. Human diseases associated with triplet repeats

There are seven human diseases known to be associated with triplet repeats; most are neurodegenerative.

Myotonic dystrophy (Steinert's disease)

This is the most common form of adult muscular dystrophy, occurring in 1 in 8000 individuals. Symptoms include progressive muscle weakness, myotonia, cataracts, cardiac arrhythmia and diabetes. The myotonin kinase gene (*DMPK*), located at chromosome locus 19q13.3 DM, is expressed in muscle and codes for a cyclic-AMP-dependent protein kinase. Normally, the trinucleotide CTG is repeated 5–7 times in the 3' untranslated region (UTR) of this gene[8,9], but 50–2000 copies occur in individuals with myotonic dystrophy. Higher copy numbers are associated with more severe disease and an earlier onset of the symptoms[8-10]. Occasionally, a reduction in the size of the repeat is seen[11].

Fragile X syndrome

This is the most common form of inherited mental retardation and occurs in 1 in 1500 males and 1 in 2500 females. Typical clinical signs include a reduced IQ, elongated facial features with large everted ears, and macro-orchidism (enlarged testicles). It is an X-linked, dominantly inherited disorder with unusual penetrance: 30% of carrier females are mentally retarded, while 20% of males inheriting the fragile X chromosome are phenotypically normal. *FMR1*, located at chromosome locus Xq27.3, is the gene involved in fragile X syndrome[1,12,13]. It is expressed in the brain and normally carries 5–52 copies of the trinucleotide CGG. Repeats occur in the 5' UTR of the first exon of the *FMR1* gene[14]. In mutations where 5–200 copies occur, known as pre-mutations, carriers of the gene are healthy. Children who inherit the gene from these carriers and develop fragile X syndrome have 200 to several thousand CGG repeats (full mutations)[15].

Huntington's disease

This is a progressive neurodegenerative disorder characterized by an insidious adult onset of chorea, dementia and personality changes. The prevalence of this autosomal dominant trait[16] may be as high as 1 in 10 000 individuals in some populations. Six to 37 CAG repeats normally occur in the protein-coding region of the *huntingtin* gene, which maps to chromosome locus 4p16.3 HD. In Huntington's disease, 35–121 repeats occur, with lower age of onset associated with higher repeat copy numbers.

Spinal and bulbar muscular atrophy (Kennedy's disease)

This is a rare, X-linked recessive motor neuron disorder with an adult onset of proximal and progressive muscle weakness and atrophy of bulbar muscles. It affects 1 in 50 000 males, resulting in gynecomastia (breast development) and reduced fertility. The spinal and bulbar muscular atrophy (SBMA) mutation (Ref. 17) occurs in the androgen receptor (*AR*) gene on chromosome locus Xq11-12 AR. CAG repeats normally number 12–34 in the first exon of the protein-coding region of the *AR* gene. Multiplication of these repeats to 40–62 copies results in the SBMA mutation.

Spinocerebellar ataxias

Three forms of autosomal dominant neurodegenerative diseases, spinocerebellar ataxia type 1 (SCA1)[18], Machado–Joseph disease (MJD)[19] and dentato-rubral pallidolusyian atrophy (DRPLA)[20], result in progressive degeneration of the cerebellum and posterior columns of the spinal cord, with varying signs of central and peripheral nervous system impairment, as a result of expansion of the trinucleotide CAG. The genes involved are on different chromosomes in each of these syndromes (6p22-23, 14q32.1 and 12p12-13), and are of unknown function. Increased severity and earlier age of onset are associated with higher numbers of repeats. DRPLA is very rare outside of Japan, where it has an incidence of 1 in 10^6 individuals. Haw River syndrome (HRS), a dominant neurodegenerative disease affecting five generations of an African–American family in the USA, has recently[20] been shown to be caused by an expanded repeat in the same allele as occurs in DRPLA. However, the clinical and neuropathological features of HRS and DRPLA differ, and the reason for this is currently unknown.

Gender source of repeat expansions

The sex of the transmitting parent also affects triplet repeat expansions. In fragile X syndrome, expansion of the pre-mutation to full mutation occurs only on the chromosome transmitted by the mother. In Huntington's disease, expansion occurs more often when the gene is transmitted by the father, and a juvenile onset of the disease is more likely to occur when the trait is inherited from the father[7].

Mechanisms of pathogenesis

In classically mendelian inherited disorders, a loss of function of the gene product usually results from inheritance of two recessive alleles. The inheritance of a mutation leading to gain of function of the resultant protein generally results from one dominant allele that causes a dominant-negative disorder. Recessive disorders can result from any mutation in DNA, such as deletions or point mutations, which result in altered transcription, translation or protein function. Dominant disorders usually result from point mutations: the gene product is still produced, but in an altered form[2–7]. When the loss of one allele results in a disease phenotype arising from half of the normal amount of protein being made, this is known as haplo-insufficiency. The triplet repeat disorders, however, do not neatly fit this classical model.

Fragile X

The X-linked recessive fragile X site contains a CpG island that is methylated in inactive normal and inactive pre-mutation X chromosomes, but not in active X chromosomes. In male pre-mutation carriers, repeat lengths are transcribed into *FMR1* mRNA that has a normal steady-state level and half-life, and translated into a protein with a similar concentration to that in normal cells[15]. Methylation occurs in the promoter region in the chromosome, with full mutation correlating with loss of *FMR1* gene expression and therefore loss of function.

Spinobulbar muscular atrophy

In spinobulbar muscular atrophy (SBMA), complete loss of function of the androgen receptor (AR) does not occur. Poly(CAG) in the coding region leads to the synthesis of polyglutamine in the AR protein, and presumably a gain of function of the protein in males[25]. It is possible that females are protected from this toxic gain-of-function disorder by low but protective androgen levels.

Other trinucleotide repeat diseases

Other polyglutamine diseases lead to neurodegenerative disorders. Poly(CAG) expansion also occurs in SCA1, DRPLA and Huntington's disease, resulting in expanded polyglutamine tracts of considerable length in the gene products. What functions are altered in the resultant proteins remain to be determined.

Myotonic dystrophy

In myotonic dystrophy, controversy exists over the effect of triplet repeats on *DMPK*, the gene encoding myotonin kinase[26–30]. Since the triplet repeats occur in the 3' untranslated portion of the affected gene, the variability of clinical presentation and the dominant mode of inheritance must be reconciled with the fact that the protein product is not expected to be altered.

Some studies report that expression of mRNA is lost[26] or decreased[27] in adult patients with late-onset disease, and others report that mRNA production is increased in early-onset congenital dystrophy patients[28]. The repeat has been shown to cause hyperstable nucleosome formation and positioning, suggesting a mechanism in which transcription is inhibited, DNA polymerases are stalled and DNA templates slip[29]. In one study, hyperstabilization of mRNA is more consistent with the dominant

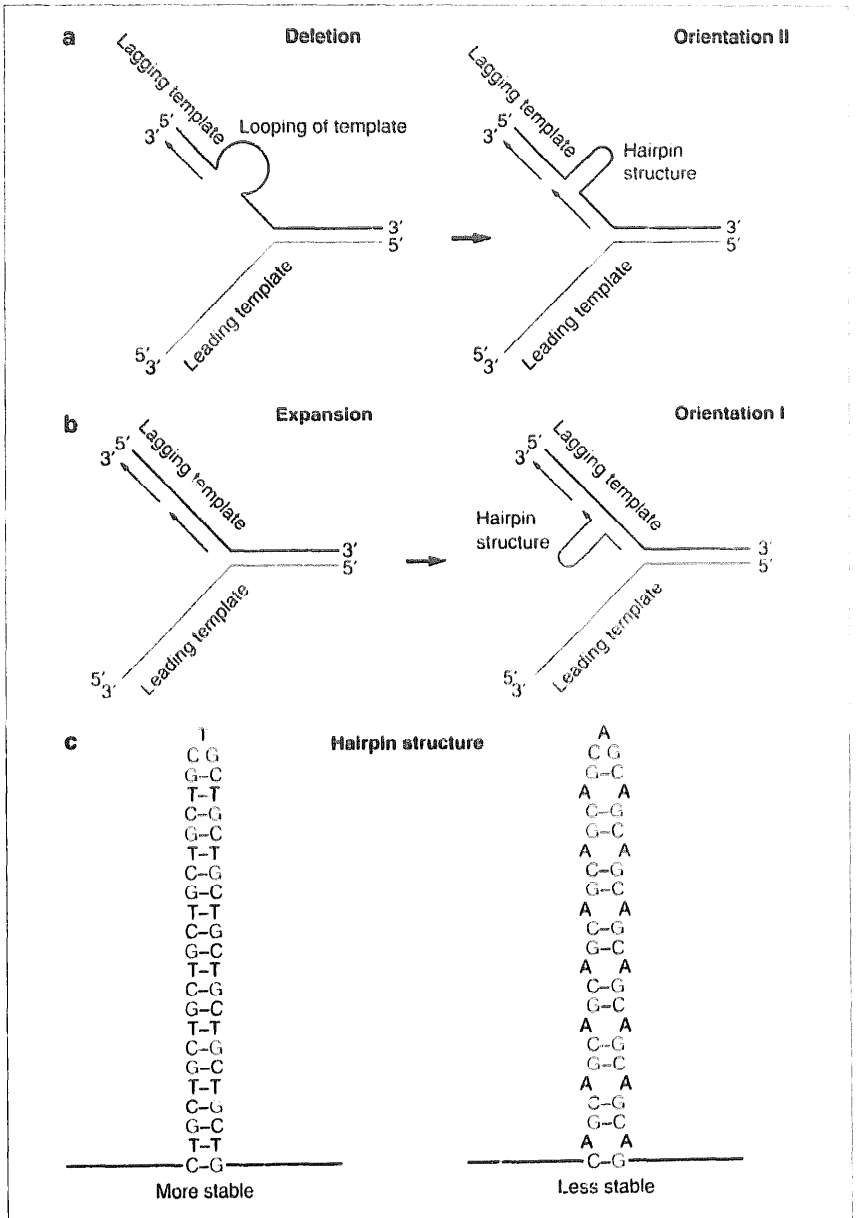


Figure 1. Model for orientation-dependent instability of triplet repeats during replication. (a) Deletion: looping out of the template on the lagging strand leads to formation of a hairpin structure that is bypassed by DNA polymerase. (b) Expansion: looping out of the newly synthesized DNA on the lagging strand causes repeated replication of the repeat-containing region. (c) Possible hairpin structures adopted by CTG and CAG single-stranded DNA. Reproduced, with permission, from Ref. 22.

Glossary

Allele – One of two (diploid) copies of a gene in the genome. An allele before it becomes mutated is called a pre-mutation allele.

Amniocytes and chorionic villus – Fetal cells that can be tested for genetic defects.

Congenital dystrophy – Myotonic dystrophy in which the triplet-repeat expansion is inherited from a parent rather than arising spontaneously in the individual.

CpG island – Regions of the genome involved in transcription initiation that are unusually rich in the nucleosides C and G and are surrounded by regions with a reduced C and G content.

Dominant-negative disease – A disease caused by a single mutant copy of a gene, the other copy being normal. If a new protein is produced, it may have functional consequences ('gain of function').

Export artifact – Nucleic acids that cannot be exported from the nucleus into the cytoplasm.

***FMR1* mRNA** – The mRNA transcribed from the fragile X gene.

Nucleosome – A nuclear structure that organizes the DNA.

Okazaki fragments – Short lengths of DNA formed on the lagging strand during DNA replication. They are rapidly joined to form a continuous DNA strand; the leading strand is synthesized continuously.

Triplet (or trinucleotide) repeat expansion – Situation where three nucleotides within a gene are repeated a number of times. This may result in a 'disease phenotype'.

mode of inheritance[28], but further studies into the function of *DMPK* will indicate whether a true gain of function results from the mutation[25].

A recent observation (D.H. Housman, pers. commun.) suggests that the phenotype in myotonic dystrophy may arise from deficiencies in the protein. Homozygous deletions ('double knock-out') of the *DMPK* gene in mice resulted in late onset disease with defects in muscle physiology and morphology. A heterozygous deletion had a more mild phenotype. This raises the possibility that the disease results from haplo-insufficiency. Perhaps this deficiency in the protein kinase also contributes to a dominant phenotype if the continued lack of the enzyme has a cumulative effect on the progression of the disease. Another possibility is that another gene is affected downstream of *DMPK* (reviewed in Ref. 30).

Do triplet repeat diseases result from RNA processing defects?

The intracellular location of transcripts from the *DMPK* gene has been analyzed by *in situ* hybridization in fibroblasts and muscle biopsies of patients with myotonic dystrophy and normal individuals[31]. In myotonic dystrophy patients, but not in normal individuals, post-transcriptional RNA containing expanded CTG repeats was detected as foci that appear to build up in the nuclei (Fig. 2). This nuclear concentration of transcripts may represent aberrant processing of the RNA in myotonic dystrophy. Recently we have found a similar effect in fragile X syndrome[32]. One possibility is that the polyadenylation of the transcript is disrupted, leading to an export artifact[33]. Accumulation and inappropriate processing of transcripts might in turn affect the processing or export of other mRNAs in the cell[34].

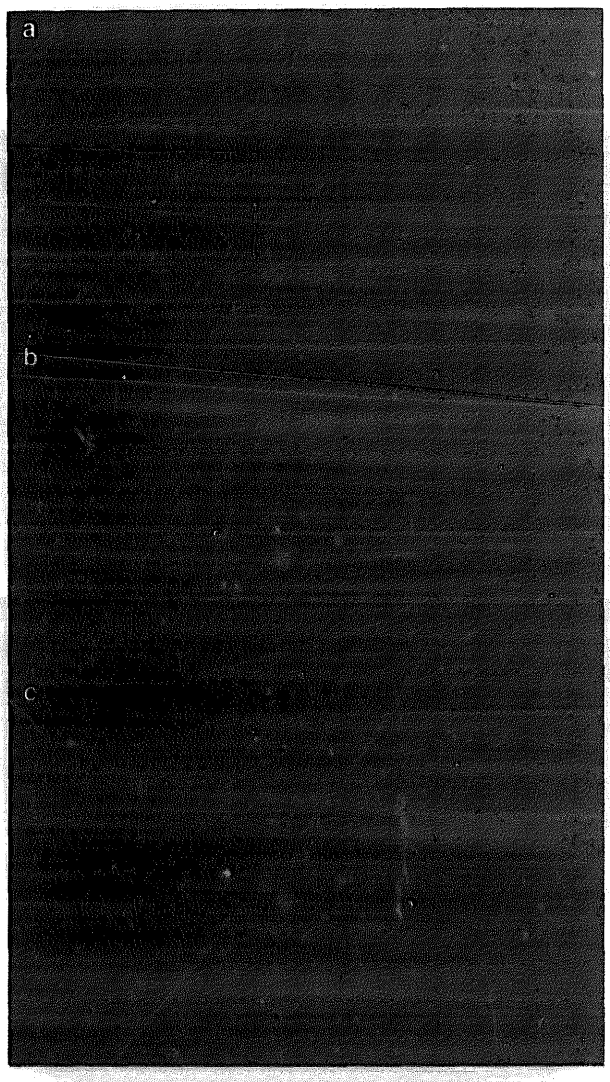


Figure 2. Detection of the expanded *DMPK* gene and its transcript by *in situ* hybridization. Fibroblasts from a myotonic dystrophy patient were hybridized simultaneously with (**a**) a sense probe [$(CTG)_{10}$] conjugated with fluorescein (green) and (**b**) an antisense probe [$(CAG)_{10}$] conjugated to Cy3 (red). Post-transcriptional RNA (red) can be distinguished from the site of the gene (green) and (**c**) its nascent transcripts (white, resulting from the superimposition of the red and green signals). DNA was stained by 4',6'-diaminophenylindole (DAPI, blue). The nuclear concentration of transcripts may represent aberrant processing of the RNA in myotonic dystrophy. See also Ref. 31.

Therapeutics

Current strategies for 'correction' of genetic disorders have focused on recessive conditions where there is a loss of function. Intervention in the triplet repeat diseases, where there may be a gain of function, presents a significant challenge. At present, the therapies for triplet repeat diseases, where they exist, are aimed at palliation of symptoms.

Prospects for the future

Since large-scale expansions of triplet repeats have been found only in humans, and not in other species, there are no known animal models of these conditions. The instability of repeats, especially in fragile X syndrome and myotonic dystrophy, has resulted in the loss of these

The outstanding questions

- What is the mechanism of triplet repeat expansion?
- How do the repeat expansions explain 'anticipation' of disease in the next generation?
- By what mechanism do the repeats result in a dominant disease phenotype?
- How many more human diseases arise from triplet repeats?
- Why is the same expanded repeat seen in dentato-rubral pallidolusyian atrophy and Haw River syndrome although the clinical and neuropathological features differ?
- How do the repeats interact with the myotonic dystrophy gene to result in disease?

sequences during attempts to clone the genes containing them. Because the repeats in spinobulbar muscular atrophy are shorter, it has been possible to clone the mutant androgen receptor gene and insert it into transgenic mice. The gene has been stably inherited[35]. No 'disease' phenotype was observed, however, possibly owing to insufficient expression of the transgene. By contrast, the human *SCA1* gene (82 repeats), which encodes the ataxia-1 protein, was expressed in much greater amounts in transgenic animal models and resulted in a neurological and histological phenotype typical of human ataxia, for example, a loss of Purkinje cells[36]. The development of this animal model should result in more rapid progress on the elucidation of the mechanism of disease pathogenesis.

The detection of characteristic mRNA foci in nuclei from the fibroblasts of individuals with myotonic dystrophy, but not from normal individuals (Fig. 2)[34], suggests that a convenient and rapid diagnostic test could be developed using cells derived from skin biopsies, or in the prenatal period from amniocytes or chorionic villus cells. Since the amount of signal in the test resulting from *in situ* hybridization is proportional to the number of triplet repeats, it could also be used, in theory, to predict the severity of disease. However, such diagnostic tests will necessitate the development of instrumentation for imaging and sample preparation.

Acknowledgements. The author thanks Lynne Lederman for help with the manuscript, David Housman for his useful advice and Krishan Taneja for his research contributions. Supported by NIH HD18066 to RHS.

References

1 Plassart, E. and Fontaine, B. (1994) **Genes with triplet repeats: a new class of mutations causing neurological diseases.** *Biomed. Pharmacother.* 48, 191–197

2 Caskey, C.T. *et al.* (1992) **Triplet repeat mutations in human disease.** *Science* 256, 784–789

3 Richards, R.I. and Sutherland, G.R. (1992) **Dynamic mutations, a new class of mutations causing human disease.** *Cell* 70, 709–712

4 Warren, S.T. and Nelson, D.L. (1993) **Trinucleotide repeat expansions in neurological disease.** *Curr. Opin. Neurobiol.* 3, 752–759

5 Hoffman, E.P. and Jaffurs, D. (1993) **An expanding enigma.** *Curr. Biol.* 3, 456–459

6 Morell, V. (1993) **The puzzle of the triple repeats.** *Science* 260, 1422–1423

7 Bates, G. and Lehrach, H. (1994) **Trinucleotide repeat expansions and human genetic disease.** *BioEssays* 16, 277–284

8 Brook, J.D. *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

9 Mahadevan, M. *et al.* (1992) **Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene.** *Science* 255, 1253–1255

10 Fu, Y-H. *et al.* (1992) **An unstable triplet repeat in a gene related to myotonic muscular dystrophy.** *Science* 255, 1256–1258

11 O'Hoy, K.L. *et al.* (1993) **Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission.** *Science* 259, 809–812

12 Verkerk, A.J.M.H. *et al.* (1991) **Identification of a gene (*FMR1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome.** *Cell* 65, 905–914

13 Oberlé, I. *et al.* (1991) **Instability of a 55-base pair DNA segment and abnormal methylation in fragile X syndrome.** *Science* 252, 1097–1102

14 Eichler, E.E. *et al.* (1994) **Length of uninterrupted CGG repeats determines instability in the *FMR1* gene.** *Nat. Genet.* 8, 88–94

15 Feng, Y., Lakkis, L., Devys, D. and Warren, S.T. (1995) **Quantitative comparison of *FMR1* gene expression in normal and pre-mutation alleles.** *Am. J. Hum. Genet.* 56, 106–113

16 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, 971–983

17 LaSpada, A.R. *et al.* (1992) **Meiotic stability and genotype–phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy.** *Nat. Genet.* 2, 301–304

18 Orr, H.T. *et al.* (1993) **Expansion of an unstable trinucleotide (CAG) repeat in spinocerebellar ataxia type 1.** *Nat. Genet.* 4, 221–226

19 Maciel, P. *et al.* (1995) **Correlation between CAG repeat length and clinical features in Machado–Joseph disease.** *Am. J. Hum. Genet.* 57, 54–61

20 Burke, J.R. *et al.* (1994) **The Haw River Syndrome: dentatorubropallidoluysian atrophy (DRPLA) in an African–American family.** *Nat. Genet.* 7, 521–524

21 Mitas, M. *et al.* (1995) **Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: $(CTG)_{15}$.** *Nucleic Acids Res.* 23, 1050–1059

22 Kang, S., Jaworski, A., Ohshima, K. and 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, 213–218

23 Kunkel, T.A. (1993) **Slippery DNA and diseases.** *Nature* 365, 207–208

24 Gacy, A.M. *et al.* (1995) **Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*.** *Cell* 81, 533–540

25 Housman, D.H. (1995) **Gain of glutamines, gain of function?** *Nat. Genet.* 10, 3–4

26 Carango, P., Noble, J.E., Marks, H.G. and Funange, V.L. (1993) **Absence of myotonic dystrophy protein kinase (DMPK) mRNA as a result of a triplet repeat expansion in myotonic dystrophy.** *Genomics* 18, 340–348

27 Fu, Y-H. *et al.* (1993) **Decreased expression of myotonin-protein kinase messenger RNA and protein in adult form of myotonic dystrophy.** *Science* 260, 235–238

28 Sabouri, L.A. *et al.* (1993) **Effect of the myotonic dystrophy (DM) mutation of mRNA levels of the *DM* gene.** *Nat. Genet.* 4, 233–238

29 Wang, Y.H. *et al.* (1994) **Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene.** *Science* 265, 669–671

30 Nowak, R. (1995) **Genome mappers have a hot time at Cold Spring Harbor.** *Science* 268, 1134–1135

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

32 Taneja, K.L., Nelson, D. and Singer, R.H. (1995) **Spatial distribution of CGG repeat sequences in nuclei of fragile X patient cells.** *Mol. Biol. Cell* (Suppl.) 6, 310a

33 Wang, J. *et al.* (1995) **Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation.** *Hum. Mol. Genet.* 4, 599–606

34 Nowak, R. (1995) **Making sense of myotonic dystrophy.** *Science* 270, 1120

35 Bingham, P.M. *et al.* (1995) **Stability of an expanded trinucleotide repeat in the androgen receptor gene in transgenic mice.** *Nat. Genet.* 9, 191–196

36 Burright, E.N. *et al.* (1995) **SCA1 transgenic mice: a model for neurodegeneration caused by an expanded CAG trinucleotide repeat.** *Cell* 82, 937–948