ORIGINAL ARTICLES

Molecular Neurogenetics Unit,
Massachusetts General Hospital,
Charlestown, Massachusetts
02129, USA.
M E MacDonald
G Barnes
J Srinidhi
M P Duyao
C M Ambrose
H MacFarlane
M A Anderson
B Jenkins
J F Gusella

Department of Neurology,
Boston University Medical
School, Boston, Massachusetts
02118, USA.
R H Myers

Department of Medical
Genetics, Indiana University
Medical Center, Indianapolis,
Indiana 46202, USA.
J Gray
P M Conneally

Department of Neurology,
Massachusetts General Hospital,
Boston, Massachusetts 02114,
USA.
A Young
J Penney
Z Hollingsworth
W Koroshetz
J-H Cha
L Dure

Department of Neurology,
University of Rochester Medical
Center, Rochester, New York
14642, USA.
I Shoulson

Brain Tissue Resource Center,
McLean Hospital and
Department of Neuropathology,
Harvard Medical School,
Belmont, Massachusetts 02178,
USA.
E Bird
J P Vonsattel

Laboratory for Molecular
Neuropathology, Massachusetts
General Hospital Neuroscience
Center, Charlestown,
Massachusetts 02129, USA.
J P Vonsattel

Universidad de Zulia,
Maracaibo, Venezuela.
E Bonilla
M de Young

Departments of Neurology and
Psychiatry, College of
Physicians and Surgeons,
Columbia University, New York
NY 10032, and Hereditary
Disease Foundation, 1427 7th St,
Suite 2, Santa Monica,
California 90401, USA.
C Moscowitz
G Penchaszadeh
L Brzustowicz
N S Wexler

Long Island Jewish Hillside
Medical Center, Glen Oaks, New
York 11004, USA.
J Alvir

Baylor University College of
Medicine, Houston, Texas 77009,
USA.
J Bickham Conde

Aguirre 621-7A, Capital 1414,
Buenos Aires, Argentina.
F Gomez

Servicio de Genética, Hospital
Virgen del Camino, 31008
Pamplona, Spain.
M Ramos-Arroyo

Department of Neurology,
University of Miami, Miami,
Florida 33136, USA.
J Sanchez-Ramos

Pediatric Neurology, University
of Mississippi Medical Center,
Jackson, Mississippi 39216, USA.
S R Snodgrass

Department of Genetics,
Harvard Medical School,
Boston, Massachusetts 02114,
USA.
J F Gusella

Correspondence to
Dr MacDonald.

Received 18 August 1993.
Revised version accepted
23 September 1993.

Gametic but not somatic instability of CAG repeat length in Huntington's disease

M E MacDonald, G Barnes, J Srinidhi, M P Duyao, C M Ambrose, R H Myers, J Gray, P M Conneally, A Young, J Penney, I Shoulson, Z Hollingsworth, W Koroshetz, E Bird, J P Vonsattel, E Bonilla, C Moscowitz, G Penchaszadeh, L Brzustowicz, J Alvir, J Bickham Conde, J-H Cha, L Dure, F Gomez, M Ramos-Arroyo, J Sanchez-Ramos, S R Snodgrass, M de Young, N S Wexler, H MacFarlane, M A Anderson, B Jenkins, J F Gusella

Abstract

Instability of a CAG repeat in 4p16.3 has been found in Huntington's disease (HD) chromosomes. Unlike a similar repeat in the fragile X syndrome, the expanded HD repeat showed no evidence of somatic instability in a comparison of blood, lymphoblast, and brain DNA from the same persons. Four pairs of monozygotic HD twins displayed identical CAG repeat lengths suggesting that repeat size is determined in gametogenesis. In contrast with the fragile X syndrome and with HD somatic tissue, mosaicism was readily detected as a diffuse spread of repeat lengths in DNA from HD sperm samples. Typically, the modal repeat size was larger in the sperm DNA than in corresponding lymphoblast DNA, with the greatest degree of gametic mosaicism coinciding with the longest somatic CAG repeats. These data indicate that the developmental timing of repeat instability appears to differ between HD and fragile X syndrome, and that the fundamental mechanisms leading to repeat expansion may therefore be distinct.

(*J Med Genet* 1993;30:982–6)

Huntington's disease (HD) is a progressive neurodegenerative disorder of mid life onset that is inherited in an autosomal dominant fashion.[1] The cardinal symptoms of HD, choreic movements, psychiatric changes, and intellectual decline, result from the loss of neurones, primarily in the striatum. The genetic basis for HD has recently been identified as an expanded, unstable CAG trinucleotide in a new 4p16.3 gene.[2–5] This mechanism of mutation has been described in the fragile X syndrome, myotonic dystrophy, spinobulbar muscular atrophy, and spinocerebellar ataxia 1.[6–21] In the fragile X syndrome, a premutation allele of 52 to 200 repeat units gives rise to the disease allele by explosive expansion to 200 to 2000 repeat units.[22] The observation of extensive somatic variation of repeat length in fragile X syndrome in affected subjects and in some premutation carriers, combined with the

absence of the full disease allele from sperm of affected males, has led to the suggestion that repeat expansion may occur during a window early in fetal development.[23 24]

In HD, affected subjects possess 37 to 86 copies of the CAG repeat.[3] Although the repeat is very unstable, with size changes occurring in more than 80% of transmissions, no explosive increases comparable to those observed in fragile X syndrome have been reported. To determine whether there is evidence of repeat instability occurring during development, we have scrutinised the HD repeat length in a variety of tissues, and matched cell lines, in four pairs of affected monozygotic twins, and in sperm from HD gene carriers. The results indicate that HD repeat length instability does not occur somatically, but in the germline, and may therefore be fundamentally different from the mechanism underlying fragile X syndrome.

Methods and subjects

All subjects typed for the HD CAG repeat were diagnosed or presymptomatic members of HD families drawn from the New England Huntington's Disease Center Without Walls, the Brain Tissue Resource Center of the McLean Hospital (Brain Bank), the HD Research Roster at Indiana University, or the US-Venezuela Collaborative HD project. Four pairs of twins were ascertained and typed with five highly polymorphic DNA loci (GABRB1, D13S120, D21S210, SPN, TOP1P2) to confirm monozygosity. The likelihood of monozygosity was computed by standard procedures[25] using known allele frequencies (Human Genome Database) and the estimated frequency of mating of different parental genotypes. The estimated likelihood of monozygosity exceeded 0·99999 for all four sets of twins. It was not possible to ascertain the timing of the twinning event in these subjects from medical history.

The DNA tested for all subjects was prepared from lymphoblastoid cell lines, fresh blood, frozen postmortem brain or peripheral tissue samples, or pelleted sperm.[3] The length

of the HD gene CAG repeat for all DNA was assessed using polymerase chain reaction (PCR) amplification, as described previously.[2] Autoradiograms of sperm DNA PCR products were scanned with an LKB UltroScan XL laser densitometer to calculate relative area under each absorbance curve, defining the range and mode of the repeat sizes observed relative to PCR products from the corresponding lymphoblast DNA run in adjacent lanes, and to sized products from sequenced cosmid DNA.

Results

ABSENCE OF SOMATIC VARIATION IN REPEAT LENGTH

The somatic variation in triplet repeat length in fragile X syndrome was first noted as an extended smear of expanded alleles in blots of DNA extracted directly from blood. Typing of blood DNA from more than 100 HD cases has not shown multiple or extensively smeared PCR products from the disease chromosome. Instead, the expanded allele yields a relatively discrete product visualised as a set of three to four closely spaced bands. This minimal heterogeneity of the PCR product is probably a consequence of the PCR assay itself and varies

slightly with DNA quality and precise reaction conditions. Certainly, this 'chatter' is comparable to that obtained with the normal allele, and with most alleles of simple sequence repeat markers. These observations indicate that somatic variation of the HD triplet, if it occurs at all, is much more limited than that of the expanded fragile X repeat. To explore the issue of somatic mosaicism in greater detail, we have examined HD DNA from a number of sources.

We first compared HD triplet repeat length in paired DNA samples from the same subject, one sample extracted from fresh blood, and the other prepared from a lymphoblastoid cell line established years earlier. Fig 1 shows typical results for samples separated by seven to ten years. In these examples, each subject possessed an HD allele, in the range of 38 to 49 units, and a normal allele, in the 18 to 21 unit range. The size of the PCR product in these four and 11 other cases (not shown) appeared to be identical within a pair. Thus, despite the shifting polyclonal nature of the blood's nucleated cells, no discernible change in repeat length occurred either on the HD or normal chromosomes.

We also compared DNA from various regions of HD brain with lymphoblast or blood DNA from the same person. In 19 independent cases, no variation in HD triplet repeat length was observed. Fig 2 shows typical results comparing blood or lymphoblast DNA with DNA from the striatum, the region most severely affected in HD, and selected other brain regions. For all 19 HD cases, repeat lengths of 42 to 54 units on the disease chromosome and 12 to 27 units on the normal chromosome were all invariant within a person, regardless of the tissue source of the genomic DNA. In a few instances, additional tissues were available, including in some cases spleen, liver, kidney, and skeletal muscle. In no case was any somatic variation of the HD triplet repeat detected.
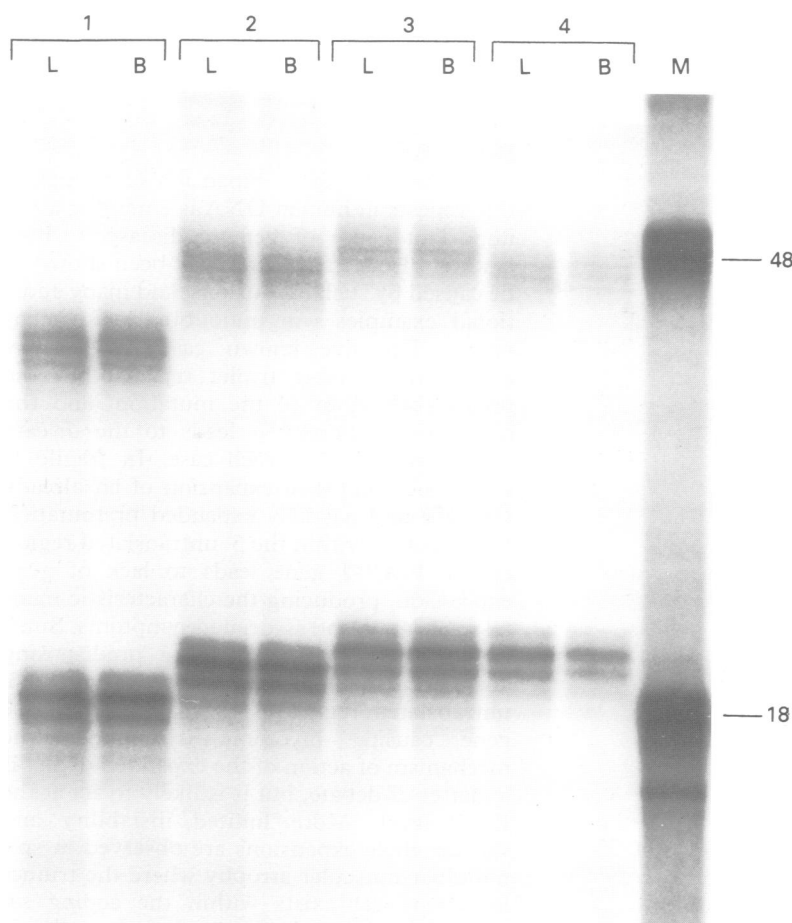

Figure 1   Comparison of repeat length in paired HD lymphoblast and blood DNA. The HD CAG repeat was PCR amplified from paired DNA samples of four subjects. L = lymphoblast DNA, B = blood DNA. The lymphoblast lines were established from independent blood samples collected 8, 7, 10, and 8 years earlier than the blood samples used for subjects 1 to 4, respectively. Lane M contains PCR products from two cosmids known to contain 48 repeat units and 18 repeat units, respectively. Repeat unit lengths of the expanded and normal alleles were estimated as: subject 1: 38 and 18; subject 2: 47 and 20; subject 3: 48 and 21; subject 4: 47 and 21.

IDENTICAL HD REPEAT LENGTHS IN MONOZYGOTIC TWINS

An alternative approach to exploring the stage at which repeat instability occurs is to examine pairs of monozygotic twins. We have analysed four such pairs affected with HD. The twins in each of pairs 1 to 4 displayed onset of the disorder within three years of each other, at ages 18, 26 to 29, 33, and 30, respectively. Fig 3 shows the results of PCR amplification of the HD triplet repeat from cerebellum, blood, or lymphoblast DNA of three of these pairs. Comparison of the HD repeat showed that within each of the four pairs, the twins displayed identical PCR products of 54, 50, 48, and 49 units, respectively. Pairs 1 and 4 inherited the HD allele from their father while pairs 2 and 3 received the disease gene from their mother. We have previously shown that more than 80% of HD transmissions from either sex involve a measurable size change in the HD repeat length.[3] Therefore, the uniformity of the repeat size within each twin pair suggests
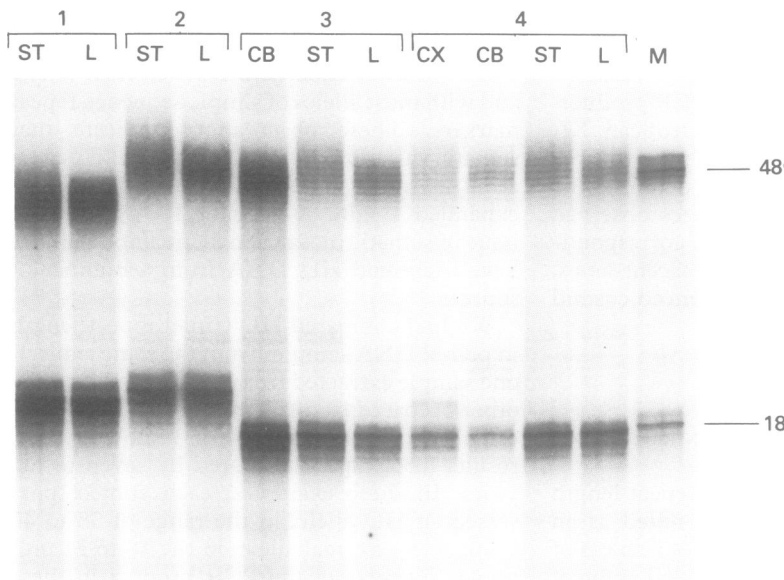
Figure 2 Comparison of repeat length in HD brain and lymphoblast DNA. The HD CAG repeat was PCR amplified from matched DNA samples of four subjects. L = lymphoblast, ST = striatum, CB = cerebellum, CX = frontal cortex. The brain DNA samples were obtained from frozen postmortem tissue with confirmed HD neuropathology. Lane M contains PCR products from two cosmids known to contain 48 repeat units and 18 repeat units, respectively. Repeat unit lengths of the expanded and normal alleles were estimated as: subject 1: 44 and 20; subject 2: 50 and 21; subject 3: 48 and 17; subject 4: 48 and 17.
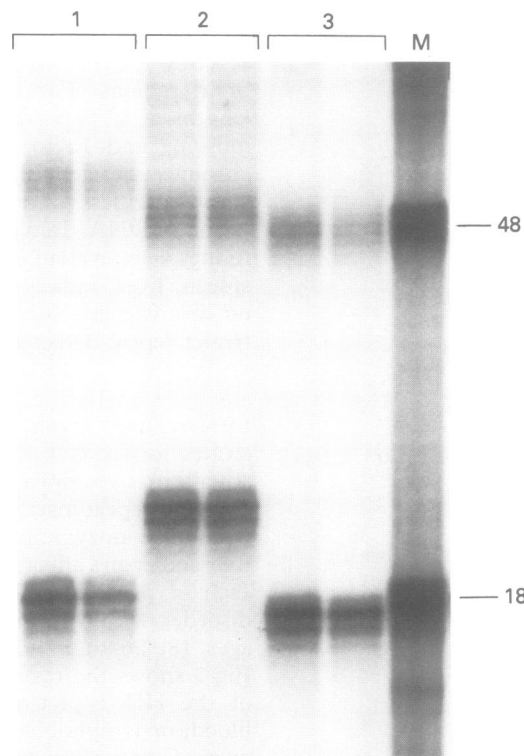

Figure 3 Comparison of HD repeat length in monozygotic twins. The HD repeat was PCR amplified from the DNA of the first three sets of identical twins described in the text. The source of DNA from each pair was frozen cerebellum, blood, and lymphoblastoid cell lines, respectively. Repeat unit lengths of the expanded and normal alleles in each twin pair were estimated as: pair 1: 54 and 18; pair 2: 50 and 23; pair 3: 48 and 17. The ladder-like pattern above the expanded alleles in pair 2 is a variable PCR artefact that appears inconsistently. Lane M contains PCR products from two cosmids known to contain 48 repeat units and 18 repeat units, respectively.

that repeat instability occurred before the twinning event in each case.

VARIATION OF HD REPEAT LENGTHS IN SPERM DNA

The apparent absence of somatic variation in HD repeat length and the identity of repeat length in monozygotic twins support the notion that the HD repeat length is determined before fertilisation. We have previously shown in sperm DNA from a limited number of HD patients that mosaicism does exist in this tissue.[3] We have now examined additional sperm samples from a total of 16 HD patients to determine the extent of the repeat length variation. The ages of the sperm donors ranged from 21 to 63 years. Although significant repeat length heterogeneity was evident in almost all sperm DNA samples, its extent was highly variable. Examples of these results, comparing HD repeat length in sperm, blood, and lymphoblast DNA are shown in fig 4. We also used scanning densitometry to compare the PCR products directly from paired sperm and lymphoblast DNA samples yielding the data shown in the table and fig 5. In general, the subjects whose somatic DNA exhibited the largest repeat lengths displayed the greatest range of repeat sizes in their sperm DNA. Moreover, in about half the cases, the modal repeat length was greater in the sperm DNA than the somatic repeat length (fig 5).

Discussion

Instability leading to expansion of trinucleotide repeats in human DNA is emerging as an important cause of genetic disease.[2-22] Five distinct disorders have already been shown to be caused by such expansions, and many additional examples will undoubtedly be uncovered. The five known cases all display expanded, unstable triplet repeats, but the precise behaviour of the mutation, and the mechanism whereby it leads to the disease phenotype differ in each case. In fragile X syndrome, explosive expansion of an already unstable and partially expanded premutation trinucleotide within the 5' untranslated region of the FMR-1 gene leads to lack of gene expression, producing the characteristic mental retardation and associated symptoms. Similar explosive expansions of a predisposing allele occur in a trinucleotide repeat in the 3' untranslated region of the myotonin kinase gene causing myotonic dystrophy. The mechanism of action of the expanded triplet is a matter of debate, but it is likely to act at the RNA level. More limited instability and shorter allele expansions are observed in spinobulbar muscular atrophy where the trinucleotide repeat exists within the coding sequence of the androgen receptor on the X chromosome.

The trinucleotide repeat on HD chromosomes is probably also located in coding sequence. The HD repeat displays a greater frequency of change and a wider range of repeat lengths on disease chromosomes than
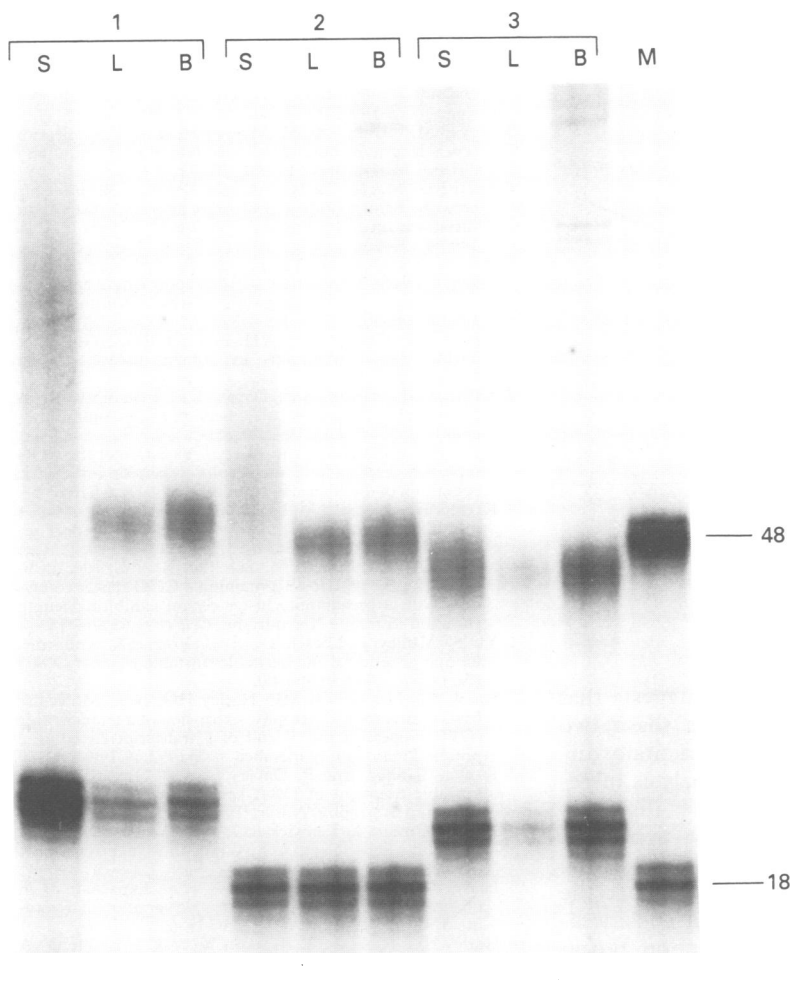
Figure 4 Comparison of HD repeat lengths in matched sperm, lymphoblast, and blood DNA. The HD CAG repeat was PCR amplified from matched DNA samples of subjects 1 to 3 from the table who span the gamut of results observed in our larger sample of 16 subjects. S = sperm DNA, L = lymphoblast DNA, B = blood DNA. Lane M contains PCR products from two cosmids known to contain 48 and 18 repeat units, respectively.

that in spinobulbar muscular atrophy, but far fewer repeat units than the disease alleles of the fragile X syndrome and myotonic dystrophy. The most recent trinucleotide repeat associated with a human disorder occurs in spinocerebellar ataxia 1 and appears to behave

Triplet repeat length in HD lymphoblast and sperm DNA.

| Subject | HD status* | Repeats units† | | | | |
| | | Lymphoblast | | Sperm | | |
		AN	AE	AN	AE range	AE mode
1	A (6 y)	24	52	24	52–90	78
2	AR	18	49	18	49–64	54
3	AR	22	44	22	44–49	44
4	AR	18	46	18	45–61	46
5	A (4 y)	18	51	18	50–80	70
6	A (2 y)	14	48	14	48–65	51
7	AR	16	42	16	42–48	43
8	AR	20	47	20	45–54	47
9	AR	15	44	15	43–47	44
10	A (6 y)	18	48	18	48–65	50
11	A (1 y)	17	45	17	41–54	46
12	A (8 y)	22	46	22	42–59	48
13	A (2 y)	18	38	18	36–47	40
14	A (7 y)	20	43	20	40–65	46
15	A (2 y)	31	43	31	41–50	44
16	AR	19	42	19	38–47	42

* HD status: number of years affected in brackets. AR = asymptomatic at risk.
† AE = expanded HD allele, AN = normal allele.

like the HD repeat, but has not yet been explored in great detail. Thus, clear differences exist in the behaviour and mechanism of action of the various trinucleotide repeats, but the question of whether the underlying mechanism of mutation is identical in all of the disorders has not been explored.

Recent studies of the fragile X syndrome have been interpreted to suggest that both the allele expansion and somatic mosaicism in this disorder result from instability of the repeat early in fetal development.[23 24] The matching pattern of allele expansion in monochorionic-monoamniotic monozygotic twins indicated that instability occurs before 10 to 20 days of fetal development, the timing of the twinning event in these cases. The absence of the fully expanded allele from the sperm of affected males suggested that instability occurs only after the laying down of the germline cells, before day 5. Thus, the apparent window for fragile X repeat instability is between day 5 and day 20 of fetal life.

Our data concerning the HD repeat are in sharp contrast to the results in the fragile X syndrome. Despite the high frequency of size changes of the repeat in HD transmissions, somatic mosaicism is rare, if it occurs at all. The tissue most affected in HD, the striatum, other regions of the brain, and peripheral tissues which are thought to be untouched by the disorder, all display the same repeat lengths within a person, as do cultured cell lines. The extremely low level or absence of somatic mosaicism in HD bodes well for using the CAG repeat length for predictive testing. Our data indicate that typing of blood DNA provides an accurate reflection of the repeat expansion found in the critical brain regions and is therefore applicable to presymptomatic diagnosis.

More than 80% of HD transmissions involve measurable changes in repeat length.[3] However, each of four pairs of HD monozygotic twins shows identical repeat lengths. Thus, the repeat instability could only have occurred before the twinning events. Since the vast majority of monozygotic twins result from an event between days 2 and 5 of fetal development, the instability of the HD repeat apparently occurs before the developmental window of fetal days 5 to 20 described for the fragile X syndrome repeat expansion.

The conclusion that the timing of repeat instability differs between these two disorders is supported by the sharp contrast in the status of the trinucleotide repeats in sperm DNA. The HD sperm display a variable degree of allele heterogeneity that appears to increase with the size of the allele present in somatic tissue. Thus, rather than arising after laying down of the germline, as in the fragile X syndrome, instability of the trinucleotide repeat in HD occurs during male gametogenesis. Since HD repeat size changes, albeit of a smaller magnitude, also occur with high frequency in transmissions from an affected mother, it is likely that repeat instability also occurs during female gametogenesis.
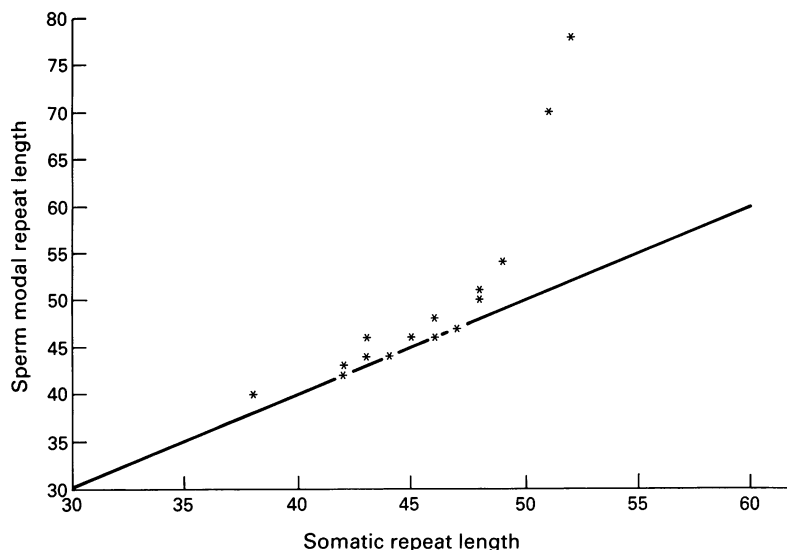
The difference in timing of repeat instability

Figure 5   Comparison of modal sperm repeat length with somatic repeat length. The modal repeat length determined by densitometry of the PCR products from sperm (table) is plotted against the somatic repeat length from the same subject.

1 Martin JB, Gusella JF. Huntington's disease: pathogenesis and management. *N Engl J Med* 1986;315:1267–76.
2 Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;72:971–83.
3 Duyao MP, Ambrose CM, Myers RH, *et al*. Trinucleotide repeat length: instability and age of onset in Huntington's disease. *Nature Genet* 1993;4:387–92.
4 Snell RG, MacMillan JC, Cheadle JP, *et al*. Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nature Genet* 1993;4:393–7.
5 Andrew SE, Goldberg YP, Kremer B, *et al*. The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nature Genet* 1993;4:398–403.
6 Caskey CT, Pizzuti A, Fu YH, Fenwick RG, Nelson DL. Triplet repeat mutations in human disease. *Science* 1992;256:784–9.
7 Suthers GK, Huson SM, Davies KE. Instability versus predictability: the molecular diagnosis of myotonic dystrophy. *J Med Genet* 1992;29:761–5.
8 Fu YH, Kuhl DPA, Pizzuti A, *et al*. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* 1991;67:1047–58.
9 Kremer EJ, Pritchard M, Lynch M, *et al*. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* 1991;252:1711–14.
10 Verkerk AJMH, Pieretti M, Sutcliffe JS, *et al*. Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 1991;65:904–14.
11 Yu S, Mulley J, Loesch D, *et al*. Fragile-X syndrome: unique genetics of the heritable unstable element. *Am J Hum Genet* 1992;50:968–80.
12 Brook JD, McCurrach ME, Harley HG, *et al*. 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* 1992;68:799–808.
13 Buxton J, Shelburne P, Davies J, *et al*. Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* 1992;355:547–8.
14 Fu YH, Pizzuti A, Fenwick RG, *et al*. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 1992;255:1256–9.
15 Harley HG, Brook JD, Rundle SA, *et al*. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* 1992;355:545–6.
16 Harley HG, Rundle SA, Reardon W, *et al*. Unstable DNA sequence in myotonic dystrophy. *Lancet* 1992;339:1125–8.
17 Mahadevan M, Tsilfidis C, Sabourin L, *et al*. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 1992;255:1253–5.
18 LaSpada AR, Wilson EM, Lubahn DB, Harding AE, Fishbeck H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 1991;352:77–9.
19 LaSpada AR, Roling D, Harding AE, *et al*. Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nature Genet* 1992;2:301–4.
20 Biancalana V, Serville F, Pommier J, *et al*. Moderate instability of the trinucleotide repeat in spino-bulbar muscular atrophy. *Hum Mol Genet* 1992;1:255–8.
21 Orr HT, Chung M, Banfi S, *et al*. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genet* 1993;4:221–6.
22 Mandel JL. Questions of expansion. *Nature Genet* 1993;4:8–9.
23 Wohrle D, Hennig I, Vogel W, Steinbach P. Mitotic stability of fragile X mutations in differentiated cells indicates early post-conceptional trinucleotide repeat expansion. *Nature Genet* 1993;4:140–2.
24 Reyniers E, Vits L, De Boulle K, *et al*. The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm. *Nature Genet* 1993;4:143–6.
25 Cavalli-Sforza LL, Bodmer WF. *The genetics of human populations*. San Francisco: WH Freeman, 1971:965.

in fragile X syndrome and HD suggests that trinucleotide repeat expansion in these two disorders might result from fundamentally different mechanisms. A detailed analysis of repeat instability in each of the other disorders that involve trinucleotide repeat expansion might shed more light on this question. If the molecular basis for expansion is in fact the same in all cases, then additional considerations, such as sperm viability, selection, repeat size range, mode of action, affected tissue, etc, must play a crucial role in specifying the distinct characteristics of repeat behaviour and transmission in each disorder.