

# Trinucleotide repeat length instability and age of onset in Huntington's disease

M. Duyao<sup>1</sup>, C. Ambrose<sup>1</sup>, R. Myers<sup>2</sup>, A. Novelletto<sup>3</sup>, F. Persichetti<sup>3</sup>, M. Frontali<sup>4</sup>, S. Folstein<sup>5</sup>, C. Ross<sup>6</sup>, M. Franz<sup>5</sup>, M. Abbott<sup>6</sup>, J. Gray<sup>7</sup>, P. Conneally<sup>7</sup>, A. Young<sup>8</sup>, J. Penney<sup>8</sup>, Z. Hollingsworth<sup>8</sup>, I. Shoulson<sup>9</sup>, A. Lazzarini<sup>10</sup>, A. Falek<sup>11</sup>, W. Koroshetz<sup>8</sup>, D. Sax<sup>2</sup>, E. Bird<sup>12</sup>, J. Vonsattel<sup>12</sup>, E. Bonilla<sup>13</sup>, J. Alvir<sup>14</sup>, J. Bickham Conde<sup>15</sup>, J.-H. Cha<sup>16</sup>, L. Dure<sup>8</sup>, F. Gomez<sup>17</sup>, M. Ramos<sup>18</sup>, J. Sanchez-Ramos<sup>19</sup>, S. Snodgrass<sup>20</sup>, M. de Young<sup>14</sup>, N. Wexler<sup>21</sup>, C. Moscovitz<sup>21</sup>, G. Penchaszadeh<sup>21</sup>, H. MacFarlane<sup>1</sup>, M. Anderson<sup>1</sup>, B. Jenkins<sup>1</sup>, J. Srinidhi<sup>1</sup>, G. Barnes<sup>1</sup>, J. Gusella<sup>1,22</sup> & M. MacDonald<sup>1</sup>

The initial observation of an expanded and unstable trinucleotide repeat in the Huntington's disease gene has now been confirmed and extended in 150 independent Huntington's disease families. HD chromosomes contained 37–86 repeat units, whereas normal chromosomes displayed 11–34 repeats. The HD repeat length was inversely correlated with the age of onset of the disorder. The HD repeat was unstable in more than 80% of meiotic transmissions showing both increases and decreases in size with the largest increases occurring in paternal transmissions. The targeting of spermatogenesis as a particular source of repeat instability is reflected in the repeat distribution of HD sperm DNA. The analysis of the length and instability of individual repeats in members of these families has profound implications for presymptomatic diagnosis.

Huntington's disease (HD) is a devastating inherited neurodegenerative disorder characterized by progressive disordered movements and psychological changes including cognitive decline and emotional disturbance<sup>1</sup>. The disorder typically has an insidious onset in middle-age (mean ~40 years), but the age span among those afflicted ranges from very young to very old. A preponderance of the juvenile onset cases results from male transmission of the disease gene<sup>2,3</sup>. The motor disturbance of HD begins as minor involuntary movements that become progressively exaggerated and incapacitating until they consume the entire body in chorea. There is no treatment to prevent the onset or to delay the relentlessly fatal course of HD.

Recently, we isolated the defective gene causing Huntington's disease and found that it contains a (CAG)<sub>n</sub> trinucleotide repeat that is expanded and unstable on disease chromosomes<sup>4</sup>. A limited sample of HD chromosomes contained 42–86 copies of the triplet whereas normal chromosomes possessed 11–34 copies. Our initial study also suggested that there might be a correlation between the length of the triplet repeat and the age of onset of HD. The discovery of an unstable trinucleotide repeat in the HD gene has implications for developing a direct presymptomatic diagnostic test as well as for explaining the phenotypic variation in the disorder. We have therefore extended our analysis to a much larger set of HD and normal chromosomes to address critical questions raised by the presence of the trinucleotide repeat.

**Repeat lengths on normal and HD chromosomes**  
Our initial report identified a (CAG)<sub>n</sub> trinucleotide repeat that appeared to be expanded and unstable on a sampling

of HD chromosomes. We have now examined the number of copies of this triplet repeat on a total of 425 HD chromosomes from 150 independent families compared with 545 normal chromosomes. This comparison was performed by polymerase chain reaction (PCR) analysis and cannot exclude that the size variations detected do not occur in flanking sequences. However, we presume that the size differences in most or all cases reflect variation in the (CAG)<sub>n</sub> sequence. The results of this analysis are displayed in Fig. 1. We observed two non-overlapping

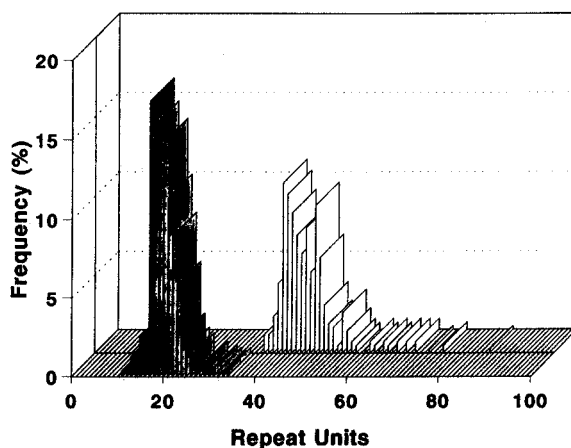


Fig. 1 Comparison of (CAG)<sub>n</sub> repeat unit number on control and HD chromosomes. Frequency distributions are shown for the number of (CAG)<sub>n</sub> repeat units observed on 425 HD chromosomes from 150 independent families and from 545 control chromosomes. Shaded bars, control chromosomes; open bars, HD chromosomes.

<sup>1</sup>Molecular Neurogenetics Unit, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA<sup>2</sup>Department of Neurology, Boston University Medical School, Boston, Massachusetts 02118, USA<sup>3</sup>Department of Biology, University of Rome "Tor Vergata", Rome, Italy<sup>4</sup>Istituto di Medicina Sperimentale, CNR, Rome, Italy<sup>5</sup>Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA<sup>6</sup>Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA<sup>7</sup>Department of Medical Genetics, Indiana University Medical Center, Indianapolis, Indiana 46202, USA<sup>8</sup>Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA<sup>9</sup>Department of Neurology, University of Rochester Medical Center, Rochester, New York 14642, USA<sup>10</sup>Department of Neurology, Robert Wood Johnson Medical School, New Brunswick, New Jersey 08903, USA<sup>11</sup>Department of Psychiatry, Emory University School of Medicine, Atlanta, Georgia 30306, USA<sup>12</sup>Brain Tissue Resource Center, McLean Hospital and Department of Neuropathology, Harvard Medical School, Belmont, Massachusetts 02178, USA<sup>13</sup>Universidad de Zulia, Maracaibo, Venezuela<sup>14</sup>Long Island Jewish Hillside Medical Center, Glen Oaks, New York 11004, USA<sup>15</sup>Baylor University College of Medicine, Houston, Texas 77009, USA<sup>16</sup>Neuroscience Laboratory, Kennedy Krieger Institute, Baltimore Maryland 21205, USA<sup>17</sup>Aguirre 621-7A, Capital 1414, Buenos Aires, Argentina<sup>18</sup>Servicio de Genética, Hospital Virgen del Camino, 31008 Pamplona, Spain<sup>19</sup>Department of Neurology, University of Miami, Miami, Florida 33136, USA<sup>20</sup>Pediatric Neurology, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA<sup>21</sup>Departments of Neurology and Psychiatry, College of Physicians and Surgeons, Columbia University, New York New York 10032 and Hereditary Disease Foundation, 1427 7th St., Suite 2, Santa Monica California 90401, USA<sup>22</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA

Correspondence should be addressed to M.M.

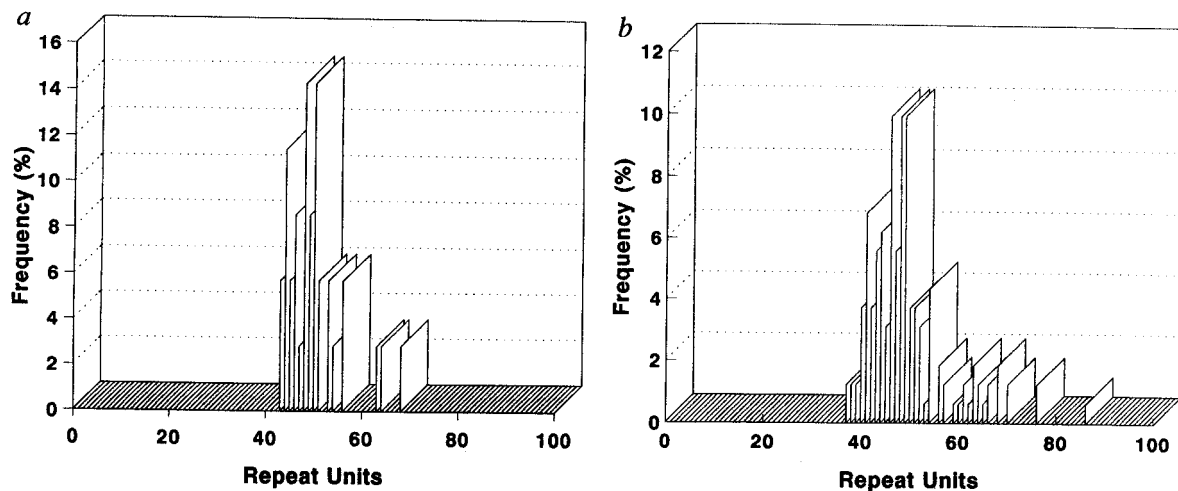
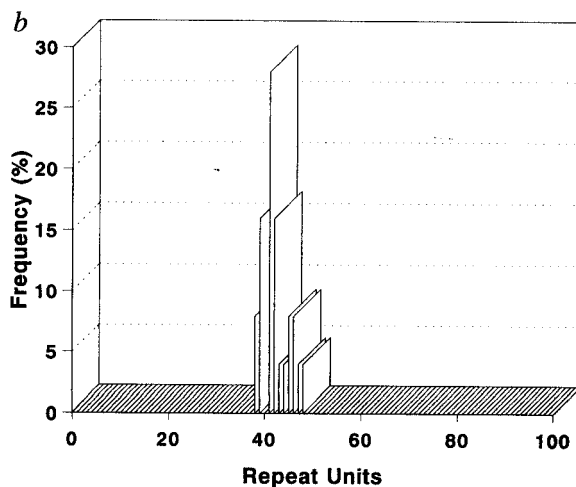
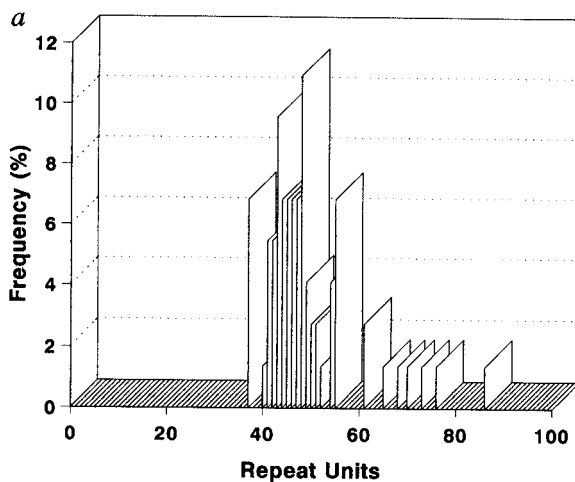


Fig. 2 Comparison of  $(CAG)_n$  repeat unit number on maternally and paternally transmitted HD chromosomes. Frequency distributions are shown for the 134 and 161 HD chromosomes from Fig. 1 known to have been transmitted from the mother (a) and father (b), respectively. The two distributions differ significantly based on a t-test ( $t_{272,3} = 5.34$ ,  $p < 0.0001$ ).

distributions of repeat length, although the upper end of the normal range and the lower end of the HD range were separated only by 3 repeat units. The normal chromosomes displayed 24 alleles producing PCR products ranging from 11–34 repeat units, with a median of 19 units (mean 19.71, s.d. 3.21). The HD chromosomes yielded 54 discrete PCR products corresponding to repeat lengths of 37–86 units, with a median of 45 units (mean 46.42, s.d. 6.68).

Of the HD chromosomes, 134 and 161 were known to be maternally or paternally-derived, respectively. To investigate whether the sex of the transmitting parent

might influence the distribution of repeat lengths, we plotted these two sets of chromosomes separately Fig. 2. The maternally-derived chromosomes displayed repeat lengths ranging from 37 to 73 units, with a median of 44 (mean 44.93, s.d. 5.14). The paternally-derived chromosomes had 37 to 86 copies of the repeat unit, with a median of 48 units (mean 49.14, s.d. 8.27). However, a higher proportion of the paternally-derived HD chromosomes had repeat lengths greater than 55 units (16% versus 2%), suggesting the possibility of a differential effect of paternal versus maternal transmission.



### Instability of the trinucleotide repeat

The data in Fig. 1 combine repeat lengths from 150 different HD families representing many potentially independent origins of the defect. To examine the variation in repeat lengths on sets of HD chromosomes known to descend from a common founder, we have separated the

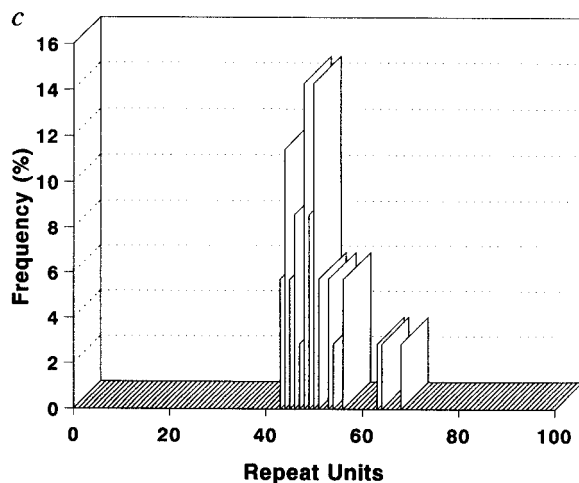


Fig. 3 Comparison of  $(CAG)_n$  repeat unit number on HD chromosomes from three large families with different HD founders. Frequency distributions are shown for 75, 25 and 35 HD chromosomes from the Venezuelan HD family (a)<sup>5,6</sup>, Family Z (b) and Family D (c)<sup>7</sup>, respectively. The Venezuelan distribution did not differ from the overall HD chromosome distribution in Fig. 1 ( $t_{79,7} = 1.58$ ,  $p < 0.12$ ). Both Family Z and Family D did produce distributions significantly different from the overall HD distribution ( $t_{22,2} = 6.73$ ,  $p < 0.0001$  and  $t_{458} = 2.90$ ,  $p < 0.004$ , respectively).

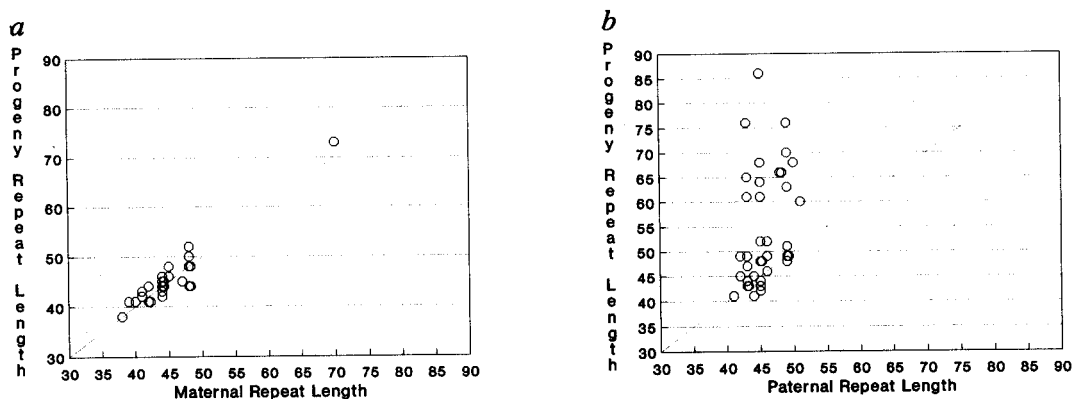


Fig. 4 Relationship of (CAG)<sub>n</sub> repeat length in parents and corresponding progeny. Repeat length on the HD chromosome in mothers (a) or fathers (b) is plotted against the repeat length in the corresponding offspring. A total of 25 maternal transmissions and 37 paternal transmissions were available for typing.

data from three large HD kindreds<sup>5-7</sup> with different chromosome 4p16.3 haplotypes<sup>8</sup>, typed for 75, 25 and 35 individuals, respectively. Despite the single origin of the founder HD chromosome within each pedigree, members of the separate pedigrees display a wide range of repeat lengths (Fig. 3). This instability of the HD chromosome repeat is most prominent in members of a large Venezuelan HD kindred (Fig. 3a) in which the common HD ancestor has produced 10 generations of descendants, numbering over 13,000 individuals. The distribution of repeat lengths in this sampling of the Venezuelan pedigree (median 46, mean 48.26, s.d. 9.3) is not significantly different from that of the larger sample of HD chromosomes from all families. Fig. 3b and c display results for two extended families in which HD was introduced more recently than in the Venezuelan kindred. These families have been reported to exhibit different age of onset distributions and varied phenotypic features of HD<sup>7</sup>. Both revealed extensive repeat length variation, with a median of 41 and 49 repeat units, respectively. The distribution of repeat lengths in the members of the family in Fig. 3b was significantly different from the distribution of all HD chromosome repeat lengths ( $p < 0.0001$ ), with a smaller mean of 42.04 repeat units (s.d. 2.82). The repeat distribution from HD chromosomes of Fig. 3c was also significantly different from the total data set ( $p < 0.004$ ), but with a higher mean of 49.80 (s.d. 5.86).

**Parental source effects on repeat length variation**

For 62 HD chromosomes in Fig. 1, the length of the trinucleotide repeat also could be examined on the corresponding parental HD chromosome. In 20 of 25 maternal transmissions, and in 31 of 37 paternal transmissions, the repeat length was altered, indicating considerable instability. A similar phenomenon was not observed for normal chromosomes, where more than 500 meiotic transmissions revealed no changes in repeat length, although the very existence of such a large number of normal alleles suggests at least a low degree of instability.

Figure 4 shows the relationship between the repeat lengths on the HD chromosomes in the affected parent and corresponding progeny. For the 20 maternally-inherited chromosomes on which the repeat length was altered, 13 changes were increases in length and 7 were decreases. Both increases and decreases involved changes of less than 5 repeat units and the overall correlation between the mother's repeat length and that of her child

was  $r = 0.95$  ( $p < 0.0001$ ). The average change in repeat length in the 25 maternal transmissions was an increase of 0.4 repeats.

On paternally-derived chromosomes, the 31 transmissions in which the repeat length changes comprised 26 length increases and 5 length decreases. Although the decreases in size were only slightly smaller than those observed on maternally-derived chromosomes, ranging from 1-3 repeat units, the increases were often dramatically larger. Thus, the correlation of the repeat length in the father with that of his offspring was only  $r = 0.35$  ( $p < 0.04$ ). The average change in the 37 paternal transmissions was an increase of 9 repeat units. The maximum length increase observed through paternal transmission was 41 repeat units, a near doubling of the parental repeat.

For both male and female transmissions, there was no correlation between the size of the parental repeat and either the magnitude or frequency of changes.

**Repeat length variation in male gametogenesis**

To determine whether the variation in the length of the repeat observed through male transmission of HD chromosomes is reflected in the male germ cells, we amplified the repeat from sperm DNA and from DNA of the corresponding lymphoblast from 5 HD gene carriers. The results reveal striking differences between the lymphoblast and sperm DNA for the HD chromosome repeat, but not for the repeat on the normal chromosome (Fig. 5). All the sperm donors are members of the Venezuelan HD family and range in age from 24-30 years. Individuals 1 and 2 are siblings with HD chromosome repeat lengths based on lymphoblast DNA of 45 and 52, respectively. Individuals 3 and 4 are also siblings, with HD repeat lengths of 46 and 49, respectively. Individual 5, from a different sibship than either of the other two pairs, has an HD repeat of 52 copies. In all 5 cases, the PCR amplification of sperm DNA and lymphoblast DNA yielded identical products from the normal chromosome. However, in comparison with lymphoblast DNA, the HD gene from sperm DNA yielded a diffuse array of products. In 3 of the 5 cases (2, 4 and 5), the diffuse array spread to much larger allelic products than the corresponding lymphoblast product. Subject 2 showed the greatest range of expansion, with the sperm DNA product extending to over 80 repeat units. Interestingly, the 3 individuals displaying the greatest variation have the longest repeats

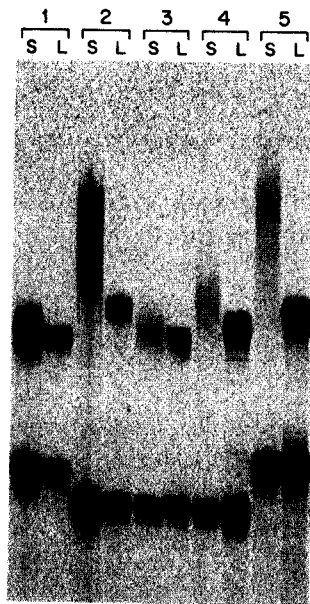


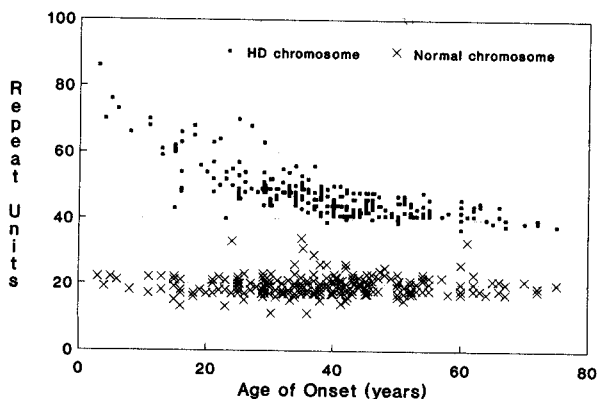
Fig. 5 Amplification of the *HD* (CAG)<sub>n</sub> repeat from sperm and lymphoblast DNA. DNA from sperm (S) and lymphoblasts (L) for 5 members (pairs 1–5) of the Venezuelan *HD* pedigree aged 24–30 were used for PCR amplification<sup>4</sup> of the *HD* (CAG)<sub>n</sub> repeat. The lower band in each lane derives from the normal chromosome.

and are currently symptomatic. The other two donors have shorter repeat lengths in the *HD* range, and remain at risk.

#### Relationship between repeat length and age of onset

We have suggested previously that increased repeat length might correlate with a reduced age of onset of *HD*. Age of onset data could be determined for 234 of the individuals represented in Fig. 1. Figure 6 displays the repeat lengths found on the *HD* and normal chromosomes of these individuals relative to their age of onset. Indeed, age of onset is inversely correlated with the *HD* repeat length. A Pearson correlation coefficient of  $r = -.75$ ,  $p < 0.0001$  was obtained assuming a linear relationship between age of onset and repeat length. When a polynomial function was used, a better fit was obtained ( $R^2 = 0.61$ ,  $F = 121.45$ ), suggesting a higher order association between age of onset and repeat length.

Although very strong statistical correlations were obtained in the analysis of all *HD* chromosomes, repeat length is not a good predictor of age of onset for any given



individual. There is considerable variation in the age of onset associated with any specific number of repeat units, particularly for trinucleotide repeats in the 37–52 unit range (88% of *HD* chromosomes) where onset ranged from 15–75 years. In this range, a linear relationship between age of onset and repeat length provided as good a fit as a higher order relationship. The 95% confidence interval surrounding the predicted regression line was estimated at  $\pm 18$  years. In the 37–52 unit range, the association of repeat length to onset age is only half as strong as in the overall distribution ( $r = -0.40$ ,  $p < 0.0001$ ), indicating that much of the predictive power is contributed by repeats longer than 52 units. In this increased range, onset is likely to be very young and consequently not relevant to most persons seeking testing.

For the 178 cases in the 37–52 repeat unit range for which it was possible to subdivide the data set based on parental origin of the *HD* gene, multivariate regression analysis suggested a significant effect of parental origin on age of onset ( $p < 0.05$ ) independent of repeat length in this range. *HD* gene carriers from maternal transmissions had an average age of onset two years later than those from paternal transmissions. Larger data sets and additional analyses will be required to confirm the validity of this preliminary observation.

In both univariate and multivariate analyses, we detected no association between age of onset and the repeat length on the normal chromosome, either in the total data set, or when it was subdivided into chromosomes of maternal or paternal origin.

#### Discussion

The overall distribution of repeat lengths on the 425 *HD* chromosomes was completely non-overlapping with that observed for control chromosomes. At first glance, these data suggest that sizing of the repeat unit could provide a completely accurate diagnostic for *HD*. However, this perception is seriously misleading, since a number of considerations complicate the diagnostic use of repeat length. First, the separation between the distributions is only 3 repeat units, suggesting that larger data sets will very likely contain chromosomes within the intervening zone, creating an overlap.

Second, our data set excluded chromosomes from a few clinically diagnosed individuals who have previously been shown not to have inherited the *HD* chromosome by DNA marker linkage studies<sup>9,10</sup>. These four individuals from three well-characterized *HD* pedigrees inherited a chromosome haplotype characteristic of the normal chromosome in their respective affected parent. As expected, these four individuals have repeat lengths well within the normal range<sup>4</sup>, despite the fact that *HD* affecteds in the remainder of each pedigree display repeat lengths well within the *HD* range. Their disease manifestations, including a typical age of onset, have not been explained, and these persons may represent phenocopies of *HD*. Regardless of the mechanism involved, the occurrence at low frequency of such individuals within known *HD*

Fig. 6 Relationship of repeat unit length with age of onset. Age of onset was established for 234 diagnosed *HD* gene carriers and plotted against the repeat length observed on both the *HD* and normal chromosomes in the corresponding lymphoblast lines.

families must be considered if diagnostic conclusions are based solely on repeat length.

Finally, our control data set excludes a number of chromosomes from phenotypically normal individuals who are related to "spontaneous" cases of HD or "new mutations"<sup>34</sup> (R.H.M. *et al.*, manuscript submitted). Chromosomes from these individuals who are not clinically affected and have no family history of the disorder cannot be designated as HD. However, these chromosomes cannot be classified as unambiguously normal because they are essentially the same chromosome as that of an affected relative, the diagnosed "spontaneous" HD proband, except with respect to repeat length. The lengths of repeat found on these ambiguous chromosomes (34–38 units) span the gap between the control and HD distributions, confounding a decision on the status of any individual with a repeat in the high normal to low HD range.

The striking difference in the high repeat length range (>55) between HD chromosomes transmitted from the father and those transmitted from the mother indicated a potential parental source effect. When this was examined directly, the HD chromosome repeat length changed in about 85% of transmissions. Most changes involved a fluctuation of only a few repeat units, with larger increases occurring only in male transmissions. The greater size increases in male transmission appear to be caused by particular instability of the HD trinucleotide repeat during male gametogenesis, based on the amplification of the repeat from sperm DNA. It remains to be determined whether this instability is inherent in the size or sequence of the repeat itself, or can be affected by paternal age, environmental influences or possibly affected status. It is also conceivable that the repeat length is similarly unstable in female gametogenesis but that larger expansions are incompatible with production of a viable oocyte.

The analysis of individual families underscores the instability of the HD trinucleotide repeat. Even though in each of three large families all HD chromosomes can be traced to a common ancestor, the related HD chromosomes display an array of different sizes of repeats. In a sampling of the Venezuelan HD pedigree, the repeat length variation mirrored that seen in the overall population, whereas in two smaller families of American origin, a tighter distribution was seen. It remains to be determined whether these differences are solely due to the number of generations since introduction of the HD chromosome into the family, or whether HD chromosomes of different haplotypes predispose to a restricted range of repeat lengths.

The discovery of the unstable trinucleotide repeat in HD offered a potential explanation for the wide variations in age of onset of the disorder with the observation that the longest repeats were found in juvenile onset HD cases. Analysis of our data set indicates that repeat length is only a partial determinant of age of onset. While the overall correlation between repeat length and age of onset is striking, a broad range of onset ages is associated with any given repeat length. The spread in ages of onset is especially large in the lower end of the HD repeat distribution (37–52 units), suggesting that in these individuals, environmental effects, genetic modifiers or stochastic processes play a significant role in determining onset.

The delineation of distinct distributions for repeat lengths on normal and HD chromosomes justifies the use

of this assay for direct prediction of HD status in most instances. However, implementation of this radical change in diagnostic capacity should be approached with a caution similar to that occasioned by the discovery of a linkage marker for HD in 1983. Hopefully, the experience gained in developing and using a genetic counseling protocol for the HD linkage test<sup>11,12</sup> will provide the basis for a modified approach that takes into account both the peculiarities of direct testing and of the findings specific to HD described above. There are numerous special concerns in dealing with repeat lengths of about 30–40 units in this disorder. Moreover, it is clear from the wide range in ages of onset associated with any particular repeat length that using the assay to predict age of onset or severity of disease is not warranted and should not be done. With the above concerns taken into consideration, the direct monitoring of the (CAG)<sub>n</sub> repeat in the HD gene should provide a less expensive and more widely applicable diagnostic procedure than linkage testing.

The trinucleotide repeat expansion in HD is the fifth example of this mechanism in human disease. Unstable trinucleotide repeats have been observed in fragile X syndrome, myotonic dystrophy and spino-bulbar muscular atrophy (SBMA)<sup>13–29</sup>. The unstable HD repeat behaves in a fundamentally different manner from the repeats in fragile X syndrome and in myotonic dystrophy. Both of these disorders often show much larger expansions of the trinucleotide stretch than are seen in HD, and instability increases with the size of the repeat. Significant changes in repeat length are usually associated with maternal transmission in fragile X syndrome and are seen from either parent in myotonic dystrophy. The largest changes in HD are associated with paternal transmission.

In general, the behaviour of the HD trinucleotide repeat most closely resembles that seen in SBMA, but there are significant differences. Both repeat segments display similar normal and disease-associated size ranges. The disease-associated repeats typically fluctuate by less than 5 units and decreases in repeat length are not uncommon. However, the HD repeat appears to be more unstable than that in SBMA. Changes in repeat length occur in most transmission from either sex across the entire range of HD repeat sizes, with the occasional larger expansions resulting exclusively from male transmission. In SBMA, size changes occur less frequently than in HD. Although instability in SBMA appears to be increased in male transmissions, the resulting size changes do not match the magnitude of some paternally-transmitted HD alleles. Thus, the details of repeat expansion in HD do not precisely duplicate the experience in SBMA, and are quite distinct from observations in fragile X syndrome and myotonic dystrophy, suggesting that different mechanisms of instability may be operating.

Very recently a fifth example of trinucleotide repeat expansion in human disease has been reported<sup>33</sup>. Spinocerebellar ataxia type 1 (SCA1) is associated with an expanded CAG repeat which displays normal and disease variation similar to that seen in HD. Like HD, the age of onset in SCA1 appears to be inversely correlated with age of onset of the disease and shows a bias toward larger size increases during male transmission. More detailed analyses will be required to define the precise parallels between the behavior of the triplet repeat in SCA1 and HD. Interestingly, a second locus for autosomal dominant cerebellar ataxia has recently been mapped to 12q (ref. 34)

**Acknowledgements**

We thank the many investigators who have supplied blood samples, pedigree information from HD families or advice and support. We are extremely grateful to the HD families for supporting and participating in this research effort. We thank the Brain Tissue Resource Center for brain samples. This work was supported by NIH grants NS16367 (Huntington's Disease Center Without Walls), NS16375, NS17978, NS02385 (HD Roster) and NS22031, by CNR grant 93.00031.PF99 and by grants from Bristol-Myers Squibb, Inc., the Hereditary Disease Foundation Collaborative Research Agreement, the Joan and William J. Schreyer/Merrill Lynch Fund to Cure Huntington's Disease of the Hereditary Disease Foundation, the Huntington's Disease Society of America, the Foundation for the Care and Cure of Huntington's disease, the W.M. Keck Foundation, the William J. Matheson Foundation, the Bay Foundation, and the Charles A. Dana Foundation. C.M.A. and M.P.D. received fellowship support from the Andrew B. Cogan Fellowship of the Hereditary Disease Foundation and from the Huntington's Disease Society of America.

and would seem a likely candidate in what is probably a long list of neurological diseases involving this mechanism.

**Methodology**

**Typing of HD and normal chromosomes.** HD chromosomes were derived from symptomatic individuals and "at risk" individuals known to be gene carriers by linkage marker analysis. All HD chromosomes were from members of well-characterized HD families of varied ethnic backgrounds used previously for genetic linkage and disequilibrium studies<sup>8,30</sup>. Three of the 150 families used were large pedigrees, each descended from a single founder. The large Venezuelan HD pedigree is an extended kindred of over 13,000 members from which we typed 75 HD chromosomes<sup>5,6</sup>. Two other large families that have been described previously as Family Z and Family D, provided 25 and 35 HD chromosomes, respectively<sup>7</sup>. Normal chromosomes were taken from married-ins in the HD families and from unrelated normal individuals from non-HD

families. The DNA tested for all individuals except four was prepared from lymphoblastoid cell lines or fresh blood<sup>5,31</sup>. In the exceptional cases, DNA was prepared from frozen cerebellum. No difference in the characteristics of the PCR products were observed between lymphoblastoid, fresh blood or brain DNAs. For five members of the Venezuelan pedigree aged 24–30, we also prepared DNA by extracting pelleted sperm from semen samples. The length of the HD gene (CAG)<sub>n</sub> repeat for all DNAs was assessed using PCR, as described<sup>4</sup>.

**Statistical analysis.** Associations between repeat lengths and onset age were assessed by Pearson correlation coefficient and by multivariate regression to assess higher order associations. Comparisons of the distributions of repeat length for all HD chromosomes and those for individual families were made by analysis of variance and t-test contrasts between groups. The 95% confidence bands were computed around the regression line utilizing the general linear models procedure of SAS<sup>32</sup>.

Received 6 May; accepted 11 June 1993.

- Martin, J.B. & Gusella, J.F. Huntington's disease: Pathogenesis and management. *New Engl. J. Med.* **315**, 1267–1276 (1986).
- Merrit, A.D., Conneally, P.M., Rahman, N.F. & Drew, A.L. Juvenile Huntington's chorea. In *Progress in neurogenetics* (eds Barbeau A. & Brunette, J.R.) 645–650 (Excerpta Medica, Amsterdam, 1969).
- Bird, E.D., Caro, A.J. & Pilling, J.B. A sex related factor in the inheritance of Huntington's chorea. *Ann. hum. Genet.* **37**, 255–260 (1974).
- Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983 (1993).
- Gusella, J.F. et al. A polymorphic DNA marker genetically linked to Huntington's Disease. *Nature* **306**, 234–238 (1983).
- Wexler, N.S. et al. Homozygotes for Huntington's disease. *Nature* **326**, 194–197 (1987).
- Folstein, S.E. et al. Huntington's Disease: Two families with differing clinical features show linkage to the G8 probe. *Science* **229**, 776–779 (1985).
- MacDonald, M.E. et al. The Huntington's disease candidate region exhibits many different haplotypes. *Nature Genet.* **1**, 99–103 (1992).
- MacDonald, M.E. et al. Recombination events suggest possible locations for the Huntington's disease gene. *Neuron* **3**, 183–190 (1989).
- Pritchard, C. et al. Recombination of 4p16 DNA markers in an unusual family with Huntington disease. *Am. J. hum. Genet.* **50**, 1218–1230 (1992).
- Meissen, G.J. et al. Predictive testing for Huntington's disease with use of a linked DNA marker. *New Engl. J. Med.* **318**, 535–542 (1988).
- Brandt, J. et al. Presymptomatic diagnosis of delayed-onset disease with linked DNA markers: The experience in Huntington's disease. *J. Am. med. Assoc.* **261**, 3108–3114 (1989).
- Suthers, G.K., Huson, S.M. & Davies, K.E. Instability versus predictability: the molecular diagnosis of myotonic dystrophy. *J. med. Genet.* **29**, 761–765 (1992).
- Fu, Y.H. et al. Variation of the CCG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67**, 1047–1058 (1991).
- Kremer, E.J. et al. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)<sub>n</sub>. *Science* **252**, 1711–1714 (1991).
- Verkerk, A.J.M.H. 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* **65**, 904–914 (1991).
- Yu, S. et al. Fragile-X syndrome: unique genetics of the heritable unstable element. *Am. J. hum. Genet.* **50**, 968–980 (1992).
- Brook, J.D. 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* **68**, 799–808 (1992).
- Aslanidis, C. et al. Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature* **355**, 548–551 (1992).
- Buxton, J. et al. Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* **355**, 547–548 (1992).
- Fu, Y.H. et al. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**, 1256–1259 (1992).
- Harley, H.G. et al. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* **355**, 545–546 (1992).
- Harley, H.G. et al. Unstable DNA sequence in myotonic dystrophy. *Lancet* **339**, 1125–1128 (1992).
- Mahadevan, M. et al. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**, 1253–1255 (1992).
- Tsilfidis, C., McKenzie, A.E., Mettler, G., Barcelo, J. & Komeluk, R.G. Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. *Nature Genet.* **1**, 192–195 (1992).
- Bruner, H.G. et al. Reverse mutation in myotonic dystrophy. *New Engl. J. Med.* **328**, 476–480 (1993).
- LaSpada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E. & Fishbeck, H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**, 77–79 (1991).
- LaSpada, A.R. et al. Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nature Genet.* **2**, 301–304 (1992).
- Biancalana, V., Serville, F., Pommier, J., Julien, J., Hanauer, A. & Mandel, J.L. Moderate instability of the trinucleotide repeat in spino-bulbar muscular atrophy. *Hum. molec. Genet.* **1**, 255–258 (1992).
- Conneally, P.M. et al. Huntington disease: No evidence for locus heterogeneity. *Genomics* **5**, 304–308 (1989).
- Anderson, M.A. & Gusella, J.F. Use of cyclosporin A in establishing Epstein-Barr virus-transformed human lymphoblastoid cell lines. *In Vitro* **20**, 856–858 (1984).
- SAS/STAT User's Guide, Version 6, 4th edn, Vol. 2 (SAS Institute Inc., Cary, North Carolina, 1989).
- Orr, H.T. et al. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genet.* **4**, 221–226 (1993).
- Gispert, S. et al. Chromosomal assignment of the second locus for autosomal dominant cerebellar ataxia (SCA2) to chromosome 12q23–24.1 *Nature Genet.* **4**, 295–299 (1993).