A study of the Huntington’s disease associated trinucleotide repeat in the Scottish population

Lilias H Barron, Jonathan P Warner, Mary Porteous, Susan Holloway, Sheila Simpson, Rosemarie Davidson, David J H Brock

Abstract
Accurate measurements of a specific CAG repeat sequence in the Huntington’s disease (HD) gene in 337 HD patients and 229 normal controls from the Scottish population showed a range from 35 to 62 repeats in affected subjects and eight to 33 in normal subjects. A link between early onset of symptoms and very high repeat number was seen. For HD patients with the most common affected allele sizes (39 to 42 repeats) absolute repeat size was a poor index for the age at onset of symptoms. There was variability in the transmitted repeat size for both sexes in the HD size range. We observed a significant increase of repeat size for paternal transmission of the disease and greater instability for paternally transmitted CAG repeats in the HD size range.

Huntington’s disease (HD) is a progressive neurodegenerative disorder characterised by motor disturbance, cognitive loss, and psychiatric manifestations.1 It is inherited in an autosomal dominant fashion and affects approximately 1 in 10,000 subjects in most European populations.2 Huntington’s disease is typified by a distinctive choreic movement disorder with an onset usually in the fourth or fifth decade, with a gradual worsening over a period of 10 to 20 years until death. An earlier onset or juvenile form is occasionally seen and is associated with a higher proportion of cases with a paternal transmission of the disease allele.3 Recently variable expansion of an unstable DNA fragment contained in the 5’ transcribed region of a new gene, IT15 or huntingtin, has been identified as the mutational mechanism causing HD.4 Initial reports have suggested that the expanded (CAG)n repeat is present on all HD chromosomes tested and, although the mechanism by which increased trinucleotide repeat length leads to the characteristic neuropathology of HD is still unknown, there is no longer any doubt that the mystery surrounding the genetic basis of HD has been resolved. The repeat is highly variable and initial reports suggested 17 discrete allele sizes ranging from about 11 to 34 copies of the CAG trinucleotide in the normal population and an expanded range from 42 to 100 copies on Huntington’s chromosomes. A direct correlation between age of onset and size of the expansion has been proposed with the largest CAG expansions being found in juvenile onset Huntington’s patients. There have been no reports of mitotic instability of the

CAG expansions and huntingtin is expressed in a wide range of tissues. Evidence for meiotic instability of the (CAG)n repeat in spermatogenesis has recently been presented (G P Bates, personal communication).

We have used our own PCR assay to determine the exact size of the (CAG)n repeat in a large set of normal and HD subjects from the Scottish population. We have used these results to address the questions of (1) the size distribution in the normal and HD populations, and (2) the effect of paternal or maternal transmission of the disease allele.

Methods
Our PCR assay has been published.5 After this publication variation in the (CCG)n rich region immediately 3’ to the (CAG)n repeat has been reported.6 This region contains at least four discrete alleles in the general population varying in size over a range of at least 9 bp. In the East Anglian population there is some evidence for linkage disequilibrium between the smallest allele size, which corresponds to the originally published sequence of this interval, and HD chromosomes. PCR assays including this further variable region will give inaccurate sizes for the CAG expansion. Although several improved PCR assays have been published since the publication of the original paper by the Huntington’s Consortium,7,8 our assay is the only method published that does not amplify across this second variable region.

Results
ACCURATE SIZE DETERMINATION OF THE HD CAG REPEAT
We present some typical results obtained using our assay (figs 1, 2, and 3). The best resolution for accurate sizing of the PCR product is obtained using 6% denaturing acrylamide gel electrophoresis. We have found that the identification of the absolute product size is hampered if both DNA strands are visible. Incorporation of either [α-32P]dTTP or dATP into the PCR product results in only one strand being labelled and halves the multiple shadow bands associated with PCR amplification across repetitive regions. The most intense band is easily identified. We have tried to compare our radiolabelled PCR products to a number of molecular weight ladders and sequencing ladders but have found that the only unambiguous control was to use PCR product from the amplification of cosmids clone L191F1,4 which has been sequenced and shown to contain 18 CAG repeats. The PCR
Figure 1  An ethidium bromide stained 2% agarose gel showing PCR products obtained using our assay. To show the high resolution of this assay using 2% agarose gel electrophoresis we selected a group of people with a common allele size of 16 CAG repeats (+1 repeat) and varied the size of the other allele. The second allele sizes are shown under the lanes and size increases from left to right. We have used standard PCR amplification conditions for all the tracks as this assay produces robust amounts of product across the whole of our normal and HD size ranges of CAG repeats. The PCR product is so small that the difference between our smallest HD allele (35 CAG) and largest normal allele (33 CAG) can be resolved (see centre and far right). Note the intense 62 bp constant band (CB) that is always produced. This band contains five CAG repeats. B = distilled water control.

Figure 2  An ethidium bromide stained 2% agarose gel. For this figure we selected a group of people with a frequently encountered HD allele size of 41 CAG repeats and varied the size of the normal allele. The second allele sizes are shown under the lanes and size increases from left to right. We have used standard PCR amplification conditions for all the tracks. Note the ease with which a very high normal repeat size (31 CAG) is resolved compared to a small but common HD allele. CB = constant band, B = distilled water blank.

Figure 3  This gel shows the resolving power of an ethidium bromide stained 2% agarose gel when CAG repeat sizes are near the top of the normal range (33 CAG repeats) and the bottom of the HD range (36 CAG). Two affected subjects with high normal allele sizes (33 and 31 repeats) are shown and a typical normal subject homozygous for 16 repeats was run as a comparison. We have used standard PCR amplification conditions for all the tracks. CB = constant band, B = distilled water blank.

The PCR assay always produces a constant band even in the blank control. This constant band corresponds to a repeat size of five CAG repeats when compared to cosmid L191F. If autoradiographs are overexposed the ladder of stutter bands starting at five CAG repeats and extending in 3 bp steps up the gel makes the size determination of the alleles a straightforward exercise (fig 4). We have fluorescently labelled one of the primers and a Pharmacia automated fluorescent DNA sequencer has been used to perform a fluorescent DNA typing assay on the PCR product resolved using 6% denaturing acrylamide electrophoresis. The analysis software scales the intensities of the fluorescent peaks with respect to the strongest signal which simplifies identification of alleles (fig 5).

The accurate sizing of the CAG repeat sequence in the huntingtin gene is an issue of very great importance. Other groups have found that the size distributions of the normal and HD CAG repeats show a small overlap at a value of about 30 repeats. The CAG repeat sizes of a very small number of affected subjects from large HD families with lower than the average family repeat size fall within the proposed region of overlap (G P Bates, personal communication). Two families where new HD mutations have occurred as a result of an expansion from the normal to the affected size range have been reported (G P Bates, personal communication). The ability to determine absolute repeat sizes and ensure accurate cross comparison of expansions...
A study of the Huntington's disease associated trinucleotide repeat in the Scottish population

1005

observed in different laboratories will be critically important, especially if the overlap of the size distributions is to be validated.

THE NORMAL AND HD RANGES OF CAG REPEAT SIZES IN SCOTLAND

The CAG repeat sizes were determined for 337 affected subjects from 145 Scottish families. Our HD families included 38 affected subjects from Aberdeenshire who are also described by Simpson et al in this issue. Our control group consisted of 229 normal subjects including 144 spouses from HD families. This gave a total number of normal alleles of 795 including the normal alleles from the affected subjects. The distribution of the normal and HD ranges are summarised in table 1 and shown graphically in fig 6.

The peaks of the normal and HD ranges are well separated at 16 and 41 repeats respectively and we have no overlap of the two distributions. The 25 discrete normal alleles range from eight to 33 repeats and the 24 discrete HD alleles range from 35 to 62 repeats. In our set of patients clinically diagnosed with HD we found 12 cases that clearly had two alleles well within the normal range (<25 repeats). These 12 were excluded from all pooled data. They included 11 misdiagnoses (Alzheimer’s disease, multi-infarct dementia, and Parkinson’s disease). We have only one patient with two normal alleles and a clinically confirmed diagnosis of HD. We decided to exclude this patient from the analysis until we had confirmed the allele sizes using a fresh DNA sample.

INTERGENERATIONAL STABILITY OF CAG REPEAT SIZE

In our normal population we have obtained CAG repeat sizes for 65 sets of both parents and their children. We have not seen any change of CAG repeat size in this group or in any other CAG repeat transmission within the normal range. We have two families where a CAG repeat of >30 repeats has been passed on three times with no change in size. When CAG repeat sizes in the HD size range are transmitted we have found a change of repeat size in 72.5% of affected parent, affected child pairs. This variability is also apparent in sibs where the affected parents’ CAG repeat size has not been determined.

THE EFFECT OF PARENTAL ORIGIN OF DISEASE

There have been many reports of an earlier age of onset when HD is of paternal origin.10 We looked at the sizes of the CAG repeats in affected parent, affected child pairs. The change of CAG repeat size upon transmission of the HD chromosome was calculated and the results are shown graphically in fig 7. We have included all of the affected children in the graphical data. The median test was used to test for a difference between the paternal and maternal distributions. A summary of the data used for this test is shown in table 2. In order to avoid a bias for any familial effect we have used affected parent, first affected child pairs only in this calculation. There is a significant positive difference between the paternal and maternal medians, p=0.026. In paternal transmissions the median repeat size increases by one repeat whereas in maternal transmissions the median is unchanged. The range and mean of CAG repeat sizes inherited from the father increased whereas the range and mean repeat sizes inherited from the mother showed little change. This statistical analysis gives unequivocal evidence for increased paternal CAG repeat instability, the overall effect being repeat expansion in paternal transmission.

AGE OF ONSET AND CAG REPEAT SIZE

We have found it impossible to obtain accurate age at onset information for many of our patients from their clinical records. If there is a correlation between age of onset of symptoms and CAG repeat size our poor age at onset data would only confuse the issue. We can, however, make a number of general observations. We see a definite link between very large expansion sizes and early onset of symptoms. All subjects with greater than 53 CAG repeats have ages of onset in their mid twenties or earlier in our population. We do not find the absolute CAG repeat size alone is a good clinical index for age of onset in affected subjects because we have found many subjects with 39 to 42 CAG repeat size range with ages of onset greater than 30 years apart. A particularly striking example of age of onset not being caused by CAG repeat length alone is that of two affected sibs with identical repeat sizes on both the normal and the paternally inherited affected chromosomes having ages of onset differing by at least 15 years. We have several similar families where sibs with identical hap-
lotypes across the HD region have ages of onset differing by up to 20 years.

**Discussion**

The recent identification of an unstable trinucleotide repeat as the causative mutation for HD allows us to address some of the unusual features of this disease. We did not find any CAG repeat number clearly in the HD range in our set of 795 normal chromosomes and neither did we find any evidence for an overlap between the normal and HD size distributions. It may be significant that we have not found anyone with a CAG repeat size of 34 in between the two distributions, but this may merely reflect the rarity of this allele size. We have discussed the question of accurate CAG repeat size determination and hope that the question of the variability in the (CAG)n rich region immediately 3' to the (CAG)n repeat will quickly be addressed by other laboratories to help clarify the degree of overlap between the two distributions. The PCR assay that we use gives the most reliable resolution and size determinations. Labelling of one strand only either radioactively or fluorescently is beneficial whenever absolute size is of critical importance.

Reappraisal of the 12 cases where patients who had been clinically diagnosed as affected but had two repeat sizes well within the normal range showed 11 misdiagnoses. We have one subject whose clinically affected status and allele repeat sizes in the normal range have been confirmed. There is no strong evidence for a family history in this case. This one subject from a sample of 338 affected persons is strong evidence that heterogeneity in the mutational events leading to HD must be very

---

**Table 1** Frequency of CAG repeat sizes on normal and HD chromosomes.

<table>
<thead>
<tr>
<th>Number of CAG repeats</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal chromosomes</td>
<td></td>
</tr>
<tr>
<td>8 to 13</td>
<td>0.033</td>
</tr>
<tr>
<td>14 to 18</td>
<td>0.079</td>
</tr>
<tr>
<td>19 to 23</td>
<td>0.189</td>
</tr>
<tr>
<td>24 to 28</td>
<td>0.054</td>
</tr>
<tr>
<td>29 to 33</td>
<td>0.015</td>
</tr>
<tr>
<td>Huntington's chromosomes</td>
<td></td>
</tr>
<tr>
<td>35 to 39</td>
<td>0.113</td>
</tr>
<tr>
<td>40 to 44</td>
<td>0.031</td>
</tr>
<tr>
<td>45 to 49</td>
<td>0.017</td>
</tr>
<tr>
<td>50 to 54</td>
<td>0.045</td>
</tr>
<tr>
<td>55 to 59</td>
<td>0.018</td>
</tr>
<tr>
<td>60 to 64</td>
<td>0.006</td>
</tr>
</tbody>
</table>

---

**Figure 5** Fluorescent PCR products resolved using a Pharmacia automatic DNA sequencer. The PCR products were resolved on a 6% denaturing polyacrylamide gel and the fluorescent profiles were collected. Only one of the primers was fluorescently labelled so the detector only sees one strand. The analysis software scales the peaks relative to the strongest signal. The cosmid clone L191F1 was run as a control. This clone has been sequenced and contains 18 CAG repeats. The individual shadow bands every 3 bp are also seen with this system and scoring the peak sizes relative to the constant band (not shown) is straightforward. We score the highest peak as the CAG repeat size.

**Figure 6** Size distribution of the CAG repeat on normal and Huntington's chromosomes in Scotland.
low and that 'HD-like' neurological conditions must be extremely rare.

Variability of the age at onset of symptoms has long been a cause for speculation with suggestions that different mutations could be responsible for the juvenile form. We, like others, have found only the CAG repeat expansion as a cause of HD in our population. We have no true juvenile onset cases but have clear evidence that early age of onset is associated with expansions greater than 53 repeats. We do not believe that CAG repeat size alone will prove to be a good clinical index for age of onset as we have observed a high degree of variability in the ages of onset in the most frequently observed HD repeat size range.

The list of diseases caused by a triplet repeat expansion is ever increasing. Evidence for a linkage with juvenile Huntington's disease has been observed for myotonic dystrophy, fragile X syndrome, spinocerebellar ataxia type 1, and FRAXE.11-14 We were able to show a significant increase of repeat size for paternal transmission of the disease and greater instability for paternally transmitted CAG repeats. This effect in its own right is largely responsible for the anticipation and earlier age of onset observed in paternally transmitted disease. There are a number of reports of the HD normal allele or a closely linked gene influencing the age of onset of HD.10,15 Interestingly, Snell et al10 observed a normal allele effect only when the normal chromosome is inherited from the mother whereas Farrer et al15 found a normal allele or closely related gene effect that is independent of the sex of the parent. We feel that the observation in our population of a number of sib pairs with identical genotypes including CAG repeat sizes across the HD region with greater than 15 years difference between the ages of onset is strongly suggestive that another mechanism or unlinked locus needs to be invoked that somehow moderates age of onset.

Finally it is of the greatest importance that accepted international guidelines and counselling protocols should still be adhered to when direct repeat size determination is being used. Of particular importance, given the ease with which a PCR assay can be performed, is that samples of those at risk should not be inadvertently tested without the patients' consent. Until the issue of a small overlap between the HD and normal size distributions has been satisfactorily resolved caution should be observed when interpreting the significance of repeat sizes falling in this interval.

We would like to thank all the clinicians who have sent us samples over the years, Brian Smith, the Huntington's family care officer for Glasgow, Alison Fordyce from the MRC Human Genetics Unit, Aileen Croizzie for help with clinical records, and Alison Rae for running the automatic DNA sequencer.