Analysis of triplet repeats in the huntingtin gene in Japanese families affected with Huntington’s disease

Naoki Masuda, Jun Goto, Nobuhiro Murayama, Masahiko Watanabe, Ikuko Kondo, Ichiro Kanazawa

Abstract
Huntington’s disease (HD) is associated with the expansion of a CAG repeat in the huntingtin gene. Molecular analysis of the repeat in Japanese HD patients and normal controls was performed. The size of the CAG repeat ranged from 37 to 95 repeats in affected subjects and from seven to 29 in normal controls. A significant correlation was found between the age of onset and the CAG expansion. The length of the expanded repeat is unstable in meiotic transmission and large increases occur in paternal transmission. At the same time the CCG repeat polymorphism adjacent to the CAG repeat was analysed and haplotypes of HD chromosomes were identified. Strong linkage disequilibrium was found between the CAG repeat expansion and an allele of (CCG)\textsubscript{10} in Japanese HD chromosomes. It is distinct from that described previously in western populations. Western HD chromosomes strongly associate with an allele of (CCG). Possible mechanisms underlying the disequilibrium in Japan are discussed.

PCR ANALYSIS
PCR amplification of the triplet repeat sequence in the huntingtin gene was performed using the primers HD-1, HD-2, HD-3, and HD-4F, which have been described previously.18 The primers HD-1 (ATGAA-GGCCTTGCAGTTCCCTCAAGTCCTTC) and HD-2 (AAACTCAAGTGCTGGTGC) produce a product containing the CAG repeat and the GC rich region, which includes the polymorphic CCG repeat. The primers HD-3 (GCCGGTGGCGCT-GTTGGCTGCTGCTGCTGG) and HD-4F (GCAGCAGCACCAAGCCGCACCGCCGCACCGCCGCACCGCCG) are complementary and are at the boundary between the CAG repeat and the CCG repeat; the primers HD-1 and HD-3 amplify only the CAG repeat sequence, and the primers HD-2 and HD-4F amplify the GC rich region. These three PCR amplifications were performed for all DNA samples analysed. The amplified alleles were separated on a 5% polyacrylamide gel containing 7 mol/l urea. Allele sizes were determined by comparison with the sizes of the amplified products of pHE5A and pHE5Na and M13mp18 sequence ladders. pHE5A and pHE5Na are the plasmids constructed as follows. Genomic DNA from a
Huntington's disease patient was amplified by PCR using primers HD-1 and HD-2. The amplified products were separated on an agarose gel and the expanded allele product and the normal sized product were purified and subcloned separately into the Smal site of pBluescriptII SK+, designated pHE5A and pHE5Na, respectively. The triple repeat sequence in pHE5A and pHE5Na were of the form TTC(CAG)$_n$CAAACAGCCGCCA (CCG)$_m$CCT and TTC(CAG)$_n$(CAACAGCGCCCA(CCG)$_m$CCT, respectively.

HAPLOTYPING
The haplotype of each chromosome studied was determined by comparison of the lengths of products of the three sets of PCR amplification, and was deduced and confirmed by phase determination where two or three generations were available for molecular analysis.

Results
DISTRIBUTION OF CAG REPEAT NUMBER
Our linkage studies showed that Japanese HD families were tightly linked to chromosome 4p16.3. All 20 HD families previously studied were analysed for determination of number of the CAG repeat. A typical autoradiograph of one of the families is shown in fig. 1. It was confirmed that all affected subjects with HD show expansion of the CAG repeat (data not shown).

Additional Japanese HD families that were not studied for linkage analysis were also analysed. Molecular data were obtained from 92 affected subjects from 58 unrelated HD families. A control group consisted of 185 unrelated non-HD chromosomes which were identified in the families studied. The distribution of normal and abnormal CAG repeat number is shown in fig. 2. The normal range varied from seven to 29 repeats (mode 17 repeats) and the observed heterozygosity and PIC in the normal Japanese population are 0.742 and 0.706, respectively. The HD range is from 37 to 95 and the mode is 42. The peaks of the normal and HD ranges are well separated and the two distributions do not overlap.

AGE OF ONSET AND CAG REPEAT LENGTH
Fig 3 shows the relationship between age of onset and the CAG repeat size in both alleles in affected subjects. A significant negative correlation (n=77, r=-0.709, p<0.001) was evident for the relationship between the age of onset and the number of longer CAG repeats. The regression curve was derived according to the formula (CAG repeat number) = $112 - 17.72 \times \ln$ (age of onset).

Data on age of onset were obtained from 22 father–child pairs and five mother–child pairs in this cohort. The Wilcoxon signed rank test showed that children with paternal transmission developed symptoms significantly earlier than their fathers (p=0.0001), whereas differences between mother–child pairs were not significant (p=0.124). The mean difference between father–child pairs was 20-23 (SE 2.69) years, and 12-20 (SE 6.29) years between mother–child pairs. Molecular data of CAG expansion were obtained from 20 father–child pairs and 14 mother–child pairs. The difference in CAG repeat size on transmission of the HD chromosome was calculated and the results are shown in fig. 4. On paternal transmission the repeat size increased by 4-05 (SE 1.98) repeats on average, whereas the size decreased by 0-36 (SE 0-86) repeats on maternal transmission. The Wilcoxon signed rank test showed that children with paternal transmission had significantly larger expansions than their father (p=0.015), whereas expansion from affected mothers to their children was not significant (p=0.685). This study contained three children with the rigid form of HD and in all cases their HD mutations were inherited from their father.

CGG REPEAT POLYMORPHISM AND HUNTINGTON'S DISEASE
The CGG repeat sequence immediately 3' adjacent to the CAG repeat is polymorphic in white populations. The frequencies of the CGG repeat alleles in the normal Japanese population is shown in the table. Alleles of seven and 10 repeats occur in the majority of...
the Japanese population. The heterozygosity and PIC for the CCG repeat polymorphism in the Japanese population is 0.483 and 0.372, respectively.

Recent studies have shown that more than 90% of HD chromosomes of western descent have the seven repeat allele and strong linkage disequilibrium between this allele and HD is evident. We first analysed the 20 families which had been previously used for linkage analysis. In only three families, the seven repeat allele cosegregated with the HD mutation; however, the 10 repeat allele cosegregated in 17 families. Fig 1 shows a family in which the 10 repeat allele unambiguously cosegregate with the expansion of the CAG repeat. Additional Japanese HD chromosomes were analysed and haplotypes were identified for a total of 58 unrelated Japanese HD chromosomes (table). Forty-nine HD chromosomes (85.5%) cosegregated with the 10 repeat allele, whereas only nine (15.5%) carried the seven repeat allele. The observed allele frequencies for the CCG repeat polymorphism in the HD population are significantly different from those obtained from the normal population ($\chi^2 = 39.39$, df = 2, $p = 0.0001$). This result supports strong linkage disequilibrium between HD and the CCG allele with 10 repeats in the Japanese population.

**HAPLOTYPY ANALYSIS OF THE CAG AND CCG REPEAT POLYMORPHISM IN THE NORMAL JAPANESE POPULATION**

Fig 5 shows the distributions of CAG repeat length of normal chromosomes which have a different CCG repeat allele. The mean CAG repeat size is 17-45 (SE 0.13) for (CCG) and chromosomes and 17-65 (SE 0.13) for (CCG)$_n$ chromosomes. The latter is slightly larger than the former, although the difference is not significantly statistically ($t$ value = 0.7, $p = 0.4839$) and the two distributions are indistinguishable ($\chi^2 = 15.78$, df = 12, $p = 0.2013$).

The primers HD-1 and -2, which have been previously described, amplify the sequence containing both CAG and CCG repeats and the products show combined polymorphism. Heterozygosity and PIC are calculated as 0.846 and 0.829, respectively, and the distribution of alleles have small twin peaks (data not shown), which result from the two predominant alleles of the CCG repeat.

**Discussion**

**EXPANSION OF THE CAG REPEAT**

Analysis of the CAG repeat showed that the range of expansion of the repeat was between 37 and 95 (mean = 42), that of normal chromosomes was between seven and 29, and alleles...
containing repeats ranging from 30 to 36 were not found. The distribution of the expanded repeat of HD was similar to that for North American and European populations described previously. It is difficult to define precisely the smallest size of the CAG repeat expansion which causes the disorder. In most studies numbers around 37 were reported as the lowest numbers for HD of European descent: 35 for the Scottish population, 36 for American, 37 for Italian, 38 for Canadian, 39 for German, 39 for Danish, and 41 for Russian. A significant correlation was found for the relationship between the age of onset and the CAG repeat expansion. The longer the CAG repeat, the younger the age of onset. A similar relationship was described for the HD population of white descent. The inverse relationship is clear for juvenile onset HD, but for adult onset cases any given number of repeats can be associated with an age of onset that varies by decades.

We obtained molecular data from 20 HD father–child pairs and 14 mother–child pairs affected with HD. A significant increase of CAG repeat size on paternal transmission of HD and greater instability of paternally transmitted repeats has been shown in western HD families. Analysis of Japanese affected parent–offspring pairs in this study also showed that on paternal transmission there appeared to be a younger age of onset together with the longer CAG repeat, but not on maternal transmission. Anticipation corresponds to expansion of the CAG repeat and it occurs in paternal transmission.

The above features of the CAG repeat are mostly similar to those of western populations.

THE CCG REPEAT POLYMORPHISM

The CCG repeat polymorphism is composed of the two predominant alleles (CCG), and (CCG), and some other minor alleles in Japanese or western populations. Frequencies of both major alleles in the Japanese population were not significantly different from those in western populations which have been described previously. Other minor alleles are rare in Japan, compared with data from western or black people, and we identified only one chromosome with the (CCG) allele in this study.

The prevalence of HD varies considerably among different ethnic groups. The frequency of the disease is very low in Japan and less than one-tenth of the prevalence in western countries. It was suggested that the low frequency of HD reflects the origins of the HD mutation. The original HD gene was considered to be western European in origin and subsequently was thought to have spread to Japan by emigration. Haplotype analysis of HD chromosomes provides a clue to the origins of the HD gene. It is interesting to note that there is a striking discrepancy between Japanese and western HD, in terms of a CCG repeat allele showing strong association with the CAG repeat expansion. In fact, strong linkage disequilibrium is found between Japanese HD chromosomes and (CCG), whereas western HD chromosomes are strongly associated with (CCG). The most plausible explanation for this discrepancy is that HD mutations in these two races originated from different ancestral lineages; the western HD mutation is mostly derived from a chromosome containing (CCG), in western Europe, while Japanese HD is mostly from a (CCG), associated chromosome in the ancestors of the Japanese population. Although the mutation rate for HD was estimated to be extremely low, recent molecular genetic findings in HD suggest that new mutation events appear to occur more frequently than considered previously. Two groups described several new mutations for HD that were shown by molecular analysis of expansion of a CAG repeat. In addition, analysis of HD chromosomes with multiallelic markers was reported to show many different haplotypes, suggesting a variety of independent HD mutations. Therefore, in addition to the hypothesis of spread of the HD gene from western Europe as the origin of Japanese HD, it is reasonable to postulate that a new mutational event in the Japanese ancestral population might contribute to the present HD patients in Japan.

In addition to the founder effect, we can consider other hypotheses to account for the disequilibrium of the CCG repeat polymorphism. First, one could assume that a cis-acting element links to the CCG repeat and predisposes to expansion of the CAG repeat. The CCG repeat sequence itself can be considered to have an effect on the CAG repeat expansion. Although it has been proposed that the allele could be critical in HD, Japanese HD chromosomes associate strongly with the allele. Therefore, it appears unlikely that a particular allele of the CCG repeat polymorphism might be involved in the mechanism of the disease. Another possibility is that relatively longer CAG repeat alleles in the normal range could be prone to mutate to expansion. The mean length of the CAG repeat of chromosomes containing (CCG), was slightly longer than that of chromosomes carrying (CCG), although the difference was not significant. Longer alleles of the CAG repeat polymorphism, which constitute a very small fraction in our cohort, could be mostly associated with an allele of (CCG), in Japan and might constitute a reservoir providing premutations. Finally, one could speculate that the threshold of CAG repeat size to predispose to expansion may vary in a chromosome having a different sized allele of the CCG polymorphism. The threshold in a chromosome carrying (CCG), might be smaller than in a (CCG), chromosome. However, the distributions of CAG repeat size appear to be similar for both (CCG) and (CCG) chromosomes (data not shown), though we have little molecular data about HD chromosomes carrying (CCG).
Analysis of triplet repeats in the huntingtin gene in Japanese families affected with Huntington's disease