Trinucleotide Repeat Analysis of Huntington's Disease gene in Singapore

H Y Law,*[D Phil], I S L Ng,**[FAMS, MBBS, M Med (Paed)], C S Yoon,***[BSc], Zhao Yi,†[PhD], M C Wong,‡[FAMS, Dip Amer Bd Neur, FRCP]

Abstract

***Introduction**: Huntington's disease (HD) is an inherited neurodegenerative disorder characterised by chorea and progressive dementia. The mutation causing the disease has been identified as an unstable expansion of a trinucleotide $(CAG)_n$. We have assessed the $(CAG)_n$ repeats in the patients and controls in our population. **Materials and Methods**: Polymerase chain reactions (PCRs) for the repeat region were carried out for 116 individuals: 10 were asymptomatic at-risk members from 5 families; 53 symptomatic patients from various hospitals; and 53 normal unrelated Singaporeans. Estimation of the number of repeats was based on Metaphor gel electrophoresis, sizing using the GeneScan on ABI 310 Genetic Analyzer, and sequencing using the same equipment. **Results**: Metaphor gel sizing generally gives an over-estimation, and GeneScan gives an under-estimation of repeat numbers compared with sequencing which is the gold standard. Of the 63 patients and family members tested, 25 had one expanded allele of 40 to 54 CAG repeats and the other allele in the normal range of 15 to 30 repeats. One patient had an allele in the intermediate range (38). **Conclusion**: The range of CAG repeats in the normal and HD alleles in our population is similar to those reported elsewhere. An accurate sizing can only be obtained with sequencing. For allele sizes in the intermediate range (37-40), sequencing should be carried out to confirm the carrier status of a patient.*

***Ann Acad Med Singapore* 2001; 30:122-7**

Key words: CAG repeats, Polymerase chain reaction, Polymorphic, Presymptomatic diagnosis, Sequencing

Introduction

Huntington's disease (HD) is an autosomal dominantly inherited disease, affecting 1 in 10,000 individuals of European origin.[1] The incidence in the Asian population is much lower,[2] and the estimated prevalence in Singapore is 3 to 15 per million.[3] Symptoms of HD include adult onset of chorea and dementia. There is presently no effective treatment and the disease is completely penetrant and progressive, resulting in death typically about 15 to 20 years after onset. As the onset of the disease is usually after reproductive age, many patients have unknowingly passed the defective gene to their offspring.

The gene responsible for HD was localised to chromosome 4 by Gusella and colleagues in 1983.[4] However, it was not until 10 years later that the gene was identified at the distal short arm of chromosome 4 at 4p16.3 in 1993.[5] The HD gene has been found to contain a 5' trinucleotide CAG repeat which is highly polymorphic in the normal population but expanded and unstable in the HD chromosome.[5,6] The CAG repeat length is inversely correlated with the age of onset of the disorder.[7] The expanded repeats are unstable, and often increase through generations particularly by paternal transmissions. This accounts for the tendency of earlier onset with succeeding generations often observed in HD pedigrees. This phenomenon is known as anticipation. The discovery of HD gene thus facilitates highly accurate diagnostic and presymptomatic testing by direct analysis of the CAG repeats.[8]

In this study, the number of CAG repeats in the HD gene was estimated for patients at-risk or suspected to have HD, and for normal unrelated random individuals. All samples were subjected to polymerase chain reaction (PCR) using

* *Principal Scientitific Officer*
** *Senior Consultant and Head*
*** *Laboratory Technologist*
Genetics Service
KK Women's and Children's Hospital
† *Research Scientist*
Department of Clinical Research
Singapore General Hospital
‡ *Senior Consultant and Head*
Department of Neurology
Singapore General Hospital
Address for Reprints: Dr H Y Law, DNA Diagnostic and Research Laboratory, Genetics Service, Department of Paediatric Medicine, KK Women's and Children's Hospital, 100 Bukit Timah Road, Singapore 229899.
E-mail: hylaw@kkh.com.sg

primers flanking the CAG repeats, followed by direct visualisation of ethidium bromide-stained PCR products on Metaphor gel and GeneScan sizing and sequencing.

Materials and Methods

Patients and Controls

A total of 116 individuals were analysed: 53 were neurologically normal unrelated Chinese individuals and 63 were either symptomatic or asymptomatic at-risk patients. The 10 asymptomatic at-risk patients comprised 6 members of an Indian family with several affected siblings reported previously,[3] 2 unrelated Indian patients whose parent/sibling were diagnosed positive in the present study and 2 unrelated Caucasian patients with a positive family history. The symptomatic group comprised 53 outpatients (41 Chinese, 7 Indians and 5 Malays) suspected to have HD. This included 4 symptomatic members from the 2 Indian families mentioned above. Twenty patients were seen in the neurology clinics in Singapore General Hospital, 20 were from Tan Tock Seng Hospital, 6 from Changi General Hospital, 3 from Woodbridge Hospital, 2 from KK Women's and Children's Hospital, one each from National University Hospital and Alexandra Hospital. Collection of patient samples started in June 1993.

PCR Amplification

The CAG repeats at the 5' end of the HD gene were estimated using a PCR method which amplifies only the CAG repeats independent of the adjacent CCG repeats 3' to it. This is important as the CCG region is known to be polymorphic.[9,10] Inclusion of CCG repeats will thus affect the accuracy of CAG repeats estimation. Figure 1 shows the orientation of the CAG and CCG repeats and the primers used in this study. The PCR was carried out in a 50 µl reaction volume with 500 ng of template DNA, 1.5 mmol/L $MgCl_2$, 50 mmol/L KCl, 10 mmol/L Tris pH 8.3, 0.2 mmol/L each of dATP, dCTP, dGTP and dTTP, 0.4 µmol of each primer, HD3 and HD447,[10] and 0.005 units *Taq* polymerase (Roche) using a Strategene Robocyler or a Perkin Elmer 9700 thermal cycler. All the components were heated at 95°C for 5 min before the *Taq* and dNTPs were added to the reaction mixture. A total of 35 cycles were carried out using the cycling condition: denaturation at 95ºC for 30 sec, annealing at 62ºC for 30 sec and extension at 72ºC for 30 sec, followed by a further extension at 72ºC for 10 min. PCR products were separated on 2.5% to 3% Metaphor gels.

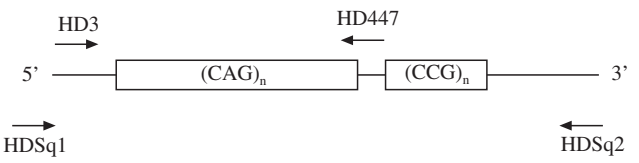

Fig. 1. The orientation of the CAG and CCG repeats at the 5' end of the HD gene and the position of the primers used for the study.

GeneScan Sizing

PCRs were carried out as above using the FAM labelled 447 primer. To size the PCR product, 1 µl of it was diluted with 5 µl of water, and 1 µl of the diluted PCR product was added to 12 µl of formamide and 0.5 µl of marker (GeneScan 500 TAMRA). The mixture was denatured at 95ºC for 5 min, cooled on ice for a further 5 min before loading onto the autosampler tray of the ABI 310 Genetic Analyzer. Electrophoresis was carried out using the GeneScan analysis software which uses POP4 as the matrix.

Sequencing

To generate sequencing template, PCR was carried out using primers HDSq1 (AGACCGCCAT GGCGACCCTGGA) and HDSq2 (TGGTCGGTGCAG CGGCTCCTCA) in a 50 µl reaction with 500 ng of template DNA, 1.5 mmol/L $MgCl_2$, 20 mmol/L $(NH_4)_2SO_4$, 7.5 mmol/L Tris pH 8.8, 0.01% Tween 20, 8% DMSO, 0.2 mmol/L each of dATP, dCTP, dGTP and dTTP, 0.4 µmol of each primer, and 0.005 units of *Taq* polymerase (Fermentas) using a Perkin Elmer 9700 thermal cycler. The PCR condition involved initial denaturation at 95ºC for 5 min, followed by 35 cycles of 98ºC denaturation for 30 sec, annealing at 64ºC for 45 sec, extension at 72ºC for 1 min, and a final extension at 72ºC for 10 min at the end of the cycle. PCR products were run on a 2% agarose gel such that the 2 alleles could be separated and the fragment in the gel can be excised and purified using Qiagen gel extraction kit. The purified PCR product was eluted in 30 µl of elution buffer and 4.5 µl of which was used for sequencing with the BigDye Terminator cycle sequencing Ready Reaction Kit according to the protocols, except that the denaturation was carried out at 98ºC. Electrophoresis of each sample was carried out for 30 min in the ABI 310 Genetic Analyzer using POP6 as the matrix.

Results

Over a period of 7 years (1993 to 2000), 63 patient samples were collected and analysed. All samples, including controls, were analysed on Metaphor gel after PCR using primers HD3 and HD447. Figure 2 shows representative electrophoresis of PCR products on a 2.7% Metaphor gel. Analysis of the PCR fragment sizes was also carried out for all except 10 patients (Fig. 4a). Sequencing of 23 samples showing expanded repeats using the above 2 methods were carried out for confirmation of the actual number of CAG repeats (Fig. 4b).

The analysis of CAG repeats for all patients tested is summarised in Table I. Of the 10 asymptomatic patients

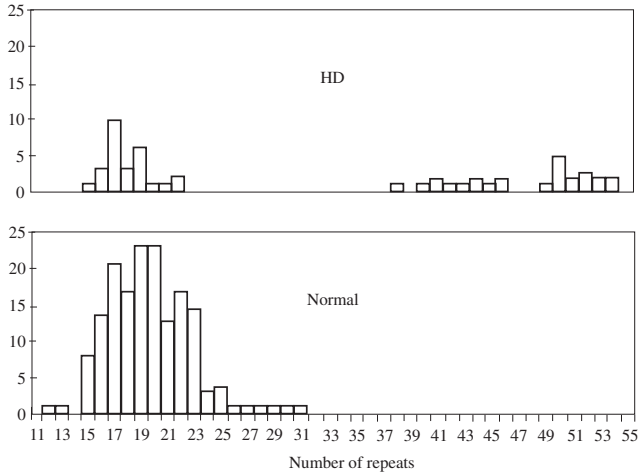

Fig. 2. Electrophoresis of PCR products amplified with HD3 and HD447 on 2.7% Metaphor gel. A product of 110bp is estimated to have 17 CAG repeats. Both M1 and M2 are size markers. M1: ϕX174 HaeIII *digest; M2: 20bp ladder (BioRad). B: Blank control. Lanes 1 and 2 show samples with expanded repeats and lane 3 a normal control. The estimated fragment sizes and repeat numbers are: Lane 1: 107 and 170bp or 22 and 43 repeats; Lane 2: 95 and 176bp or 18 and 45 repeats; Lane 3: 98 and 110bp or 19 and 23 repeats.*

Fig. 3. Distribution of CAG repeats sizes of HD patients and normal controls.

from 5 separate families, 3 were found to have expanded repeats and 1 with repeat size in the intermediate range (38) which was confirmed by sequencing. Of the large Indian family, only 1 asymptomatic member out of the 6 tested was identified as carrying the HD gene. Of those not found to carry the HD gene, 4 were heterozygous (repeat sizes were 18/20, 18/20, 17/21 and 16/22) and 1, a 7-year-old child, was homozygous for 19 repeats. As the non-affected parent of this child was not tested, the chance that this child might have a repeat larger than 54, the largest repeat identified in our study, cannot be ruled out. The positively diagnosed cousin who was tested at the same age, had 53 repeats. The ages of the 53 patients who presented with clinical symptoms ranged from 7 to 84 years. The oldest patient was diagnosed by CAG repeats at 67 years of age. However, it was not certain how long the patient has been symptomatic.

Figure 3 shows the range of $(CAG)_n$ repeats in 26 patients who were found to carry at least one expanded repeat allele (38 or more) and in 90 individuals who were found to carry only normal-sized repeats. These 90 individuals included 53 normal Chinese controls, and 37 symptomatic patients (25 Chinese, 7 Indians, 5 Malays) who were not found to carry expanded repeats. In the HD group, the smallest repeat region in the HD range was found in a Chinese (41 repeats) and 2 Indian patients (one with 40 repeats, 1 with 41 repeats). These repeat sizes were confirmed by sequencing (Fig. 4b). The largest expanded repeat region was 54, found in a 43-year-old Indian lady with a positive family history and a 32-year-old Chinese lady whose

TABLE I: ANALYSIS OF CAG REPEATS FOR ASYMPTOMATIC AT-RISK PATIENTS AND SYMPTOMATIC PATIENTS SUSPECTED TO HAVE HD†

Patient group	Ethnic group	Range of $(CAG)_n$			Subtotal	Total
		Normal (n ≤37)	Intermediate (n = 38)	HD (n ≥40)		
Asymptomatic at-risk	Indian*	6	0	2	8	10
	Caucasian	0	1	1	2	
Symptomatic	Chinese	25	0	16	41	
	Indian	1	0	6**	7	53
	Malay	5	0	0	5	
Total		37	1	25		63

† The number of repeats was based on sequencing results for all alleles with more than 30 repeats. Sequencing was not carried out for samples with repeats of 30 or less.

* Includes 6 asymptomatic members from the large Indian family with multiple affected individuals and an asymptomatic daughter of a symptomatic mother who was diagnosed to carry a HD allele in the present study.

** Includes 3 symptomatic members from the same family.

Fig. 4a. ABI 310 electropherogram of PCR product from a patient with expanded CAG repeats. The two alleles were estimated to be 86bp and 163bp in size or 15 and 41 repeats for peak A and B respectively. b) Electropherogram of a sequencing reaction showing 40 CAG repeats in the expanded allele of a HD patient. HDSq2 was used as the sequencing primer.

family history was not available. The normal alleles in this group ranged from 15 to 22 repeats. This is comparable to the normal group of individuals where more than 82% (137 out of 166) of the alleles were in this range. The smallest number of repeats in the normal range was 12. The largest normal alleles were found in a Chinese control (31 repeats) and a Chinese patient (30 repeats) whose other allele was also in the normal range (17 repeats). No allele was found in the intermediate size range of 32 to 37 repeats.

To determine which sizing method gives more accurate results, 37 alleles were sequenced and the results compared with Metaphor and GeneScan sizing. Figure 5 shows that both methods gave inaccurate allele sizes. Generally, Metaphor gel analysis gives an over-estimation and GeneScan an under-estimation. The discrepancy increased

Fig. 5. Comparison of CAG repeats sizes in 37 alleles using Metaphor gel electrophoresis estimation, GeneScan analysis and sequencing.

with allele size, especially for the GeneScan analysis. Thus, accurate sizing can only be obtained by sequencing the relevant fragment to assess the actual number of CAG repeats.

Discussion

Earlier reports of CAG repeats in the HD gene have been based on PCR products generated by the traditional method of amplifying both the CAG and CCG repeats.[4,10] The number of CAG repeats was estimated by subtraction, assuming that CCG repeats were invariably 7 copies.[11] Since it has been shown that CCG repeats are polymorphic, ranging from 7 to 12 triplets,[8] CAG repeats have been assessed independent of the CCG repeats, especially when the CAG repeats are in the upper limit of the normal size range and the lower limit of the HD range. Our study, based on the sizing of CAG repeats independent of CCG repeats, has shown a size range (12-31 repeats) of normal alleles consistent with reports in Chinese[2] and other populations.[7,11,12] All the expanded CAG repeats in the reported range for HD (36 repeats or more[12]) were found in the patient group which consisted of individuals with clinical symptoms of HD, or with family history but without clinical symptoms. The lower range of these expanded repeats (38, 40 and 41) which was confirmed by sequencing, is well above 36 repeats, the smallest number of repeats reported in HD;[13] although the individuals carrying them are not affected. These expanded alleles are reported to be unstable and possibly cause HD in the next generation. The patient who was asymptomatic and found to carry 38 repeats had a symptomatic younger sibling with 50 repeats. As both their parents were healthy and lived to an old age, it is possible that the larger allele carried in this family became unstable and expanded to the disease range in the next generation. The smallest normal parental allele reported to have expanded into the HD range in an offspring with clinical symptoms was 27 repeats.[14] While most studies have shown an allele size of 36[12] or more repeats to be associated with HD, normal elderly individuals with 36 to 39 repeats have been reported which might define a range of incomplete penetrance of HD gene mutation.[15]

Although we have not attempted to assess the number of CCG repeats in our study population, sequencing of normal and expanded regions of 37 alleles has shown the adjacent CCG to be mostly 7 repeats, though some were found to have 10 repeats. The 7 CGG repeats were found in more than 90% of HD chromosomes, but in only 67% of normal chromosomes.[8] While our study population is too small to come to any conclusion, it would be of interest to find out the frequency of this allele in normal Asian individuals.

While assessment of allele size is most accurate with sequencing, it is time consuming and expensive. Metaphor gel analysis is efficient but subjective and usually results in an over-estimation. GeneScan sizing is objective but gives an under-estimation. This is in agreement with Bruland and colleagues' findings that using the GenesScan 500 TAMRA standard, an under-estimation of 3 repeats[16] was consistently observed. However, we found the discrepancy in size estimation between the GeneScan and sequencing result increased with increasing size of repeats. As screening of 63 patients showed a clear separation of normal range and HD alleles as reported in other population, it is thus important that the repeat numbers in the intermediate size range (37-40) be confirmed by sequencing.

Most of the patients and their family members tested were heterozygous for the triplet repeats, except for 8 of them. As the largest repeat detected in this study was 54, we are unable to conclude if larger repeats could be efficiently amplified with the present protocol. Southern analysis using the HD probe may thus be necessary to confirm the absence of an upper range of repeats when only 1 allele fragment size is seen. Thus, the status of the 7-year-old child from the Indian HD family who was found to be homozygous for 19 repeats needs to be confirmed with southern analysis.

The size of CAG repeats has been reported to be inversely correlated with the age of onset.[7] However, correlation of the size of repeat with clinical symptoms and the age of onset has not been carried out as our sample size is too small, and full clinical histories were not available for all patients. Some might have been symptomatic for many years before molecular testing was available. More data need to be collected to draw any conclusion for our population.

In conclusion, the present study shows that the PCR method described provides an efficient way in assessing CAG repeats in the HD gene. For estimation of repeat size in the intermediate range, sequencing must be carried out to obtain an accurate result. For patients with a positive family history, presymptomatic diagnosis can be carried out but proper counselling must be offered. For patients presenting with clinical symptoms but without family history, assessment of the CAG repeats can be useful in confirmation of the diagnosis of HD.

Acknowledgements

This work was supported by funds from the National Medical Research Council and the Department of Clinical Research, Singapore General Hospital.

REFERENCES

1. Harper P S. The epidemiology of Huntington's disease. Hum Genet 1992; 89:365-76.
2. Chan V, Yu Y L, Chan T P, Yip B, Chang C M, Wong M T, et al. DNA

analysis of Huntington's disease in southern Chinese. J Med Genet 1995; 32:120-4.

3. Wong M C, Ng I S L, Lim S H. Huntington's disease in five siblings. Ann Acad Med Singapore 1993; 22:428-30.
4. Gusella J F, Wexler N S, Conneally P M, Naylor S L, Anderson M A, Tanzi R E, et al. A polymorphic DNA marker genetically linked to Huntington's disease. Nature 1983; 306:234-8.
5. The 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.
6. Zuhlke C, Riess O, Schroder K, Siedlaczck I, Epplen J T, Engel W, et al. Expansion of the (CAG)n repeat causing Huntington's disease in 352 patients of German origin. Hum Mol Genet 1993; 2:1467-9.
7. Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, et al. Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat Genet 1993; 4:387-92.
8. Davis M B, Bateman D, Quinn N P, Marsden C D, Harding A E. Mutation analysis in patients with possible but apparently sporadic Huntington's disease. Lancet 1994; 344:714-7.
9. Rubinsztein D C, Barton D E, Davison B C, Ferguson-Smith M A. Analysis of the Huntington gene reveals a trinucleotide-length polymorphism in the region of the gene that contains two CCG-rich stretches and a correlation between decreased age of onset of Huntington's disease and CAG repeat number. Hum Mol Genet 1993; 2:1713-5.
10. Andrew S E, Goldberg Y P, Theilmann J, Zeisler J, Hayden M R. A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. Hum Mol Genet 1994; 3:65-7.
11. Goldberg Y P, Andrew S E, Clarke L A, Hayden M R. A PCR method for accurate assessment of trinucleotide repeat expansion in Huntington disease. Hum Mol Genet 1993; 2:635-6.
12. Kremer B, Goldberg P, Andrew S E, Theilmann J, Telenius H, Zeisler J, et al. A worldwide study of the Huntington's disease mutation. The sensitivity and specificity of measuring CAG repeats. N Engl J Med 1994; 330:1401-6.
13. Rubinsztein D C, Leggo J, Coles R, Almqvist E, Biancalana V, Cassiman J J, et al. Phenotypic characterization of individuals with 30-40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36-39 repeats. Am J Hum Genet 1996; 59:16-22.
14. Goldberg Y P, Andrew S E, Theilmann J, Kremer B, Squitieri F, Telenius H, et al. Familial predisposition to recurrent mutations causing Huntington's disease: genetic risk to sibs of sporadic cases. J Med Genet 1993; 30:987-90.
15. McGlennen R C, Alinson P S, Matthias Hagen V L, Parker T I, Lovell M A, Kelly T E. Evidence of an unstable paternal 27 CAG repeat allele in the Huntington gene giving rise to clinically overt Huntington's disease in a patient with the genotype (17/48). Am J Hum Genet Suppl 1995; 57:A246.
16. Bruland O, Almqvist E W, Goldberg Y P, Boman H, Hayden M R, Knappskog P M. Accurate determination of the number of CAG repeats in the Huntington disease gene using a sequence-specific internal DNA standard. Clin Genet 1999; 55:198-202.