Trinucleotide Repeats and Haplotypes at the Huntingtin Locus in an Indian Sample Overlaps with European Haplogroup A

Nagaraj S Molly, Lakshmi Narayanan Kota, Ram Murthy Anjanappa, Sowmya Venugopal, Radhika Vaidyanathan, Pramod Pal, Meera Purushottam, Sanjeev Jain, Mahesh Kandasamy


Abstract

Huntington’s disease (HD), an autosomal dominant neurodegenerative syndrome, has a world-wide distribution. An estimated 2.5-10/100,000 people of European ancestry are affected with HD, while the Asian populations have lower prevalence (0.6-3.8/100,000). The epidemiology of HD is not well described in India, and the distribution of the pathogenic CAG expansion, and the associated haplotype, in this population needs to be better understood. This study demonstrates a distribution of CAG repeats, at the HTT locus, comparable to the European population in both normal and HD affected chromosomes. Further, we provide an evidence for similarity of the HD haplotype in Indian sample to the European HD haplogroup.

Introduction

The diagnosis of Huntington’s disease (HD) is based on estimation of the CAG repeat length at the HTT locus. The normal HTT gene contains less than 27 CAG repeats, and a few normal individuals have intermediate CAG(27-35) repeat expansion and display no symptoms suggestive of HD. Subjects with borderline CAG(36-39) repeats may or may not develop symptoms. Individuals affected with HD typically have at least one HTT allele containing CAG repeat size of 40 or greater.

The age at onset (AAO) is inversely correlated with length of the pathogenic CAG stretch in the HTT gene. Almost 50-70% of the variation observed is determined by the CAG repeat length, the remaining maybe explained by the additional influence of other cis and trans elements, as well as environmental factors. Highly expanded CAG sequences cause disease onset at a younger age. The fundamental mechanisms of CAG repeat instability are poorly understood.

The prevalence of HD varies among different populations, with prevalence rates of 2.5 -10 per 100,000 in people of European ancestry, while the Japanese (0.11–0.45 per 100,000), Chinese (0.5-1 per 100,000) and African populations (<0.01 per 100,000) show significantly lower prevalence. Indian and other South Asian populations are expected to have intermediate prevalence of HD. Prevalence studies of Indian immigrants in UK, predominantly from the northern regions of the Indian subcontinent, suggest that HD occurs in 1.75 per 100,000 individuals. It is generally accepted from clinical experience, and family studies of different geographical regions, that HD is distributed widely in India.

The origin of the pathogenic CAG expansion in India is not well understood. Multiple founder effects, and admixture with European populations probably has caused the spread of HD into diverse populations across the world.

Pramanik et al. reported CAG repeat expansions in 28 HD samples from eastern part of India ranging from 41-56 while the range of the normal allele on the homologous chromosome was 13-29 repeats. In a previous study of 117 individuals we reported that the mean CAG repeats on normal alleles was 18±2.4 with a range of 14–30, and in the 26 subjects with HD, the mean CAG repeat length in expanded alleles was 48.4 ± 8.7 with the range of 39 to 82. Using another marker (CCG repeats) adjacent to the CAG repeats in the HTT gene, it has been observed that (CCG)10 repeat size was linked to the majority (84.5%) of expanded alleles in the Japanese populations, while in the Caucasian populations (CCG)7 repeat was associated with the expanded alleles. There was no significant correlation found between CCG repeats and expanded CAG in Indian samples.

A polymorphic ?2642 (GAG insertion/deletion) of the HTT locus indicated that majority of the insertion (i.e. the presence of GAG) is associated with the normal alleles while deletion is associated with the pathological expansions. In addition a
marker D4S127 (CA repeat allele 2-151bp) has significantly been associated with HD patients originated from the southern part of India, while this was not seen in HD patients originated from northern India.

Recent reports suggests that differences in the prevalence of HD may be linked to differences in haplotypes. Warby et al hypothesized that HTT haplogroups (A, B and C) are associated with differences in mutational rates, thereby influencing the prevalence across populations. Haplogroup A is largely seen in HD patients from Europe, while in East Asian populations (China and Japan), HD alleles are associated with haplogroup C. The HTT haplotypes in Indian population are not determined yet. In this study, we describe the genetic characteristics of the HD mutation and the prevalence of the common HTT haplotypes in the Indian population.

Methodology

Enrolment of Clinical Samples:
The clinical samples (N=164) were derived from both outpatients and inpatient referrals to Genetic Counseling and Testing Center (GCAT) at the National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore, India. The healthy control samples (N=103) were recruited by purposive sampling of volunteers from different parts of India (Total, N=267). Written informed consent was obtained after detailed explanation and genetic counseling of all the subjects and family members prior to enrollment into the study. This study was approved by the Institutional Ethics Committee, National Institute of Mental Health and Neurosciences, (NIMHANS) Bangalore, India.

Genotyping of CAG repeats CCG repeats, 2642 and D4S127 polymorphisms at HTT locus:
The genomic DNA was extracted from whole blood using standard protocols. The extracted DNA was subjected to polymerase chain reaction (PCR) using appropriate primers. In patients and normal chromosomes, the CAG repeats, CCG2 2642 and D4S127 polymorphisms were analyzed as described previously. The PCR products were characterized by high resolution agarose gel electrophoresis and fragment analysis was performed using ABI 3500xL genetic analyzer. Repeat sizing was done using Gene Mapper v3.5. PHASE was used to determine the most likely haplotype for the samples.

Analysis of Haplogroup A, B C in HD cases
Of the reported twenty-two predisposing tag SNPs, three SNPs (rs762855 (SNP1), rs3856973 (SNP2) and rs4690073 (SNP3) were selected to distinguish haplogroups A, B and C. The PCR was carried out by designing appropriate sets of primers. The SNP1 (rs762855) was determined by Sanger’s sequencing in ABI 3500xL genetic analyzer. The SNP2 and SNP3 were detected by PCR RFLP using restriction enzymes TaqI and AseI respectively.

Statistical analysis
The mean CAG repeat numbers was compared by Mann-Whitney U test using R software and QTI- plot, and statistical significance was defined as P<0.05. Spearman’s rank correlation and partial correlation coefficients were used for correlation analysis.

Results

Description of the clinical sample:
Of 164 clinical samples, pathological expansion of more than 39 CAG repeats was confirmed in 116 (71%) samples (symptomatic, N=102, and pre-symptomatic, N=14). The remaining samples (N=48, 29%) had choreiform movements, but did not show any CAG expansion at the HTT locus.

Almost half of the confirmed sample showed paternal inheritance (N= 46; 45.1%), while maternal inheritance (N=25; 24.5%) was seen in a quarter. Both parents were affected in two (1.9%) cases. 'Sporadic' mutations without any reported family history were seen in 11 (11%) cases. Inheritance information was not available for 18 cases (17.6%) (Fig 1A).

Distribution of CAG repeats at HTT locus
The control group showed a mean CAG distribution of 17.6±2 (range: 11-34) in the lower allele (LA) and 20.1±3.6 (range: 15-34) in the upper allele (UA) (Fig. 1B,C,D). The symptomatic HD cases (N=102) showed a mean distribution of 17.9±2.8 in LA and 45.9±7.3 in UA (Fig. 1B,C,D). There were 14 pre-symptomatic individuals with expanded alleles. (LA= 17.6±1.7; UA=42.9±3.1; range 36-48).

The distribution of LA among the HD cases and healthy controls (Fig 1B) did not differ significantly (Control:17.6; HD:17.9). The number of CAG repeats in HD sample ranged from 40 to 85 repeats in the UA. Expansions between 40-50 repeats were most common (N=86, 84.3%), with expansions greater than 50 CAG repeats being less frequent (N=16, 15.7%) (Fig1 D).
Fig. 1: Mode of inheritance and distribution of CAG repeats in the study group and Distribution of (CAG)n size of the Huntington disease (HD) gene in 438 chromosomes of the Indian population.

Fig.1A: Mode of inheritance based on detailed clinical interviews of symptomatic HD individuals and family members. Number of individuals (%)

Fig.1B: Distribution of CAG repeats at the lower allele (LA) in healthy controls, HD symptomatic and pre symptomatic individuals.

Fig.1C: Distribution of CAG repeats at the upper allele (UA) in healthy controls, HD symptomatic and pre symptomatic individuals. *** p<0.001

Fig.1D: The CAG distribution in the Indian population shows a pattern of 5 distinct spreads.

The average AAO was 37.6±12 years, with the CAG repeat size correlating inversely with the AAO (r = -0.7 P<0.0001) (Fig.2), and accounted for about 51.4% of the variance at onset. HD individuals with CAG 40-50 repeats had an AAO of 40.2±10.4 years, while those with more than 50 CAG repeats had a much lower AAO (19.5± 5.0) (r = -0.96 P<0.0001). In these subjects, CAG repeats accounted for about 93.3% of the variance at onset (Fig. 2). Notably, the highest CAG repeat size (85 repeats) was seen in a child aged 4 years.
Fig. 2: Inverse correlation of AAO and HD CAG repeat length.

For each individual, the measured CAG repeat length in HTT allele (Y-axis) is plotted against AAO (years) (X-axis). The solid line represents the linear fit to the data. The CAG repeat length accounts for approximately 51.4% of the overall variation in AAO. The dotted line shows the linear fit for repeats >50. The CAG repeat length accounts for approximately 93.3% of the overall variation in AAO.

Comparison of CAG repeats in control and HD cases from different geographical regions of India:

The healthy control group had 88 individuals of southern Indian descent (LA= 17.6±2.1; UA= 19.8±3.5) and 15 individuals from northern and eastern parts of India (LA=18.0±1.3; UA=21.6±3.9). In the symptomatic HD sample (N=102), 75 were of southern Indian origin (LA= 17.8±3.1; UA= 46.5±7.6; AAO 37.8±11.7) and 27 were from the northern and eastern parts of India (LA= 18.3±2.3; UA= 45.8±7.8; AAO 37.2±13.2). The distribution of CAG repeats did not differ in number of CAG repeats, or in the AAO, between these regions.

Distribution of CCG repeats, ?2642 and D4S127 polymorphisms at the HTT locus in controls and HD patients:

The (CCG)7 repeat length was the most frequent followed by the (CCG)10 repeat length regardless of genotype. The (CCG)10 repeat size had a much lower frequency in HD cases compared to controls. The ?2642 marker was found in 90% of the HTT alleles corresponding to ‘In’ (insertion) polymorphism in both cases and controls. In the remaining 10% of subjects, this ?2642 marker was absent (deletion) in 8% of cases compared to 1.7% of controls. Moreover, more than one third of high end normal alleles had this deletion. The D4S127 microsatellite repeat distribution was similar with 5(153bp) being the most common allele in both cases and controls. As parental genotypes could not be ascertained, PHASE was used to ascertain the most probable haplotype for these 3 markers. Both cases and controls showed the 7-In-5(153bp) to be the most likely combination.

Distribution of haplogroup A, B and C in HD patients:

Due to the non availability of all parental DNA samples and genotype information, it was difficult to establish haplotype A, B, C for all samples. However, we found that out of 90 samples with expanded CAG repeats which had genotype data available, those that were homozygous (60%) for these three SNPs, had the AGG haplotype. Thus region wide analysis showed 62.8% of the southern Indian, 60% of the northern Indian and 46.6% of the eastern Indians tested had the haplogroup A.

Discussion

The average number of CAG repeats in the normal individuals is reported to be 17 to 20. The commonest CAG repeat in the present sample is 18, and the mean CAG repeat size is 18.3, which is modestly higher than that found in populations from Africa (16.2), Japan (16.9) and China (17.0); but closer to that found in European (17.8) population suggesting a genetic structure at this locus closer to the European population. Interestingly, the tribal population seems to have the lowest number of CAG repeats, with absence of any repeats greater than 18. This is perhaps consistent with lower repeat sizes observed in older populations (e.g sub-Saharan African populations).

The size of CAG repeats are evolutionarily controlled expansions which might benefit neuro-developmental processes, and may be conserved in the healthy population. The mechanism of transition from normal CAG to pathogenic expanded CAG can be envisaged by stepwise occurrence of intermediate CAG repeats followed by the genetic anticipation which would drive them into the pathological range of > 39 CAG repeats. This is reflected in the large 40-50 CAG repeat group which constitutes the majority (84%) of HD samples with late AAO (40.2 years). Thus the multistep increase in CAG number appears to play a role in CAG expansion in Indian population.

The CAG expansions beyond 50 CAG (N=16) constituted a smaller group, and had, correspondingly, a more severe illness phenotype with an early AAO (19.5 years). This is similar to those reported in other samples. The mechanisms of the CAG expansion, and abnormal folding of mutant Htt leading to neurodegeneration need to be explored further using animal and cellular models. Allele specific associations were not observed for the (CCG)7 or (CCG)10 and ?2642 deletion was found to be over-represented on expanded HTT alleles. The present study validates previous findings. Thus the underlying genetic diversity of HTT alleles needs to be assessed in other Indian populations.

In this sample, based on the three SNPs (rs762855- rs3856973- rs 4690073), haplogroup A was most common which corresponds to the Caucasian HD haplogroup. In addition, individuals from different parts of India shared this haplogroup, indicating that it is not a region specific effect. Sequencing the region further to permit finer mapping of these haplogroups is essential to comment upon the origin and spread of this mutation in India.

In summary, more than a 100 individuals have tested positive for HD at this centre, and this implies almost a 1000 subjects at risk (including siblings and children). The current population of India is estimated to be 1.27 billion. There is no population survey based data on the prevalence of HD from India. Extrapolating from the prevalence reported in south Asian population in
UK; HD prevalence = 15.9 per million), or a more reasonable estimate of around 30/million) it can be speculated that between 20,000-40,000 people, are suspected to be affected with HD in Indian population, and more than 0.2 million may be at risk°. It is suggested that the prevalence of HD in India should not be underestimated, since the occurrence of HD in India was proposed to be higher within the Asian population°. Therefore, establishing an Indian HD registry in a combined effort with diagnostic centers across the country and development of special care services in India needs to be considered. Based on HD associations from western countries, the average cost for genetic confirmation is about 1,000 USD/sample and the average annual cost of the medical service per HD patient is estimated to be 10,500-40,000 USD. Thus a considerable investment in clinical care, as well as family and social care services needs to be resolved in India. Though these are ‘rare’ syndromes, it is essential that we try to understand the genetic and clinical progression of this ‘single’ gene disorder with complex consequences and develop services focused to address the various problems the patients and their families face at different stages.

Corresponding author

Dr Mahesh Kandasamy PhD, Research Scientist, Molecular Genetics Laboratory, Neurobiology Research Centre, Department of Psychiatry, National Institute of Mental Health and Neurosciences (NIMHANS), Hosur Road, Bangalore, 560 029, Karnataka, INDIA. Email: pkmahesh5@gmail.com, Tel. No. +91 80 26995263

Keywords

Huntington's disease, HTT, CAG repeats, Ethnicity, Haplotype, India

Acknowledgements

The authors would like to thank Dr. Uday Muthane, Parkinson’s & Aging Research Foundation, Bangalore, India for his support and Muralidharan J for excellent technical support. The authors would also like to thank all the patients, their family members and volunteers for their participation and cooperation.

References


