RESEARCH ARTICLE

AMERICAN JOURNAL OF medical genetics PART B
Neuropsychiatric Genetics

Haplotype Analysis of the 4p16.3 Region in Portuguese Families With Huntington's Disease

Eliana Marisa Ramos,[1,2] Tammy Gillis,[1] Jayalakshmi S. Mysore,[1] Jong-Min Lee,[1] Martin Gögele,[3] Yuri D'Elia,[3] Irene Pichler,[3] Jorge Sequeiros,[2,6,7] Peter P. Pramstaller,[3,4,5] James F. Gusella,[1] Marcy E. MacDonald,[1] and Isabel Alonso[2,6,7]*

[1]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA

[2]UnIGENe, IBMC–Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal

[3]Center for Biomedicine, European Academy of Bozen/Bolzano (EURAC), Bolzano, Italy

[4]Department of Neurology, Central Hospital, Bolzano, Italy

[5]Department of Neurology, University of Lübeck, Lübeck, Germany

[6]CGPP, IBMC–Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal

[7]ICBAS–Instituto de Ciências Biomédicas Abel Salazar, University of Porto, Porto, Portugal

Manuscript Received: 15 October 2014; Manuscript Accepted: 25 November 2014

Huntington's disease (HD) is a neurodegenerative disorder characterized by involuntary choreic movements, cognitive impairment, and behavioral changes, caused by the expansion of an unstable CAG repeat in *HTT*. We characterized the genetic diversity of the HD mutation by performing an extensive haplotype analysis of ~1Mb region flanking *HTT* in over 300 HD families of Portuguese origin. We observed that haplotype A, marked by *HTT* delta2642, was enriched in HD chromosomes and carried the two largest expansions reported in the Portuguese population. However, the most frequent HD haplotype B carried one of the largest (+12 CAGs) expansions, which resulted in an allele class change to full penetrance. Despite having a normal CAG distribution skewed to the higher end of the range, these two core haplotypes had similar expanded CAG repeat sizes compared to the other major core haplotypes (C and D) and there was no statistical difference in transmitted repeat instability across haplotypes. We observed a diversity of *HTT* region haplotypes in both normal and expanded chromosomes, representative of more than one ancestral chromosome underlying HD in Portugal, where multiple independent events on distinct chromosome 4 haplotypes have given rise to expansion into the pathogenic range. © 2015 Wiley Periodicals, Inc.

Key words: **Haplotype analysis; Huntington's disease; Instability; Genetic diversity; Delta 2642 allele**

How to Cite this Article:
Ramos EM, Gillis T, Mysore JS, Lee J-M, Gögele M, D'Elia Y, Pichler I, Sequeiros J, Pramstaller PP, Gusella JF, MacDonald ME, Alonso I. 2015. Haplotype analysis of the 4p16.3 region in Portuguese families with Huntington's disease.
Am J Med Genet Part B 168B:135–143.

INTRODUCTION

Huntington's disease (HD) is a dominantly inherited neurodegenerative disorder, usually of adult onset, characterized by involuntary choreic movements, cognitive impairment, and behavioral changes. HD is caused by the expansion of an unstable polymorphic CAG repeat in *HTT*, located in chromosome

Conflict of interest: The authors declare no conflict of interest.
Grant sponsor: NINDS; Grant number: NS16367, NS32765; Grant sponsor: Fundação para a Ciência e a Tecnologia; Grant number: SFRH/BD/44335/2008; Grant sponsor: Programa Ciência, POPH - QREN - Tipologia 4.2 - Promocão do Emprego Científico, ESF and MCTES (Ministério da Ciência e Ensino Superior); Grant sponsor: Ministry of Health and Department of Educational Assistance.

*Correspondence to:
Isabel Alonso, UnIGENe and CGPP-IBMC, University of Porto, Rua Campo Alegre 823, Porto 4150-180, Portugal.
Email: ialonso@ibmc.up.pt

Article first published online in Wiley Online Library (wileyonlinelibrary.com)
DOI 10.1002/ajmg.b.32289

4p16.3, [The Huntington's Disease Collaborative Research Group, 1993] with alleles between 35–39 repeats showing incomplete penetrance, while alleles with 40 or more repeats are fully penetrant. Alleles with fewer than 35 CAGs are considered to be normal alleles that produce no symptoms of HD. Age-at-onset (AO) of neurological symptoms is inversely correlated with the number of expanded CAGs; repeat length alone explains about 50–70% of the variance in AO, while the remaining variance is highly heritable. [Duyao et al., 1993] Expanded CAG repeats are prone to further length variation, while most normal size CAG repeats are stably inherited. There are, however, conflicting reports about the degree of instability of high-normal alleles, a subclass of normal alleles with repeats between 27 and 35 CAGs. Several studies have shown that these alleles can also be unstable, [Goldberg et al., 1993; Myers et al., 1993; Semaka et al., 2010] while in the Venezuelan kindred and Portuguese population they were mostly stably inherited. [Brocklebank et al., 2009; Sequeiros et al., 2010]

Several genetic studies in HD families of different ethnic populations have investigated the origins of HD by examining the linkage phase of a few polymorphisms in the *HTT* region, mainly the CCG repeat adjacent to the CAG tract and the deletion of a glutamic acid residue at position 2642 (delta2642). [Squitieri et al., 1994; Almqvist et al., 1995; Hecimovic et al., 2002; Saleem et al., 2003; Paradisi et al., 2008; Norremolle et al., 2009] Genetic diversity in the *HTT* region has been studied in more detail. Warby et al. found that 95% of 268 European disease chromosomes were associated with the same haplogroup (a linear arrangement of polymorphisms that showed broad allelic similarity but contained internal allele mismatches), haplogroup A, which was also significantly enriched on high-normal alleles relatively to the general population. [Warby et al., 2009] They hypothesized that *cis*-acting factors have a major predisposing influence on HD CAG instability. In a subsequent study, they found that in East Asia (China and Japan) most HD chromosomes were associated with a minor HD European haplogroup (haplogroup C). [Warby et al., 2011] More recently, these authors found that HD chromosomes from white and mixed African HD populations were mostly associated with haplogroup A, while in black Africans HD occurred predominantly on haplogroups B and C. [Baine et al., 2013] The European HD haplogroups (A1 and A2) were absent both in the Asian and African general population. The authors argued that different *HTT* haplogroups had a different mutation rate and that geographic differences in haplogroups explain the difference in the worldwide prevalence of HD.

In a different study, Lee et al. identified 40 single nucleotide polymorphisms (SNPs) centered on *HTT* that showed genome-wide significance for association with HD. Their findings revealed that, although the *HTT* region exhibits many different haplotypes (a linear arrangement of polymorphisms that showed allelic identity), one apparent ancestral haplotype marked by the *HTT* delta2642 accounted for about 50% of HD chromosomes. [Lee et al., 2012] However, in this study, neither the extended shared haplotype nor the individual local *HTT* haplotype of disease chromosomes were associated with altered CAG-repeat length distribution or residual AO for motor symptoms. Nevertheless, the major haplotype had been previously associated with sporadic cases of HD in "new mutation" families. [Myers et al., 1993]

A small haplotype study in 140 HD Portuguese families showed three different HD founder haplotypes, associated with 7-, 9-, and 10-CCG repeats, which suggested different origins for the HD mutation in Portugal. [Costa et al., 2006] Following up on these findings, in order to characterize in greater detail the genetic diversity of HD in the Portuguese population and gain more insight into its origins, we performed an extensive haplotype analysis of the region flanking the *HTT* CAG repeat, in over 300 HD families of Portuguese origin.

MATERIALS AND METHODS

Portuguese HD families

DNA samples were collected at CGGP, a reference laboratory for HD testing in Portugal, which receives biological samples from physicians all over the country for confirmation or exclusion of HD, pre-symptomatic and prenatal testing. A total of 334 apparently independent Portuguese HD families, comprising 531 individuals, were used for haplotype studies: 441 individuals had an expanded *HTT* CAG allele (16 with a repeat in the reduced penetrance range) and 90 were unaffected relatives. Informed consent was obtained from all individuals that participated in this study.

Non-HD Families: MICROS Study

The MICROS study is a population-based survey carried out in three small isolated villages of the Val Venosta in Italy. A detailed description of the MICROS study can be found at Pattaro et al. [Pattaro et al., 2007] Briefly, DNA was collected from over 1,400 volunteers and genealogical information was reconstructed by means of church records and municipality lists, going back to the 1600s.

Selection and Marker Genotyping

Selection of the majority of the markers used in this study was based upon previous work from our lab that showed that they were overrepresented on HD chromosomes. [Lee et al., 2012] Genotyping of nine SNPs (rs6814736, rs7667745, rs1751848, rs762847, rs7658462, rs3129317, rs1730768, rs12641989, and rs16844364 - Fig. 1) flanking the CAG repeat was performed by real-time polymerase chain reaction (PCR), using the commercially available Taqman Genotyping probes (Applied Biosystems, Foster City, CA), carried out on the LightCycler® 480 (Roche Diagnostics, Mannheim, Germany), following manufacturer's instructions. Repeat sizes of the *HTT* CAG repeat, CCG repeat and *D4S127* and genotyping of the delta2642 polymorphism were determined using previously established, but slightly modified PCR amplification assays with fluorescently labelled primers. [Taylor et al., 1992; Rubinsztein et al., 1993; Warner et al., 1993; Ambrose et al., 1994] The fragments' size was then determined using the ABI PRISM 3730*xl* automated DNA Sequencer (Applied Biosystems, Foster City, CA) and GeneMapper version 3.7 software. For each marker, a set of sequenced samples was used as standards.

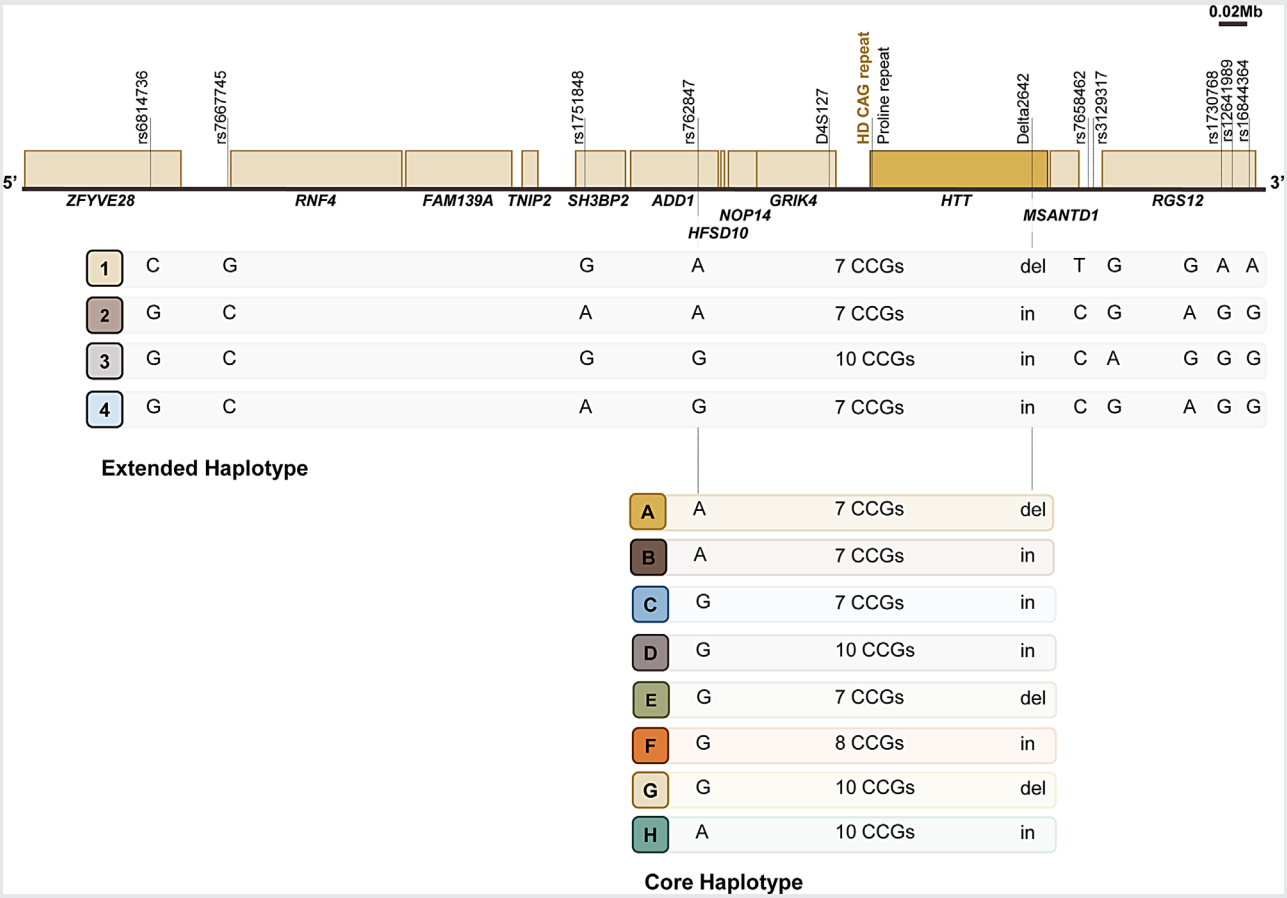

Fig. 1. Schematic map of the polymorphisms used for haplotype reconstruction. A region of 1.17Mb flanking *HTT* is represented, showing the location of nine SNPs highly associated with HD chromosomes (rs6814736, rs7667745, rs1751848, rs762847, rs7658462, rs3129317, rs1730768, rs12641989, and rs16844364) and microsatellite marker (*D4S127*). Within *HTT*, in addition to the CAG repeat itself, the adjacent proline-encoding CCG repeat and delta2642 polymorphisms were also included. The allelic structure of the four most common HD extended haplotypes (1–4) and the eight different core haplotypes (A–H) found on HD chromosomes are shown.

Haplotype Phasing

PHASE software (http://stephenslab.uchicago.edu/software.html) version 2.1 was used to reconstruct haplotypes from genotypic data, when phase could not be directly inferred from family structure. Haplotypes were reconstructed by incorporating the genotypes of SNPs rs6814736, rs7667745, rs1751848, rs762847, rs7658462, rs31293, rs1730768, rs12641989 and rs16844364, as well as the repeat sizes at *D4S127* and the CCG repeat (as a microsatellite *locus*). The genotypes of the *HTT* CAG repeat mutation (expanded or normal) and the delta2642 polymorphism (insertion or deletion) were also included. Phase-known haplotypes were taken in account, and only haplotype pairs with a PHASE probability greater than 0.8 were used for further analyses.

Statistical Analysis

Since the *HTT* CAG distribution and repeat-length variation among the different core haplotypes did not display a Normal distribution, these two variables were analyzed and compared between haplotypes by using nonparametric tests: (i) the independent samples Kruskal-Wallis test; and (ii) the independent samples median test. Statistical analyses were performed using PASW Statistics 18 (SPSS Inc., Chicago, IL).

RESULTS

Extended ~1Mb Haplotype Shared by 19% of Portuguese HD Chromosomes

We extracted genotypes from a region of 1.04Mb around the *HTT* gene for 805 independent chromosomes, from which we were able to generate 786 haplotypes after phasing: 317 HD and 469 control chromosomes. In the extended haplotype, which incorporates the entire 1.04Mb region around *HTT*, we included all nine SNPs, the proline repeat and the delta2642 polymorphism. We were able to construct complete phased haplotypes, at all these eleven *loci*, for 193 HD and 267 control chromosomes. The remaining HD and

control chromosomes were not considered in this first analysis because some chromosomes were missing genotypes or could not be unambiguously phased at all loci.

In the Portuguese HD chromosomes, we found 39 different extended haplotypes, with the top 4 haplotypes (haplotypes 1–4 in Fig. 1) accounting for almost 50% of the 193 HD chromosomes, but only 18% of the 267 control chromosomes (Table S1 and S2).

The most common of these extended HD haplotypes, haplotype 1, which was shared by 18.7% of HD chromosomes, but only 3.8% of control chromosomes, includes the delta2642 allele (shown previously to be enriched in HD chromosomes) [Ambrose et al., 1994; Novelletto et al., 1994; Squitieri et al., 1994; Almqvist et al., 1995] and the 7-CCG repeat (most frequent in both mutated and normal Western-European chromosomes) [Almqvist et al., 1994; Squitieri et al., 1994; Hecimovic et al., 2002; Garcia-Planells et al., 2005; Costa et al., 2006]. The second (14.5%) and third (6.7%) most common HD haplotypes (haplotype 2 and 3) accounted, respectively, for 7.1% and 5.6% of control chromosomes. The other HD major haplotype, haplotype 4, was present in 6.2% of all HD chromosomes and was relatively rare in control chromosomes (1.9%). Of the 39 haplotypes found in the HD chromosomes, 21 were not found at all in control chromosomes; yet, they accounted for about ¼ of all HD chromosomes. On the other hand, we found 68 different extended haplotypes in our Portuguese control chromosomes (Table S2), of which only 18 were also found in HD chromosomes. The two most frequent extended haplotypes in control chromosomes, haplotype 9 (15.4%) and haplotype 20 (7.9%), were shared by only 4.6% of all HD chromosomes; neither included the delta2642 allele. Notably, we found a greater genetic diversity among control chromosomes, where 50 low frequency haplotypes accounted for 45% of normal chromosomes.

Enrichment of Core Haplotype A Among Portuguese HD Chromosomes

In order to understand whether some of the rarer HD haplotypes could have derived from any of the four major extended haplotypes, we grouped the haplotypes based on a smaller region surrounding the CAG repeat, including rs762847 (164 Kb upstream of the *HTT* CAG repeat), the adjacent CCG proline repeat and the delta2642 polymorphism (154 Kb downstream the repeat), as shown in Figure 1. The 317 independent haplotypes found among the HD chromosomes were combined into eight different "core" haplotypes (haplotypes A–H, Fig. 1). Haplotypes found among the control chromosomes but not in HD chromosomes, were grouped together and designated "other".

Two major core haplotypes (A and B) accounted for more than 60% of all Portuguese HD chromosomes (Fig. 2). Core haplotype A, which included (among others) the most common extended haplotype (haplotype 1), was present in about ¼ of all HD chromosomes while only in 6% of the control chromosomes. Notably, almost all chromosomes (63 out of 68 genotyped for the *D4S127* marker) within core haplotype A included the 157-bp allele at *D4S127*, a single tandem repeat (STR) marker previously shown to

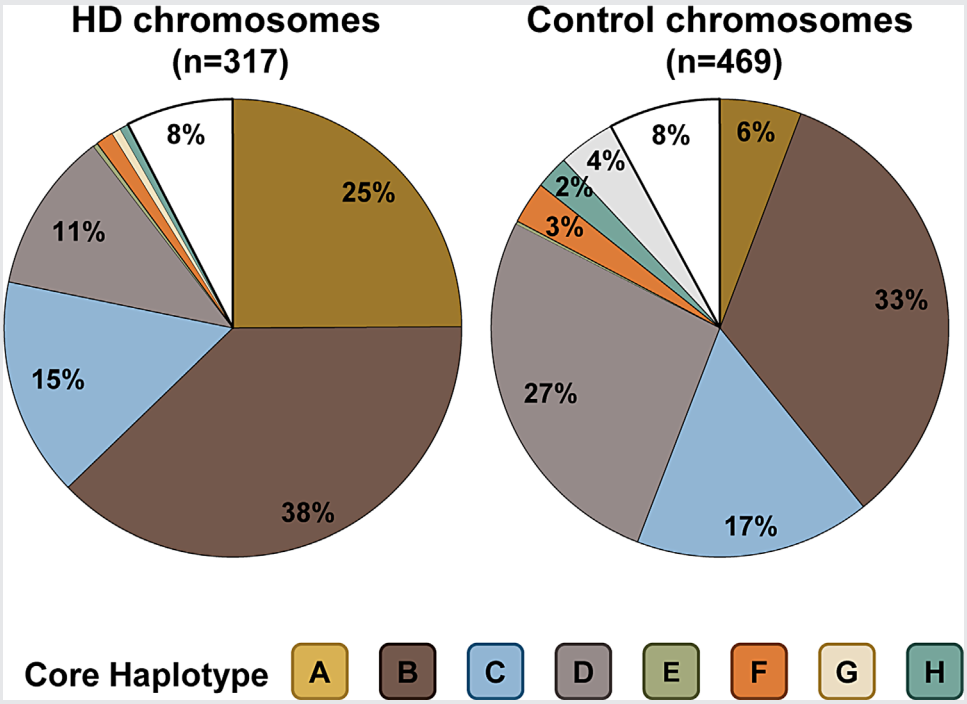

Fig. 2. Frequency of core haplotypes among HD and control chromosomes. The frequencies of each core haplotype (A–H) on HD and control chromosomes are shown, as well as the frequency of control chromosomes with "other" core haplotypes (light grey). The remaining (white) represents the frequency of chromosomes where the core haplotype could not be unambiguously determined.

be highly associated with the HD mutation, [MacDonald et al., 1992] suggesting that these chromosomes are indeed ancestrally related. Also, we observed that ~80% of chromosomes with this core haplotype shared the complete downstream region (up to 362 Kb downstream the HD CAG repeat), while there was more diversity in the region upstream of marker rs762847 (Fig. S1). Even though core haplotype A included the most common extended haplotype, core haplotype B is actually the most frequent core haplotype and is equally represented among HD (38%) and control (33%) chromosomes. By typing the additional STR marker proximal to the CAG repeat, we observed that most of the HD chromosomes (83 out of 107 genotyped for *D4S127*) with this core haplotype shared the same 155-bp allele, while in the control chromosomes we observed three alleles: 153-, 155-, and 157-bp, corresponding to respectively, 24.1%, 28.6%, and 42.9% of the control chromosomes genotyped for *D4S127* within core haplotype B. This finding reflects the higher genetic diversity observed in control chromosomes, both upstream and downstream of the CAG repeat (Fig. S1). On the other hand, one of the most frequent core haplotypes among control chromosomes (27%), core haplotype D, was present only in 11% of the HD chromosomes, and the majority of HD and normal chromosomes with this core haplotype shared the 151-bp STR allele. Most of the genetic variability in this core haplotype was observed upstream of the *HTT*CAG repeat (Fig. S1). HD chromosomes with core haplotype C had a similar frequency to control chromosomes (15% and 17%, respectively). However, while in HD chromosomes, the STRs revealed the presence of two possible variants: 151-bp (38%) and 155-bp allele (42%), most of the control chromosomes (54 out of 65 genotyped for *D4S127*) had the 151-bp allele.

High-Normal Alleles in Major Haplotypes are Genetically Closer to HD Alleles

The CAG repeat sizes for all chromosomes within core haplotypes A–D are shown in Figure 3. We observed that core haplotype A occurred in chromosomes with repeat sizes ranging from 9 to 73, haplotype B from 12 to 87, haplotype C from 14 to 64, and

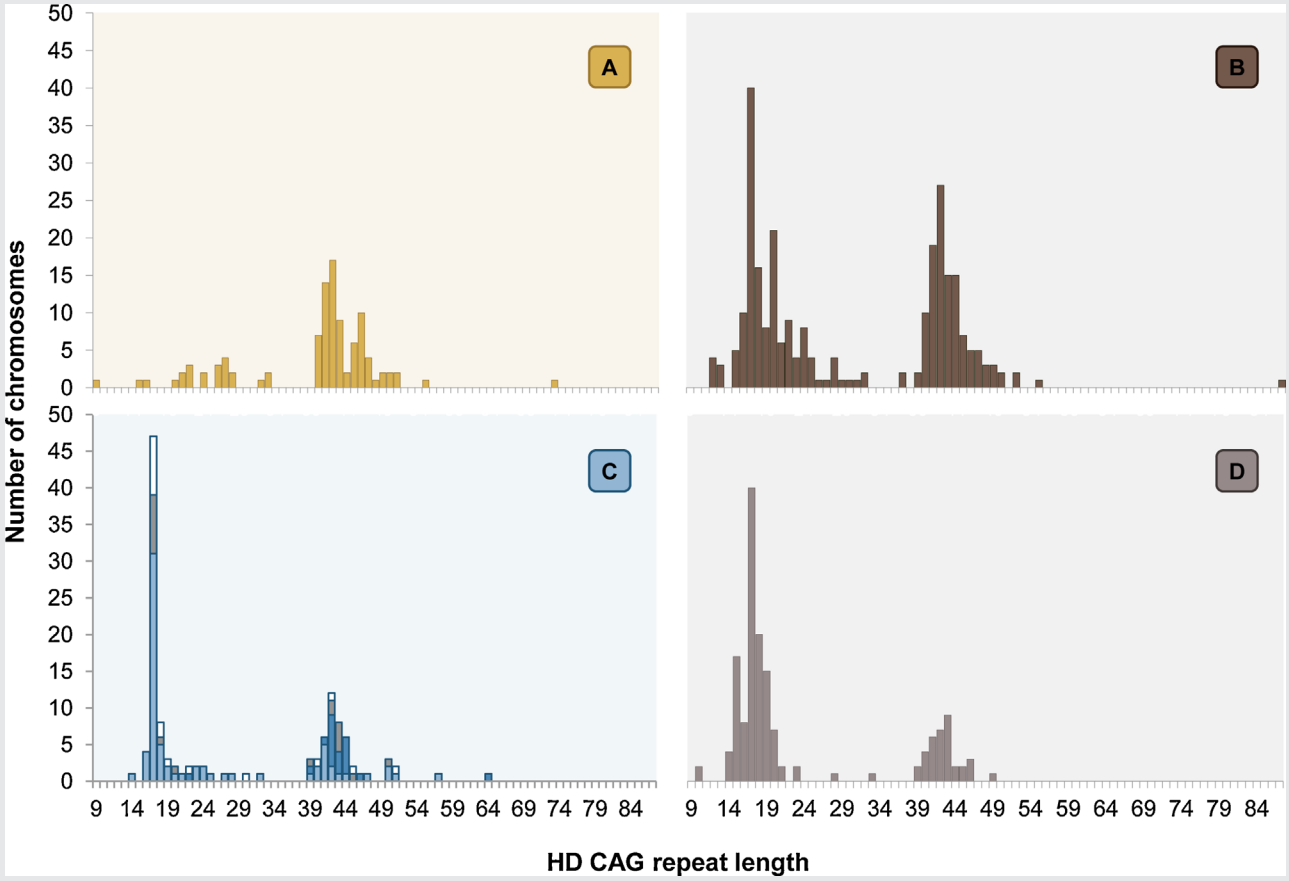

Fig. 3. ***HTT*** **CAG distribution for all chromosomes with major core haplotypes A–D. In our set of HD chromosomes, the median CAG repeat for core haplotype A was 43.0 (n = 78), B was 42.0 (n = 119), C was 43.0 (n = 49), and D was 42.0 (n = 36). For normal range allele sizes, the median CAG size for core haplotype A was 26.0 (n = 23), 18.0 for B (n = 149), 17.0 for C (n = 77), and 17.0 for D (n = 119). For haplotype C, we separated the 126 chromosomes based upon their *D4S127* genotype: 151-bp, 155-bp, other in grey and missing genotype in white, from bottom to top. The median CAG repeat for haplotype C associated with 151-bp *D4S127* allele was 17.0 (n = 70) and with 155-bp was 42.5 (n = 20).**

haplotype D from 10 to 49 repeats. In HD chromosomes (>35 CAGs) the median and distribution of *HTT* CAG repeats were the same across the major core haplotypes (Independent-Samples Median Test, $P = 0.979$; Independent-Samples Kruskal-Wallis Test, $P = 0.443$). On the other hand, when considering only chromosomes in the normal *HTT* CAG repeat range (<36 CAGs), there were significant differences in the medians and distribution of *HTT* CAG repeats across core haplotypes (Independent-Samples Median Test, $P < 0.001$; Independent-Samples Kruskal-Wallis Test, $P < 0.001$), core haplotype A and B showing a higher median CAG repeat size (26.0 and 18.0, respectively), than core haplotypes C and D (17.0 CAGs).

When comparing the extended haplotype of control chromosomes with larger CAG repeats (20–35 CAGs), we observed that, for core haplotypes B and C, these were more similar to the corresponding HD extended haplotypes (haplotype 2 and 4, respectively) than chromosomes with lower CAG repeats. Noticeably, in control chromosomes with haplotype B and shorter CAG repeats (<20 CAGs), the most frequent *D4S127* allele was 157-bp, whereas control haplotype B chromosomes with longer CAG repeats (20–35 CAGs) showed a higher percentage of the 155-bp STR allele, present in 81% of all haplotype B HD chromosomes.

HD Haplotypes and Repeat Instability

We were able to collect intergenerational instability data for 83 of the families included in this study and examined the HD-associated haplotype in 41 maternal and 42 paternal transmissions (Table I). The mean repeat-length change was $+0.59 \pm 4.14$ (median = 0.00) for maternal and $+2.69 \pm 4.46$ (median = +1.50) for paternal transmissions. In both types of transmissions the median repeat length-variation and its distribution were the same across all core haplotypes (Independent-Samples Median Test, $P = 0.985$ for maternal and $P = 0.632$ for paternal transmissions; Independent-Samples Kruskal-Wallis Test, $P = 0.933$ for maternal and $P = 0.364$ for paternal transmissions), despite similar paternal CAG repeat sizes across haplotypes (Independent-Samples Median Test, $P = 0.232$ for maternal and $P = 0.384$ for paternal CAG repeat sizes). Both the largest maternal (+25) and paternal (+24) transmissions observed among Portuguese families [Ramos et al., 2012] shared the same core haplotype A. Another large paternal transmission (+12) expansion, involving a 36 CAGs (reduced penetrance) allele that expanded to a full penetrance allele, [Ramos et al., 2012] occurred on core haplotype B.

Unstable Normal Alleles and Haplotypes

We determined the *HTT* CAG repeat for ~1,300 members of an Italian population isolate with complete genealogy information up to 15 generations. [Pattaro et al., 2007] We were able to trace 472 transmissions of the *HTT* CAG repeat from 120 closely related nuclear families. From these, we only found two unstable transmissions of normal range *HTT* CAG alleles in one family (Fig. 4, Family 1). In this consanguineous family (with a common ancestor within four meiosis), the paternal 23 CAG repeat allele on haplotype C seemed to have contracted by one repeat, while the maternal 20 CAG repeat allele on haplotype B contracted by two CAGs. Even though the occurrence in a single family of two events of unstable transmission of *HTT* CAG alleles is highly unlikely, extensive genotyping showed no evidence for non-paternity or non-maternity. We also observed an unstable transmission of a high-normal 32 CAG repeat that expanded to 33 CAGs, which occurred on a haplotype C chromosome inherited from a father (Fig. 4, Family 2). We were able to trace 22 transmissions of the *HTT* CAG repeat,

TABLE I. Transmission of the Expanded HD Repeat in Major Core Haplotypes A to D

Core haplotype	All	A	B	C	D
Maternal transmissions					
Transmissions, n	41	10	12	10	8
Range of transmitted allele (CAGs)	40–73	42–73	40–54	40–65	41–46
Mean	44	47	43	45	43
Unstable transmissions, n	28	7	7	7	6
Contractions, n	15	4	3	4	3
Expansions, n	13	3	4	3	3
Repeat-length variation, mean	+0.59	+2.20	+0.50	-0.10	-0.13
Median	0.00	0.00	0.00	0.00	0.00
Range	−2 to +25	−2 to +25	−2 to +4	−2 to +2	−2 to +1
Paternal transmissions					
Transmissions, n	42	9	19	5	7
Range of transmitted allele (CAGs)	36–52	41–52	36–51	39–46	38–43
Mean	42	44	42	42	42
Unstable transmissions, n	31	8	12	4	5
Contractions, n	4	0	3	0	1
Expansions, n	27	8	9	4	4
Repeat-length variation, mean	+2.69	+4.89	+2.00	+1.40	+1.71
Median	+1.50	+3.00	0.00	+1.00	+1.00
Range	−2 to +24	0 to +24	−2 to +12	0 to +3	−1 to +7

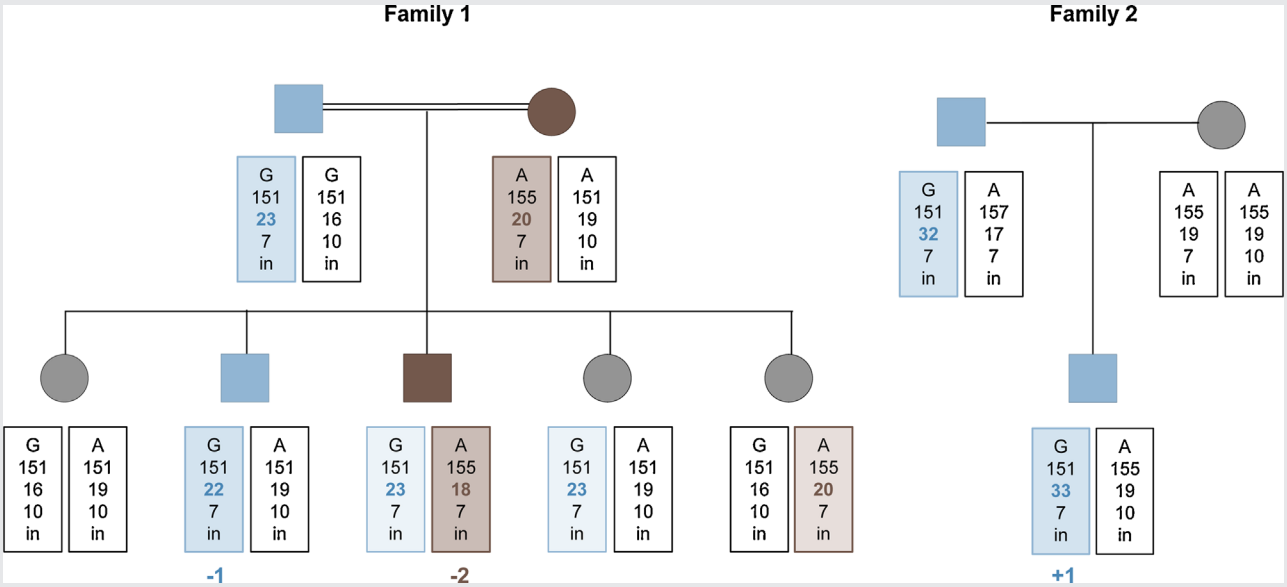

Fig. 4. Pedigrees and haplotypes from two families in the MICROS study that carried unstable non-pathogenic HD alleles. Colored symbols indicate individuals that carried an unstable allele (blue represents unstable allele associated with haplotype C and brown corresponds to haplotype B). Haplotypes of 4 markers spanning the *HTT* gene (core-haplotype: rs762847 - *D4S127* - *HTT* CAG repeat - CCG repeat - delta2642, from top to bottom) are shown, with the unstable allele haplotype boxed in the corresponding haplotype color (brown: haplotype B, blue: haplotype C).

from 13 closely related nuclear families, that were associated with the delta2642 allele (that marked haplotype A and also the rare haplotypes E and G); the transmitted *HTT* CAG repeat had mostly repeat sizes between 20–24 CAG repeats, while in two sibships was associated with a 12 CAG repeat allele. None of these *HTT* CAG repeat transmissions associated with the delta2642 allele was unstable.

DISCUSSION

Originally, it was believed that the HD mutation emerged among Western Europeans, [Hayden, 1981] mainly due to the highest rates of HD in these populations. However, after the discovery of the HD causal mutation, subsequent analysis of the *HTT* region showed that the geographical variation in HD prevalence might be correlated with the frequency of certain polymorphic alleles that resulted in a variety of independent and not easily related HD haplotypes, suggesting different independent origins for the HD mutation. [Squitieri et al., 1994; Almqvist et al., 1995; Andrew and Hayden, 1995] Although our study was focused on a small population in the South of Europe, we found 39 different extended haplotypes in the Portuguese HD chromosomes. Despite this diversity, about 19% of all HD chromosomes shared an identical haplotype (haplotype 1), extending over 680 Kb downstream and 362 Kb upstream the *HTT* CAG repeat, and marked by the delta2642 allele. This major HD haplotype seems to correspond to the most frequent HD haplotype defined in earlier studies and it is extremely similar, if not exactly the same, to the extended haplotype ("Ext_Hap_1") found to account for about 25% of a set of 699 European HD chromosomes. [Lee et al., 2012] Indeed, by comparing the common SNPs and corresponding haplotypes between the two studies (Fig. S2), we observed that our extended haplotype 1 includes, among other low frequency Western European HD haplotypes, the "Ext_Hap_1". The frequency of this haplotype (Fig. S3) among Portuguese HD chromosomes (23%) is thus lower than in the Western European HD chromosomes (38%), indicating that other haplotypes (extended haplotypes 2, 3, and 4) have contributed to a greater extent to HD in Portugal than in European chromosomes overall.

As in previous studies, [Ambrose et al., 1994; Novelletto et al., 1994; Squitieri et al., 1994; Almqvist et al., 1995] we found an increased frequency of the delta2642 allele among the Portuguese HD chromosomes, in almost 27% of all HD chromosomes, whereas only 6% of the control chromosomes had this rare allele. This suggests that most or all of those HD chromosomes with the delta2642 allele may be ancestrally related, but that in many cases the extended haplotype decayed by recombination events that produced genetic differences across the ~1Mb region around *HTT*. Therefore, we reduced the region of interest to a smaller core of 318 Kb around the *HTT* CAG repeat. We were thus able to group related extended haplotypes and recognize eight different core haplotypes associated with HD chromosomes. The four major core haplotypes are not easily related to each other: haplotype A (enriched in HD) is the only haplotype with the delta2642 allele, while haplotype D (enriched in controls) is the only one with a 10-CCG repeat. Furthermore, by looking at a STR marker upstream of the *HTT* CAG repeat, we observed a nonrandom allelic association on HD chromosomes between core haplotypes (A, B, and D) and *D4S127*. This marker seems to be relatively stable as allele changes

occurred at a rate low enough that linkage disequilibrium could still be detected with the background haplotype. A more detailed analysis of the four major haplotypes revealed that multiple occurrences of the CAG expansion seemed to have happened within the same haplotype. For example, HD within core haplotypes B and C could be due to more than one independent expansion on each haplotype as normal chromosomes bearing these haplotypes are frequent in the Portuguese population. This is probably also true for haplotype A, where its enrichment in high normal alleles would support decay of the extended haplotype on normal chromosomes prior to more than one generation of an HD allele.

Early HD studies have suggested that particular haplotypes, mainly marked by the delta2642 allele and/or the 7-CCG allele, seemed to be particularly predisposed to expansion that eventually would result in a repeat allele within the HD range, supported by an enrichment of repeats in the high normal range in these particular haplotypes. [Squitieri et al., 1994; Almqvist et al., 1995; Hecimovic et al., 2002; Costa et al., 2006; Agostinho et al., 2012] We observed that, also in our Portuguese control chromosomes, the haplotype marked by 7-CCG and delta2642 alleles (haplotype A) had a higher median normal CAG repeat length than any of the other three major haplotypes (with most of the normal alleles ranging from 20 to 33 repeats). Even though we cannot rule out the hypothesis that *cis*-acting factors have a major predisposing effect on the instability of the *HTT* repeat, it is also possible that haplotype A, present in only 6% of the control chromosomes, has a bias toward higher repeat alleles as a result of genetic drift. [Falush, 2009] Repeat-length variation is strongly related to the initial parental CAG size; therefore, a larger average repeat size would make haplotype A more likely to expand and become a new HD mutation. It is of interest that even though this specific haplotype was shared by two of the biggest expansions observed in the Portuguese population there was no statistical difference in repeat instability across haplotypes. Indeed, we also observed a large paternal expansion on haplotype B (+12 CAGs) that resulted in a repeat class change to full penetrance range and both haplotypes B and C carried alleles with very large repeats. Indeed, the highest repeat size in this study was observed in haplotype B (87 CAG repeats). This suggests that haplotypes B and C also seem to be prone to large expansions that we may not have observed due to missing genotypes on these pedigrees. Our data suggest that, instead of a predisposing haplotype, there is a pool of high normal alleles in each different founder haplotype more prone to give rise to pathogenic alleles.

To evaluate this possibility, in an isolate population from Italy, [Pattaro et al., 2007] we were able to trace 472 transmissions of *HTT* CAG repeats in the normal range, and only in three instances was the repeat unstably transmitted. These unstable transmissions happened on normal chromosomes with haplotypes B and C, showing that in the non-pathogenic alleles instability also occurs in haplotypes other than haplotype A. Furthermore, the fact that two of these events involved siblings suggests that there might be a genetic factor that acts in the zygote, perhaps due to recessive transmission/consanguinity. On the other hand, in all normal chromosomes associated with the delta2642 allele the associated *HTT* CAG repeat was stably transmitted. These data provide evidence that, in the normal population, haplotypes other than haplotype A predispose to CAG repeat expansions.

In summary, our study showed a diversity of haplotypes in both normal and expanded chromosomes, representative of more than one ancestral chromosome underlying HD in Portugal, where multiple independent events have given rise to expansion into the pathogenic range. In contrast to what has been reported in other studies, our data does not support the hypothesis of haplotype A (marked by the *HTT* delta2642 allele and enriched in HD chromosomes) having a major predisposing influence on HD CAG instability, as there was no statistical difference in repeat instability across the different major haplotypes. In fact, normal *HTT* alleles with larger CAG repeats were more similar to extended HD haplotypes than alleles with smaller CAG repeats, suggesting that, as expected, these range of alleles are a reservoir for new HD mutations, independently of their haplotype background.

ACKNOWLEDGMENTS

The authors would like to thank all patients and their families who accepted to participate in this study. We also thank physicians who referred patients for HD testing to our center: A Ramos, A Fernandes, JM Leite, F Nogueira, S Beck, T Mestre, FR Martins, F Magalhães, GC Santo, B Botelho, G Neves, JC Pereira, E Monteiro, AS Morgadinho, AI Rebelo, AM Rodrigues, M Rodrigues, A Geraldo, B Rodrigues, C Januário, MA Ferro, MC Macário, NMM da Silva, MHC Fernandes, A Morganho, L Nunes, T Lourenço, J Guimarães, J Vale, MJ Feijó, P Tavares, MTC Pereira, AB Sousa, F Falcão, H Santos, I Cordeiro, IM Gaspar, J Ferreira, MM Rosa, M Lauterbach, R Geraldes, C Moura, C Garrett, JM Campillo, P Esperança, J Figueiredo, M Veloso, JL Loureiro, P Coutinho, F Piloto, A Henriques, AP Carvalho, C Moço, A Palos, J Becho, JMC Guerra, MG Mira, C Costa, J Coimbra, J Proença, JCP Ferreira, RJ Guerra, AP Correia, A Tuna, M Magalhães, M Mota, V Passão, FV Ramos, JJS de Sá, JM Saraiva, M Amorim, SB de Sousa, S Maia, D Alves, AC Araújo, H Águas, MS Dias, M Lourenço, M Rocha, D Martins and J Marinho. For the MICROS study, we thank the primary care practitioners R Stocker, S Waldner, T Pizzecco, J Plangger, U Marcadent and the personnel of the Hospital of Silandro (Department of Laboratory Medicine) for their participation and collaboration in the research project.

REFERENCES

Agostinho LD, Rocha CF, Medina-Acosta E, Barboza HN, da Silva AF, Pereira SP, da Silva ID, Paradela ER, Figueiredo AL, Nogueira ED, Alvarenga RM, Hernan Cabello P, Dos Santos SR, Paiva CL. 2012. Haplotype analysis of the CAG and CCG repeats in 21 Brazilian families with Huntington's disease. J Hum Genet 57(12):796–803.

Almqvist E, Andrew S, Theilmann J, Goldberg P, Zeisler J, Drugge U, Grandell U, Tapper-Persson M, Winblad B, Hayden M, Anvret M. 1994. Geographical distribution of haplotypes in Swedish families with Huntington's disease. Hum Genet 94(2):124–128.

Almqvist E, Spence N, Nichol K, Andrew SE, Vesa J, Peltonen L, Anvret M, Goto J, Kanazawa I, Goldberg YP, Hayden MR. 1995. Ancestral differences in the distribution of the delta 2642 glutamic acid polymorphism is associated with varying CAG repeat lengths on normal chromosomes: insights into the genetic evolution of Huntington disease. Hum Mol Genet 4(2):207–214.

Ambrose CM, Duyao MP, Barnes G, Bates GP, Lin CS, Srinidhi J, Baxendale S, Hummerich H, Lehrach H, Altherr M, Wasmuth J, Buckler A, Church D, Housman D, Berks M, Micklem G, Durbin R, Dodge A, Read A, Gusella J, MacDonald ME. 1994. Structure and expression of the Huntington's disease gene: Evidence against simple inactivation due to an expanded CAG repeat. Somat Cell Mol Genet 20(1):27–38.

Andrew SE, Hayden MR. 1995. Origins and evolution of Huntington disease chromosomes. Neurodegeneration 4(3):239–244.

Baine FK, Kay C, Ketelaar ME, Collins JA, Semaka A, Doty CN, Krause A, Jacquie Greenberg L, Hayden MR. 2013. Huntington disease in the South African population occurs on diverse and ethnically distinct genetic haplotypes. Eur J Hum Genet 21(10):1120–1127.

Brocklebank D, Gayan J, Andresen JM, Roberts SA, Young AB, Snodgrass SR, Penney JB, Ramos-Arroyo MA, Cha JJ, Rosas HD, Hersch SM, Feigin A, Cherny SS, Wexler NS, Housman DE, Cardon LR. 2009. Repeat instability in the 27-39 CAG range of the HD gene in the Venezuelan kindreds: Counseling implications. Am J Med Genet B Neuropsychiatr Genet 150B(3):425–429.

Costa MC, Magalhaes P, Guimaraes L, Maciel P, Sequeiros J, Sousa A. 2006. The CAG repeat at the Huntington disease gene in the Portuguese population: Insights into its dynamics and to the origin of the mutation. J Hum Genet 51(3):189–195.

Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, Folstein S, Ross C, Franz M, Abbott M, Gray J, Conneally P, Young A, Penney J, Hollingsworth Z, Shoulson I, Lazzarini A, Falek A, Koroshetz W, Sax D, Bird E, Vonsattel J, Bonilla E, Alvir J, Bickham Conde, Cha J, Dure J-H, Gomez L, Ramos F, Sanchez-Ramos M, Snodgrass J, de Young S, Wexler M, Moscowitz N, Penchaszadeh C, MacFarlane G, Anderson H, Jenkins M, Srinidhi B, Barnes J, Gusella G, MacDonald J. 1993. Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat Genet 4(4):387–392.

Falush D. 2009. Haplotype background, repeat length evolution, and Huntington's disease. Am J Hum Genet 85(6):939–942.

Garcia-Planells J, Burguera JA, Solis P, Millan JM, Ginestar D, Palau F, Espinos C. 2005. Ancient origin of the CAG expansion causing Huntington disease in a Spanish population. Hum Mutat 25(5):453–459.

Goldberg YP, Kremer B, Andrew SE, Theilmann J, Graham RK, Squitieri F, Telenius H, Adam S, Sajoo A, Starr E, Heiberg A, Wolff G, Hayden MR. 1993. Molecular analysis of new mutations for Huntington's disease: Intermediate alleles and sex of origin effects. Nat Genet 5(2):174–179.

Hayden MR. 1981. Huntington's chorea. New York: Springer-Verlag.

Hecimovic S, Klepac N, Vlasic J, Vojta A, Janko D, Skarpa-Prpic I, Canki-Klain N, Markovic D, Bozikov J, Relja M, Pavelic K. 2002. Genetic background of Huntington disease in Croatia: Molecular analysis of CAG, CCG, and Delta2642 (E2642del) polymorphisms. Hum Mutat 20(3):233.

Lee JM, Gillis T, Mysore JS, Ramos EM, Myers RH, Hayden MR, Morrison PJ, Nance M, Ross CA, Margolis RL, Squitieri F, Griguoli A, Di Donato S, Gomez-Tortosa E, Ayuso C, Suchowersky O, Trent RJ, McCusker E, Novelletto A, Frontali M, Jones R, Ashizawa T, Frank S, Saint-Hilaire MH, Hersch SM, Rosas HD, Lucente D, Harrison MB, Zanko A, Abramson RK, Marder K, Sequeiros J, Macdonald ME, Gusella JF. 2012. Common SNP-Based Haplotype Analysis of the 4p16.3 Huntington Disease Gene Region. Am J Hum Genet 90(3):434–444.

MacDonald ME, Novelletto A, Lin C, Tagle D, Barnes G, Bates G, Taylor S, Allitto B, Altherr M, Myers R, Lehrach H, Collins FS, Wasmuth JJ, Frontali M, Gusella JF. 1992. The Huntington's disease candidate region exhibits many different haplotypes. Nat Genet 1(2):99–103.

Myers RH, MacDonald ME, Koroshetz WJ, Duyao MP, Ambrose CM, Taylor SA, Barnes G, Srinidhi J, Lin CS, Whaley WL, Lazzarini AM, Schwarz M, Wolff G, Bird ED, Vonsattel J-PG, Gusella JF. 1993. De novo expansion of a (CAG) n repeat in sporadic Huntington's disease. Nat Genet 5(2):168–173.

Norremolle A, Budtz-Jorgensen E, Fenger K, Nielsen JE, Sorensen SA, Hasholt L. 2009. 4p16.3 haplotype modifying age at onset of Huntington disease. Clin Genet 75(3):244–250.

Novelletto A, Persichetti F, Sabbadini G, Mandich P, Bellone E, Ajmar F, Squitieri F, Campanella G, Bozza A, MacDonald ME, Gusella JF, Frontali M. 1994. Polymorphism analysis of the huntingtin gene in Italian families affected with Huntington disease. Hum Mol Genet 3(7):1129–1132.

Paradisi I, Hernandez A, Arias S. 2008. Huntington disease mutation in Venezuela: Age of onset, haplotype analyses and geographic aggregation. J Hum Genet 53(2):127–135.

Pattaro C, Marroni F, Riegler A, Mascalzoni D, Pichler I, Volpato CB, Dal Cero U, De Grandi A, Egger C, Eisendle A, Fuchsberger C, Gogele M, Pedrotti S, Pinggera GK, Stefanov SA, Vogl FD, Wiedermann CJ, Meitinger T, Pramstaller PP. 2007. The genetic study of three population microisolates in South Tyrol (MICROS): Study design and epidemiological perspectives. BMC Med Genet 8:29.

Ramos EM, Cerqueira J, Lemos C, Pinto-Basto J, Alonso I, Sequeiros J. 2012. Intergenerational instability in Huntington disease: Extreme repeat changes among 134 transmissions. Mov Disord 27(4):583–585.

Rubinsztein DC, Leggo J, Barton DE, Ferguson-Smith MA. 1993. Site of (CCG) polymorphism in the HD gene. Nat Genet 5(3):214–215.

Saleem Q, Roy S, Murgood U, Saxena R, Verma IC, Anand A, Muthane U, Jain S, Brahmachari SK. 2003. Molecular analysis of Huntington's disease and linked polymorphisms in the Indian population. Acta Neurol Scand 108(4):281–286.

Semaka A, Collins JA, Hayden MR. 2010. Unstable familial transmissions of Huntington disease alleles with 27-35 CAG repeats (intermediate alleles). Am J Med Genet B Neuropsychiatr Genet 153B(1):314–320.

Sequeiros J, Ramos EM, Cerqueira J, Costa MC, Sousa A, Pinto-Basto J, Alonso I. 2010. Large normal and reduced penetrance alleles in Huntington disease: Instability in families and frequency at the laboratory, at the clinic and in the population. Clin Genet 78(4):381–387.

Squitieri F, Andrew SE, Goldberg YP, Kremer B, Spence N, Zeisler J, Nichol K, Theilmann J, Greenberg J, Goto J, Kanazawa I, Vesa J, Peltonen L, Almqvist E, Anvret M, Telenius H, Lin B, Napolitano G, Morgan K, Hayden MR. 1994. DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. Hum Mol Genet 3(12):2103–2114.

Taylor SA, Barnes GT, MacDonald ME, Gusella JF. 1992. A dinucleotide repeat polymorphism at the D4S127 locus. Hum Mol Genet 1(2):142.

The Huntington's Disease Collaborative Research Group. 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72(6):971–983.

Warby SC, Montpetit A, Hayden AR, Carroll JB, Butland SL, Visscher H, Collins JA, Semaka A, Hudson TJ, Hayden MR. 2009. CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. Am J Hum Genet 84(3):351–366.

Warby SC, Visscher H, Collins JA, Doty CN, Carter C, Butland SL, Hayden AR, Kanazawa I, Ross CJ, Hayden MR. 2011. HTT haplotypes contribute to differences in Huntington disease prevalence between Europe and East Asia. Eur J Hum Genet 19(5):561–566.

Warner JP, Barron LH, Brock DJ. 1993. A new polymerase chain reaction (PCR) assay for the trinucleotide repeat that is unstable and expanded on Huntington's disease chromosomes. Mol Cell Probes 7(3):235–239.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.