CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Digital.CSIC

Ancient origin of the CAG expansion causing Huntington disease in a Spanish population

Javier García-Planells[1], Juan A. Burguera[2], Pilar Solís[2], José M. Millán[3], Damián Ginestar[4], Francesc Palau[1], Carmen Espinós[1]

[1] Laboratory of Genetics and Molecular Medicine, Department of Genomics and Proteomics, Instituto de Biomedicina, CSIC, C/ Jaume Roig 11, 46010 Valencia, Spain

Departments of [2]Neurology and [3]Genetics, Hospital Universitari La Fe, Av. Campanar 21, 46009 Valencia, Spain

[4]Department of Applied Mathematics, Universitat Politècnica de València, Camí de Vera s/n, 46022 Valencia, Spain

Running title: Huntington disease in Valencia, Spain

Corresponding author:

Dr. Francesc Palau

C/ Jaume Roig, 11, 46010 Valencia, Spain

Tel.: + 34 96 339 37 73

Fax: + 34 96 369 08 00

E-mail: fpalau@ibv.csic.es

ABSTRACT

Huntington disease (HD, MIM# 143100) is an autosomal dominant neurodegenerative disorder characterized clinically by progressive motor impairment, cognitive decline, and emotional deterioration. The disease is caused by the abnormal expansion of a CAG trinucleotide repeat in the first exon of the *huntingtin* gene in chromosome 4p16.3. HD is spread worldwide and it is generally accepted that few mutational events account for the origin of the pathogenic CAG expansion in most populations. We have investigated the genetic history of HD mutation in 83 family probands from the Land of Valencia, Eastern Spain. An analysis of the HD/CCG repeat in informative families suggested that at least two main chromosomes were associated in the Valencian population, one associated with allele 7 (77 mutant chromosomes) and one associated with allele 10 (2 mutant chromosomes). Haplotype A-7-A (H1) was observed in 47 out of 48 phase-known mutant chromosomes, obtained by segregation analysis, through the haplotype analysis of rs1313770-HD/CCG-rs82334, as it also was in 120 out of 166 chromosomes constructed by means of the PHASE program. The genetic history and geographical distribution of the main haplotype H1 were both studied by constructing extended haplotypes with flanking STRs *D4S106* and *D4S3034*. We found that we were able to determine the age of the CAG expansion associated with the haplotype H1 as being between 4,700 and 10,000 years ago. Furthermore, we observed a non-homogenous distribution in the different regions associated with the different extended haplotypes of the ancestral haplotype H1, suggesting that local founder effects have occurred.

KEYWORDS: ***Huntingtin*** **gene; CAG trinucleotide repeat; Land of Valencia; allele age; founder effect**

DATABASES

HD, MIM# 143100

genome.ucsc.edu (UCSC GENOME BROWSER)

www.ncbi.nlm.nih.gov/entrez (NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION)

www.rannala.org/labpages/software.html (BDMC21 V2.1)

INTRODUCTION

Huntington disease (HD; MIM# 143100) is an autosomal dominant neurodegenerative disorder of mild-life onset that results in progressive neurological movement symptoms accompanied by psychiatric alterations and cognitive decline (Martin and Gusella, 1986; Harper, 1996). HD occurs with a frequency of approximately 5-10 patients out of 100,000 individuals in most Caucasian populations, although its estimated frequency exhibits extreme variations among populations (Martin and Gusella, 1986; Harper, 1996). The HD genetic defect consists of an expansion of CAG triplets within exon 1 of the *huntingtin* gene on chromosome 4p16.3 (The Huntington Collaborative Research Group, 1993). As a general rule it is assumed that this highly polymorphic CAG repeat ranges from 8-35 copies on normal chromosomes, whereas the copy number ranges within 36-121 on disease chromosomes, although it depends on the populations (Zühlke et al., 1993; Kremer et al., 1994; The American College of Medical Genetics et al., 1998; Costa et al., 2003).

Dynamic mutations are caused by an expansion of trinucleotide sequences in or adjacent to a protein-coding gene (Ashley and Warren, 1995; Wells et al., 1998; Cummings and Zoghbi, 2000), and are characterised by the intergenerational instability of trinucleotide repeats and by an increasing bias during transmission. In this way, the existence of intermediate-sized nonpathologic alleles near the upper limit of the normal-sized range is thought to act as a reservoir from which *de novo* mutations arise in several generations, given that the larger the expanded alleles are, the more instable the expansion becomes (Richards et al., 1992; Goldberg et al., 1993; Imbert et al., 1993; Myers et al., 1993; Kunst and Warren, 1994; Paulson and Fischbeck, 1996). That is, a new mutation arises in an already long HD chromosome that acquires a pathological size due to an increase of the initial copy number of the repeat, mainly occurring during spermatogenesis and even before meiosis is completed

(De Rooij et al., 1993; Zühlke et al., 1993; Chong et al., 1997; Yoon et al., 2003). In many cases, expanded alleles and the longest normal alleles can be related thanks to flanking markers (Richards, 2001). In fact, strong linkage disequilibrium between disease alleles and close markers resulting from ancestral events has been referred to in many populations for several diseases caused by dynamic mutations, such as fragile-X syndrome (MIM# 309550) (Richards et al., 1992; Kunst and Warren, 1994), myotonic dystrophy (MIM# 160900) (Imbert et al., 1993; Liquori et al., 2003) or Friedreich's ataxia (MIM# 229300) (Cossée et al., 1997; Labuda et al., 2000). In reference to HD, haplotype studies carried out in different populations of a defined ancestry have revealed that multiple mutation events underlie this disorder even within each population (McDonald et al., 1992; Morrison et al., 1993; Squitieri et al., 1994; Almqvist et al., 1995). In this context, it is difficult to establish in most cases whether HD cases are caused by new mutations generated from longest normal expanded alleles or else, they descent from single founders.

Here we present allele and haplotype studies in a series of 83 HD family probands from the Land of Valencia (Eastern Spain) that represents 10% of the total Spanish population (approximately 4 million people). We have analysed five polymorphic markers within or close to the *huntingtin* gene and we have observed that more than 50 per cent of mutant chromosomes may have a very ancient common origin.

MATERIAL AND METHODS

Patients

This study involved 115 patients with molecular confirmation of HD. Patients were ascertained at the Neurology and Genetics Departments of the University Hospital La Fe. Both departments are reference services for clinical diagnosis and genetic counselling. All

patients and relatives were aware of the investigative nature of the studies and gave their consent. In order to learn about their geographical origin (town and county), the interview included information about the place of birth of both the proband and their parents, and when known, of grandparents. The HD patients belonging to 83 families originated from the Land of Valencia, in Eastern Mediterranean Spain. One hundred unrelated and healthy Valencian individuals, whose parents' samples were available (trio series), were also analysed as a control group. According to the geographical origin of the patients, frequencies have been estimated for different counties of the Land of Valencia.

Analysis of CAG-HD repeats

Genomic DNA was obtained from peripheral white blood cells by standard methods. The CAG repeat responsible for the disease was amplified by PCR independently of the adjacent CCG repeat, as previously described (Andrew et al., 1994), using the following primers: HD-C1: 5'-FAM-CCT TCG AGT CCC TCA AGT CCT TC-3', and HD-C2: 5'-CGG CGG TGG CGG CTG TTG-3'. The expected size of amplicons depends on the number of CAG repeats. It ranged from 59 bp to 146 bp for normal alleles (6-35 repeats), and from 149 bp to 404 bp for expanded alleles (36-121 repeats). The number of repeats was determined exactly by sequencing differently sized samples, and via comparative analyses in an autoanalyser ABI Prism 3100 (*Applied Biosystems*, Foster City, CA).

Allelic and Haplotype Analyses

Five polymorphic markers were studied: pter-*D4S126*-rs1313770(NT_006081.17:g.444387A>G)-HD/CCG-rs82334(NT_006081.17:g.611947A>C)-*D4S3034*-cen (Table 1, Fig. 1). Markers *D4S126*, HD/CCG and *D4S3034* were labelled and analysed in an autoanalyser ABI Prism 3100 (*Applied Biosystems*, Foster City, CA). SNPs

rs1313770 and rs82334 were analysed by restriction analyses using the endonuclease *Mbo*I and *BseD*I, respectively. Allelic frequencies of every marker were estimated in both the control and affected populations. A χ^2 test was performed to determine a significant allelic association between normal and affected chromosomes in cases with notable differences of allelic frequencies,. Haplotypes were established based on the five polymorphic markers in both populations by inheritance or by homozigosity, whenever possible. Otherwise, haplotypes were reconstructed using version 2.0 of the PHASE (PHASE v2.0) program running on MS-DOS.

Dating the mutation

Two mathematical approaches were applied to date the mutation: (1) The method reported by Serre et al. (1990) was applied, based on intra-allelic variability of linked markers, for which three mathematical models were performed: rs1313770_HD locus_rs82334, HD locus_rs82334_*D4S3034*; and rs1313770_HD locus_*D4S3034*; (2) a Monte Carlo likelihood method implemented in the program BDMC21 v2.1 was used, (Slatkin and Rannala, 1997; www.rannala.org/labpages/software.html), to consider information provided by the multiallelic markers *D4S126* and *D4S3034*. The program parameter settings were: growth rate = 0.05; sample fraction = 0.045 (assuming a frequency of HD= 5/100,000) (Burguera et al. 1997); and mutation rate of the linked markers = 0.0001. Program data settings were: mutant copies = 17; and segregating sites = 5. Program option settings were: number of Monte Carlo replicates = 10,000; and initial, final and increments = 10, 200, and 10, respectively. The resulting output file is converted in a log likelihood curve using MATLAB v6.5.

RESULTS

Mutation analysis

Searching for mutations has made it possible to establish the exact length of the pathogenic CAG repeats, and also to determine the distribution of mutant and normal alleles in the affected population and the control group (Fig. 2). The CAG length ranged from 37 to 86 for the HD population (between 38 and 46 units have been found in 93.60% of the disease chromosomes), and from 8 to 31 for the control group (between 13 and 21 units have been found in 87.17% of the control chromosomes). The allele with 15 repeats has been by far the most common normal allele.

Allelic and haplotype analyses

To perform allelic association studies, we estimated the allele frequencies of two SNPs, rs131770 and rs82334, and of the CCG repeat adjacent to HD mutation (Table 2) in both control and patient groups. Whenever possible, we established the associated allele to the HD mutation for every marker, by either familial segregation or by allelic homozygosity. The HD mutation was significantly associated with allele A of the marker rs131770 ($p<0.0001$), allele 7 of the CCG triplet ($p<0.0001$) and with allele A of the marker rs82334 ($p<0.0001$).

A haplotype analysis with the HD/CCG repeat and flanking SNPs was performed to determine the number of mutational events. Firstly, we constructed a core haplotype with markers rs1313770-HD/CCG-rs82334. Seventy nine out of 83 probands were correctly phased for the HD/CCG marker. In the two cases in which the HD mutation was associated with the 10-CCG allele, the haplotypes were A/G-10-C and A/G-10-A/C. Unfortunately, this phase could not be established more accurately. We could establish the exact phase of the core rs1313770-HD/CCG-rs82334 in 48 patients out of 77 probands that were associated with 7-CCG allele. In accordance with the variability of both flanking SNPs, only two haplotypes

associated with HD were unambiguously established among the four possible haplotypes: A-7-A (haplotype H1) in 47 families (56.6%) and A-7-C (haplotype H2) in 1 family (1.2%). In order to obtain more genetic information, we proceeded to infer haplotypes in the 83 available patients (166 chromosomes) by using the PHASE v2.0 program. We included both the 48 phase-known mutant haplotypes and the 35 phase-unknown mutant haplotypes. Eight different haplotypes were discerned and the two phase-unequivocal haplotypes were also computationally constructed. Haplotype H1 was by far the best option yielded by the program and was generated in 120 out of 166 chromosomes, and haplotype H2 did so in 2 chromosomes (2/166). Two other haplotypes generated by the PHASE v2.0 program are worthy of mention since they were represented in a notable number: G-7-A (H3) and G-7-C (H4) in 20/166 and 11/166 chromosomes, respectively. These new haplotypes generated by the PHASE program suggested that there might possibly more haplotypes and therefore, some of them might cause disease. Moreover, the striking number of the H1-associated cases, both phased haplotypes and computationally-deduced haplotypes, may note at least that these H1 cases could have the same origin.

Two flanking microsatellites, *D4S126* and *D4S3034* (Table 1), were studied in order to go into the genetic history of HD mutation in our population in depth. Allelic frequencies were estimated for these STR markers, and the pathogenic mutation was observed as being significantly associated with allele 3 of marker *D4S126* ($p<0.0001$), and also with allele 2 of marker *D4S3034* ($p<0.0001$) (Table 2). Affected chromosomes bearing other alleles for both markers were nevertheless identified as well. In fact, with regard to marker *D4S126*, the HD mutation was associated with allele 6 in nine cases, allele 5 in one case, and allele 4 in six cases. With regard to marker *D4S3034*, the HD mutation was associated with allele 3 in seven cases and allele 4 in one case (Table 2). Upon the basis of *D4S126*-rs1313770-HD/CCG-rs82334-*D4S3034*, we constructed extended haplotypes in the 47 H1 families. This phase

could be established by inheritance in only 17 out of 47 chromosomes bearing the H1 haplotype. We obtained five different haplotypes: two of them, H1A and H1D, were present in seven cases each, and the remaining definite haplotypes were constructed in one chromosome each (Fig. 3). We proceeded to computationally generate more haplotypes in the 47 H1 cases by the PHASE program. A total of 24 different haplotypes were discerned, and among them all the unambiguous haplotypes could be generated again. Moreover, haplotype H1B must be emphasised since it was yielded as one of the best construction by the program (11/74), although it was not established with total certainty in any case.

To investigate the geographical distribution of the haplotype H1, we placed the extended haplotypes H1A and H1D, the two more frequent phase-known haplotypes, and the haplotype H1F, that was once unequivocally established and computationally generated in a notable number (11/74), on a map of the Land of Valencia. The geographical distribution turned out to be non-uniform (Fig. 4). H1A was basically concentrated on the coast; H1D is located in the inner Southern counties; and finally, H1F was distributed along the Land of Valencia and in the more inner Southern regions. When estimated frequencies were considered for several of these counties, we observed that the highest ones corresponded to the areas where haplotypes H1A, H1D and H1F were localised (Fig. 4), which could provide us evidence of a narrow historical relation among these haplotypes.

Dating the mutation

In order to estimate the original time of the HD mutation associated to the H1 core haplotype actually appeared in the Land of Valencia, we applied the method developed by Serre et al. (1990), based on the intra-allelic variability at two markers for which at least one recombination between them has occurred. Depending on the mathematical model, we obtained different results. According to the first model proposed

(rs1313770_HD/CAG_rs82334), the mutation is 235 generations old; the second model (HD/CAG_rs82334_*D4S3034*) suggested the origin of the mutation occurred 350 generations ago; and finally, the third model (rs1313770_HD/CAG_*D4S3034)*, the mutation is 417 generations old.

We also performed a second analysis by means of the BDMC21 v2.1 program (Slatkin and Rannala, 1997). This approach relies on the assumption that the genetic variation among a group of highly linked polymorphic markers, defining a haplotype on which a novel non recurrent mutation arose, is a function of the mutation frequencies of those linked markers, as well as an indication of the time since this unique mutation first occurred. To achieve this approach, we considered information from multiallelic STRs *D4S126* and *D4S3034* and therefore, families with the haplotype H1 associated with the disease were taken into account. This approach showed us that HD mutation associated with the H1 haplotype is approximately 500 generations old in our population (Fig. 5). Assuming a generation time of 20 years, these results led us to estimate that the origin for this mutation is extremely ancient, between 4,700 and 10,000 years old.

De novo **mutation**

A *de novo* mutation caused by an unstable transmission, was observed in a patient with 46 CAG repeats. The mutant allele was a consequence of the meiotic expansion of a paternal 30 CAG repeat. Paternal false paternity was excluded by using ten non-linked STRs. The haplotype associated with this HD/CAG expansion was 5-A/G-7-A-3. We have not previously observed this haplotype in either phase or non-phase haplotypes associated with the HD mutation.

DISCUSSION

Classical studies supported by the variations in the geographical distribution, along with epidemiological studies, have postulated that the current HD distribution could be caused by a single ancestral mutation that would have had its origins in Western Europe, and that would have been spread worldwide through emigration (Hayden, 1981). Haplotype analyses in different populations have however shown a wide variety of HD haplotypes, that is suggestive of distinct independent mutations as the origin for disease (McDonald et al., 1992; Andrew et al., 1993; Almqvist et al., 1994 and 1995). Allele and haplotype analyses in a series of 83 familial probands from the Land of Valencia (Spain), have shown evidences which indicate that the disease had more than one origin in the population. Haplotypes in which the HD mutation was associated to the 10-CCG allele with absolute certainty, or with ambiguity, are worthy of mention. Collectively, all these patients could have the same haplotype A/G-10-C associated to disease, and since this allelic combination is not very common in the Valencian population, a single recent origin for this mutation might be suggested.

Allele and segregation analysis showed that the vast majority of chromosomes (77 out of 83 cases) causing disease bore a 7-CCG allele. Some previous reports have concluded that between 90-95% of HD patients of both Caucasian and Asiatic descent have a 7-CCG allele, while only approximately 5-10% of patients have a 10-CCG allele on the affected chromosome (Almqvist et al., 1994; Andrew et al., 1994; Squitieri et al., 1994; Yapijakis et al., 1995; Pramanik et al., 2000), which is quite similar to our data. Two haplotypes were unambiguously phased: H1 and H2 for 47 and 1 out of 83 cases, respectively. Computationally inferred haplotypes rendered two more haplotypes (H3 and H4) so that, at least, other possibilities could be computationally valid as well. If all four haplotypes are associated with HD, two hypotheses are plausible: firstly, only one mutational event occurred

on the ancestral haplotype H1, and haplotypes H2, H3 and H4 are the result of genetic events (recombination or mutation) on H1; secondly, haplotypes represent different origins as the consequence of independent mutational events, or because they come from other populations due to migratory movements. Taking into account the extremely low mutation rate of SNPs, the second hypothesis would carry more weight. However, in light of the notable number of haplotypes H1, and also because of their distribution, the first hypothesis should not be discarded. It is not possible to discern more accurately which option is correct with the data that is available. Similar results have been reported in previous studies on the genetic history and evolution of the HD mutation (McDonald et al. 1992; Andrew et al. 1993; Morrison et al., 1993, 1995; Squitieri et al. 1994). In Sweden, segregation analyses showed that 89% of HD families shared the same haplotype constructed with two polymorphic markers within the *huntingtin* gene, so that at least three origins for the HD mutation were suggested in the aggregate (Almqvist et al., 1994 and 1995). In the same way, a similar finding has been reported in other disorders associated with dynamic mutations. In Friedreich ataxia, more than 50% of the GAA expanded chromosomes were associated with a single haplotype in families of a European origin, whereas the rest of the expanded alleles segregate were associated with haplotypes that could be derived from the founder one by genetic mechanisms such as recombination or mutation (Monrós et al., 1996; Cossée et al., 1997).

When haplotypes H1 were discerned by including microsatellites, six variant haplotypes were identified. These extended haplotypes have shown a non-homogeneous geographical distribution (Fig. 4). In fact, it is striking that many HD cases sharing a common core haplotype H1 are placed in a reduced geographical area in the Land of Valencia. Haplotype H1A is basically placed in two coastal regions: *La Marina Alta* and *La Safor*. Haplotype H1D is more distributed and it could be in favour of different timings. However, since haplotype H1D is mainly found in inner countryside regions, concentrated in *l'Alcoià*

and its natural routes of communication, this wider distribution could be a consequence of migratory movements rather than the multiple origins of HD mutation. Something similar could have occurred to haplotype H1F. This is situated throughout the Land of Valencia, and in the most inner Southern counties, mainly in the *Vinalopó* region. Moreover, although a complete census was not available, according to the origin of the studied HD families, we have found that the three more representative haplotypes in our series (H1A, H1D and H1F) are located in counties with high estimated frequenciess. This high frequency may therefore be the result of a founder effect (Fig. 4).

To further investigate the natural history of HD in our population, we estimated the age of the founding haplotype H1 and the associated HD expansion by using two mathematical approaches. Estimations made with computer iterations based on the method developed by Serre et al. (1990), gave a wide range of generations, between 235 and 417 (4,700 and 8,340 years old, based on the assumption that a generation lasts 20 years). This wide range of generations is not unique to the present results. In fact, the different mutations ages were estimated by using Risch's formula (Risch et al., 1995), which assumes an exponential decay of linkage disequilibrium with time, as shown by Serre's method (Serre et al., 1990), where a large variability was also shown, depending on the markers used for the dating analysis (Díaz et al., 1999; Hashemi et al., 2001). When the method implemented by Slatkin and Rannala (1997) was applied, an age estimation of approximately 500 generations was obtained (10,000 years). The variability obtained is subjected to a statistical fluctuation resulting from the methods used, which in turn depend strongly on genetic (mutation rate, selection) and demographic (population size, expansion dynamics) parameters (Díaz et al., 1999; Ciotti et al., 2000; Broeks et al., 2003). In the aggregate, all these calculations led us to conclude that the most likely scenario is that the time when the founding haplotype H1, associated with the HD mutation, actually appeared in the Land of Valencia is extremely

ancient. This is also supported by the low frequency of the H1 haplotype in the Valencian control chromosomes. Reported data on dating a HD mutation are not available. On the other hand, it is noteworthy to mention the results obtained by Bachinski et al. (2003) on the evolution of the type 2 myotonic dystrophy (DM2) expansion mutation in geographically separate populations of a European origin. They estimated the age of both the founding haplotype and the DM2 mutation to be ~4,000-11,000 years. In both cases the dynamic mutation would give arise after the Neolithic expansion of the modern humans into Europe, roughly 10,000-15,000 years ago (Cavalli-Sforza et al., 1994). At this point it is intriguing to set out the possibility that most of HD mutations in European populations came about in that period of the European history. Further studies on the allelic age of the HD expansion are needed.

De novo mutations in diseases associated with dynamic expansions used to be originated from large normal alleles that may act as a reservoir (Imbert et al., 1993). We have found one *de novo* CAG expansion in the present series. The mutation is associated with a rare haplotype 5-A/G-7-A-3 at first sight non-related with the ancestral haplotype H1. Despite the ambiguity at marker rs1313770 however, it is still possible that this rare haplotype might derived from H1. In such a case, it could be argued that the HD chromosome associated with H1 arrived into Valencia as a large normal allele that underwent further mutated expansion. This ancient large allele would have remained as the reservoir for the disease through the Valencian history.

REFERENCES

Almqvist E, Andrew S, Theilmann J, Goldberg P, Zeisler J, Drugge U, Grandell U, Tapper-Persson M, Winblad B, Hayden M. 1994. Geographical distribution of haplotypes in Swedish families with Huntington's disease. Hum Genet 94: 125-128.

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 Δ2642 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: 207-214.

Andrew S, Theilmann J, Almqvist E, Norremolle A, Lucotte G, Anvret M, Sorensen SA, Turpin JC, Hayden MR. 1993. DNA analysis of distinct populations suggests multiple origins for the mutation causing Huntington disease. Clin Genet 43: 286-294.

Andrew S, Goldberg YP, Theilmann J, Zeisler J, Hayden MR. 1994. A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. Hum Mol Genet 3: 65-67.

Ashley CT, Warren ST. 1995. Trinucleotide repeat expansion and human disease. Annu Rev Genet 29: 703-728.

Bachinsky LL, Udd B, Meola G, Sansone V, Bassez G, Eymard B, Thornton CA, Moxley RT, Harper PS, Rogers MT, Jurkat-Rott K, Lehmann-Horn F, Wieser T, Gámez J, Navarro C, Bottani A, Kohler A, Shriver MD, Sallinen R, Wessman M, Zhang S, Wright FA, Krahe R. 2003. Confirmation of the type 2 myotonic dystrophy (CCTG)n expansion mutation in patients with proximal myotonic myopathy/proximal myotonic dystrophy of different European origins: A single shared haplotype indicates an ancestral founder effect. Am J Hum Genet 73: 835-848.

Broeks A, Urbanus JHM, de Knijff P, Devilee P, Nicke M, Klöpper K, Dörk T, Floore AN, van't Veer LJ. 2003. IVS10-6T>G, an ancient ATM germline mutation linked with breast cancer. Hum Mut 21: 521-528.

Burguera JA, Solís P, Salazar A. 1997. Estimación de la prevalencia de la enfermedad de Huntington en la Comunidad Valenciana mediante el método de captura-recaptura. Rev Neurol (Barc) 25: 1845-1847.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Pricenton, New York: Pricenton University Press.

Chong SS, Almqvist E, Telenius H, LaTray L, Nichol K, Bourdelat-Parks B, Goldberg YP, Haddad BR, Richards F, Silence D, Greenberg CR, Ives E, Van den Engh G, Hughes MR, Hayden MR. 1997. Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analyses. Hum Mol Genet 6: 301-309.

Ciotti P, Struewing JP, Mantelli M, Chompret A, Avril M-F, Santi PL, Tucker MA, Bianchi-Scarrà G, Bressac-de-Paillerets B, Goldstein AM. 2000. A single genetic origin for the G101W CDKN2A mutation in 20 melanoma-prone families. Am J Hum Genet 67: 311-319.

Cossée M, Schmitt M, Campuzano V, Reutenauer L, Moutou C, Mandel J-L, Koening M. 1997. Evolution of the Friedreich's ataxia trinucleotide repeat expansion: Founder effect and premutations. Proc Natl Acad Sci USA 94: 7452-7457.

Costa MC, Magalhaes P, Ferreirinha F, Guimaraes L, Januário C, Gaspar I, Loureiro L, Vale J, Garrett C, Regateiro F, Magalhaes M, Sousa A, Maciel P, Sequeiros J. 2003. Molecular diagnosis of Huntington disease in Portugal: implications of genetic counselling and clinical practice. Eur J Hum Genet 11: 872-878.

Cummings CJ, Zoghbi HY. 2000. Fourteen and counting: unravelling trinucleotide repeat diseases. Hum Mol Genet 9: 909-916.

De Rooij KE, De Koning Gans PA, Skraastad MI, Belfroid RD, Vegter-Van Der Vlis M, Roos RA, Bakker E, Van Ommen GJ, Den Dunnen JT, Losekoot M. 1993. Dynamic mutation in Dutch Huntington's disease patients: increased paternal repeat instability extending to within the normal size range. J Med Genet 30: 996-1002.

Díaz A, Montfort M, Cormand B, Zeng B, Pastores GM, Chabás A, Vilageliu L, Grinberg D. 1999. Gaucher disease: The N370S mutation in Ashkenazi Jewish and Spanish patients has a common origin and arose several thousand years ago. Am J Hum Genet 64: 1233-1238.

Goldberg YP, Kremer B, Andrew SE, Theilmann J, Graham RK, Squitieri F, Telenius H, Adam S, Sajoo A, Starr E. 1993. Molecular analysis of new mutations for Huntington’s disease: intermediate alleles and sex of origin effects. Nat Gent 5: 174-179.

Harper PS. 1996. Huntington’s disease. Major problems in neurology. London, Philadelphia, Toronto, Sidney, Tokio: Saunders.

Hashemi J, Bendahl P-O, Sandberg T, Platz A, Linder S, Stierner U, Olsson H, Ingvar C, Hansson J, Borg A. 2001. Haplotype analysis and age estimation of the 113insR CDKN2A Founder mutation in Swedish melanoma families. Genes, Chromosomes and Cancer 31: 107-116.

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

Imbert G, Kretz C, Johnson K, Mandel JL. 1993. Origin of the expansion mutation in myotonic dystrophy. Nat Genet 4: 72-76.

Kremer B, Goldberg P, Andrew SE, Theilmann J, Telenius H, Zeisler J, Squitieri F, Lin B, Bassett A, Almqvist E, Bird TD, Hayden MR. 1994. A worldwide study of Huntington’s

disease mutation: The sensitivity and specify of measuring CAG repeats. N Engl J Med 330: 1401-1406.

Kunst CB, Warren ST. 1994. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. Cell 77: 853-861.

Labuda M, Labuda D, Miranda C, Poirier J, Soong B-W, Barucha NE, Pandolfo M. 2000. Unique origin and specific ethnic distribution of the Friedreich ataxia GAA expansion. Neurology 54: 2322-2324.

Liquori CL, Ikeda Y, Weatherspoon M, Ricker K, Schoser BGH, Dalton JC, Day JW, Ranum PW. 2003. Myotonic Dystrophy type 2: Human Founder Haplotype and Evolutionary Conservation of the Repeat Tract. Am J Hum Genet 73: 849-862.

Martin JB, Gusella JF. 1986. Huntington's disease: pathogenesis and management. N Engl J Med 315: 1267-1276.

McDonald 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 disease candidate region exhibits many different haplotypes. Nat Genet 1: 99-103.

Monrós E, Cañizares J, Moltó MD, Rodius F, Montermini L, Cossée M, Martínez F, Prieto F, de Frutos R, Koenig M, Pandolfo M, Bertranpetit J, Palau F. 1996. Evidence for a common origin of most Friedreich ataxia chromosomes in the Spanish population. Eur J Hum Genet. 4:191-198.

Morrison PJ, Graham CA, Nevin NC. 1993. Significant linkage disequilibrium between the Huntington's disease locus and markers at loci D4S10, D4S95, and D4S111 in Northern Ireland. J Med Genet 30: 1018-1019.

Morrison PJ, Johnston WP, Nevin NC. 1995. The epidemiology of Huntington's disease in Northern Ireland. J Med Genet 32: 524-530.

Myers RH, MacDonald ME, Koroshetz WJ, Duyao MP, Ambrose CM, Taylor SA, Barnes G, Srinidhi J, Lin CS, Whaley WL. 1993. De novo expansion of a (CAG)n repeat in sporadic Huntington’s disease. Nat Genet 5: 168-173.

Paulson HL, Fischbeck KH. 1996. Trinucleotide repeats in neurogenetic disease. Annu Rev Neurosci 19: 79-107.

Pramanik S, Basu P, Gangopadhaya PK, Sinha KK, Jha DK, Sinha S, Das SK, Maity BK, Mukherjee SC, Roychoudhuri S, Majumder PP, Bhattacharyya NP. 2000. Analysis of CAG and CCG repeats in Huntingtin gene among HD patients and normal populations of India. Eur J Hum Genet 8: 678-682

Richards RI, Holman K, Friend K, Kremer E, Hillen D, Staples A, Brown WT, Goonewardena P, Tarleton J, Schwartz C, Sutherland GR. 1992. Evidence of founder chromosomes in fragile X syndrome. Nat Genet 1: 257-260.

Richards RI. 2001. Dynamic mutations: a decade of unstable expended repeats in human genetic disease. Hum Mol Genet 10: 2187-2194.

Risch N, de Leon D, Ozelius L, Kramer P, Almasy L, Singer B, Fahn S, Breakefield X, Bressman S. 1995. Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. Nature Genet 9: 152-159.

Serre JL, Simon-Bouy B, Mornet E, Jaume-Roig B, Balassopoulou A, Schwartz M, Taillandier A, Boué J, Boué A. 1990. Studies of RFLP closely linked to the cystic fibrosis locus throughout Europe lead to new considerations in populations genetics. Hum Genet 84: 449-454.

Slatkin M, Rannala B. 1997. Estimating the age of alleles by use of intraallelic. Am J Hum Genet 60: 447-458.

Squitieri F, Andrew SE, Goldberg YP, Kremer B, Spence N, Zeisler J, Nichol K, Theilmann J, Greenberg, 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's disease reveals clues to the origin and mechanisms of CAG expansion and reasons for geographic variations of estimated frequency. Hum Mol Genet 3: 2103-2114.

The American College of Medical Genetics / American Society of Human Genetics / Huntington's Disease Genetic Testing Working Group. 1998. Laboratory Guidelines for Huntington's Disease Genetic Testing. Am J Hum Genet 62: 1243-1247.

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

Wells RD, Warren ST, Sarmiento M. 1998. Genetic Instabilities and Hereditary Neurological Disease. San Diego: Academic.

Yapijakis C, Vassilopoulos D, Tzagournisakis M, Maris T, Fesdjian C, Papageorgiou C, Plaitakis A. 1995. Linkage disequilibrium between the expanded (CAG)n repeat and an allele of the adjacent (CCG)n repeat in Huntington's disease patients of Greek origin. Eur J Hum Genet 3: 228-234

Yoon SR, Dubeau L, de Young M, Wexler NS, Arnheim N. 2003. Huntington disease expansion mutations in humans can occur before meiosis is completed. Proc Natl Acad Sci U S A 100: 8834-8838

Zühlke C, Riess O, Bockel B, Lange H, Thies U. 1993. Mitotic stability and meiotic variability of the $(CAG)_n$ repeat in the HD gene. Hum Mol Genet 2: 2063-2067.

Locus	Marker Systematic name[a]	Sequence (5'→3')	PCR size (bp)	Heterozigosity (alleles)
D4S126	dinucleotide	HEX-GGATCCTGTCACTGTACTCCAGCC TGCTTAACCAGTTTGACCATGAGG	155-177	83.50 % (12)
rs1313770	SNP NT_006081.17:g.444387A>G	CCAAGAGAGGACTTATCC CTGTCAGAAGTGGGATCC	138	---- (2)
HD/CCG	trinucleotide	AGCAGCAGCAACAGCC HEX-GGCTGAGGAAGCTGAGGAG	61-79	---- (6)
rs82334	SNP NT_006081.17:g.611947A>C	GCTGCTTGGAGCAGCAGC GGAGGCCACCTTTGGGTC	219	37.8 % (2)
D4S3034	dinucleotide	CTGCCAATAAACTGGGT TET-TTGCTCACCAAAGAGGTT	180-188	62 % (5)

[a]The nomenclature format followed is a GenBank accession number and version number, the letter "g" what means it is a genomic sequence, the position number of the change and the nucleotide change itself.

TABLE 1. Characteristics of the analysed markers. Data were obtained from http://genome.ucsc.edu and http://www.ncbi.nlm.nih.gov/entrez.

Marker	Allele	Control chromosomes	HD chromosomes
		N (%)	N (%)
D4S126	149 (12)	1 (0.5)	
	151 (11)		
	153 (10)	7 (3.5)	
	155 (9)	20 (10)	
	157 (8)	26 (13)	
	159 (7)	38 (19)	
	161 (6)	44 (22)	9 (32.2)
	163 (5)	24 (12)	1 (3.6)
	165 (4)	30 (15)	6 (21.4)
	167 (3)	9 (4.5)	12 (42.8)
	169 (2)	1 (0.5)	
	Total	200	28
rs1313770	A	111 (55.5)	50 (98)
	G	89 (44.5)	1 (2)
	Total	200	51
CCG repeat	6	1 (0.5)	
	7	147 (74.3)	77 (97.5)
	8	3 (1.5)	
	9	2 (1.0)	
	10	45 (22.7)	2 (2.5)
	Total	198	79
rs82334	A	143 (71.5)	63 (95.5)
	C	57 (28.5)	3 (4.5)
	Total	200	66
D4S3034	180 (1)	22 (11.1)	
	182 (2)	74 (37.4)	23 (74.2)
	184 (3)	77 (38.9)	7 (22.6)
	186 (4)	23 (11.6)	1 (3.2)
	188 (5)	2 (1.0)	
	Total	198	31

TABLE 2. Allelic distribution of the five analysed markers in both controls and the HD population. Only unequivocal-phased chromosomes associated with the HD mutation were considered.

FIGURE 1. Distribution of the CAG-HD repeat in all the alleles genotyped. The CAG repeat number is represented on the X axis, and the number of alleles detected for each repetition is represented on the Y axis.

FIGURE 2. Physical map of HD locus (NM_002111.3). *Huntingtin* gene is indicated as an arrow. Distances between each pair of markers are indicated at the top. Markers are placed at the bottom.

FIGURE 3. Distribution of HD-associated haplotypes segregating with the allele 7 of the HD/CCG polymorphism. The number of resulting cases for each haplotype either established by inheritance (known phase) or reconstructed by the PHASE v2.0 program (unknown phase) is shown. (**A)** Main haplotypes constructed with markers rs1313770-HD/CCG-rs82334; (**B)** Haplotypes based on the core H1 (A-7-A). A total of 17 extended haplotypes associated to the HD mutation were established by inheritance. Haplotype H1B is also included because it was yielded by the computational analysis as one of the best reconstructions. Vertical bars denote hypothetical recombinational events.

FIGURE 4. Location of the HD cases in the Land of Valencia. (**A)** Geographical distribution of the haplotypes H1A, H1D and H1F. Both the phased (grey symbols) and unambiguous (black symbols) haplotypes are placed. (**B)** Estimated frequencies estimated for several Valencian counties.

FIGURE 5. Maximum likelihood curves obtained by means of BDMC21. Graphics shows the log-likelihood values and the corresponding number of generation estimates regarding the

age of the founding haplotype H1. Up to 10 seeds were performed to validate the resulting data, and no significant differences were observed.