

A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes

The Huntington's Disease Collaborative Research Group*

Summary

The Huntington's disease (HD) gene has been mapped in 4p16.3 but has eluded identification. We have used haplotype analysis of linkage disequilibrium to spotlight a small segment of 4p16.3 as the likely location of the defect. A new gene, IT15, isolated using cloned trapped exons from the target area contains a polymorphic trinucleotide repeat that is expanded and unstable on HD chromosomes. A (CAG)_n repeat longer than the normal range was observed on HD chromosomes from all 75 disease families examined, comprising a variety of ethnic backgrounds and 4p16.3 haplotypes. The (CAG)_n repeat appears to be located within the coding sequence of a predicted ~348 kd protein that is widely expressed but unrelated to any known gene. Thus, the HD mutation involves an unstable DNA segment, similar to those described in fragile X syndrome, spino-bulbar muscular atrophy, and myotonic dystrophy, acting in the context of a novel 4p16.3 gene to produce a dominant phenotype.

Introduction

Huntington's disease (HD) is a progressive neurodegenerative disorder characterized by motor disturbance, cognitive loss, and psychiatric manifestations (Martin and Gusella, 1986). It is inherited in an autosomal dominant fashion and affects ~1 in 10,000 individuals in most populations of European origin (Harper et al., 1991). The hallmark of HD is a distinctive choreic movement disorder that typically has a subtle, insidious onset in the fourth to fifth decade of life and gradually worsens over a course of 10 to 20 years until death. Occasionally, HD is expressed in juveniles, typically manifesting with more severe symptoms including rigidity and a more rapid course. Juvenile onset of HD is associated with a preponderance of paternal transmission of the disease allele. The neuropathology of HD also displays a distinctive pattern, with selective loss of neurons that is most severe in the caudate and putamen. The biochemical basis for neuronal death in HD has not yet been explained, and there is consequently no treatment effective in delaying or preventing the onset and progression of this devastating disorder.

The genetic defect causing HD was assigned to chromosome 4 in 1983 in one of the first successful linkage analyses using polymorphic DNA markers in humans (Gusella

*The Huntington's Disease Collaborative Research Group comprises:

Group 1:

Marcy E. MacDonald,¹ Christine M. Ambrose,¹ Mabel P. Duyao,¹ Richard H. Myers,² Carol Lin,¹ Lakshmi Srinidhi,¹ Glenn Barnes,¹ Sherryl A. Taylor,¹ Marianne James,¹ Nicolet Groot,¹ Heather MacFarlane,¹ Barbara Jenkins,¹ Mary Anne Anderson,¹ Nancy S. Wexler,³ and James F. Gusella^{1†}

¹Molecular Neurogenetics Unit
Massachusetts General Hospital
and Department of Genetics
Harvard Medical School
Boston, Massachusetts 02114

²Department of Neurology
Boston University Medical School
Boston, Massachusetts 02118

³Hereditary Disease Foundation
1427 7th Street, Suite 2
Santa Monica, California 90401

Group 2:

Gillian P. Bates, Sarah Baxendale, Holger Hummerich, Susan Kirby, Mike North, Sandra Youngman, Richard Mott, Gunther Zehetner, Zdenek Sedlacek, Annemarie Poustka, Anna-Maria Frischauf, and Hans Lehrach

Genome Analysis Laboratory
Imperial Cancer Research Fund
Lincoln's Inn Fields
London, WC2A 3PX, England

Group 3:

Alan J. Buckler,¹ Deanna Church,¹ Lynn Doucette-Stamm,¹ Michael C. O'Donovan,¹

Laura Riba-Ramirez,¹ Manish Shah,¹ Vincent P. Stanton,¹ Scott A. Strobel,² Karen M. Draths,² Jennifer L. Wales,² Peter Dervan,² and David E. Housman¹

¹Center for Cancer Research
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

²Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, California 91125

Group 4:

Michael Altherr, Rita Shiang, Leslie Thompson, Thomas Fielder, and John J. Wasmuth
Department of Biological Chemistry
University of California
Irvine, California 92717

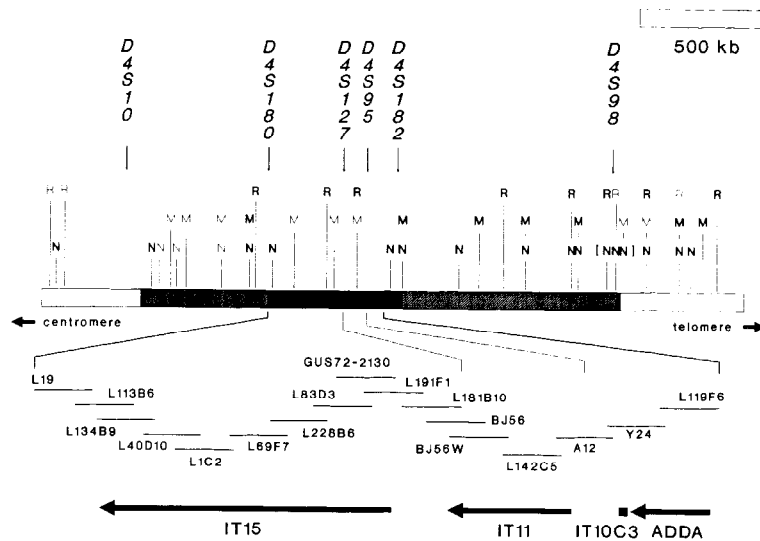
Group 5:

Danilo Tagle, John Valdes, Lawrence Elmer, Marc Allard, Lucio Castilla, Manju Swaroop, Kris Blanchard, and Francis S. Collins
Department of Internal Medicine and Human Genetics
and The Howard Hughes Medical Institute
University of Michigan
Ann Arbor, Michigan 48109

Group 6:

Russell Snell, Tracey Holloway, Kathleen Gillespie, Nicole Datson, Duncan Shaw, and Peter S. Harper
Institute of Medical Genetics
University of Wales College of Medicine
Cardiff, CF4 4XN, Wales

[†]Correspondence should be addressed to James F. Gusella.



and NruI (R). Sites displaying complete digestion are shown in boldface, while sites subject to frequent incomplete digestion are shown as lighter symbols. Brackets around the N symbols indicate the presence of additional clustered NotI sites.

Figure 1. Long-Range Restriction Map of the HD Candidate Region

A partial long-range restriction map of 4p16.3 is shown (adapted from Lin et al. [1991]). The HD candidate region determined by recombination events is depicted by hatched bars between D4S10 and D4S98. The portion of the HD candidate region implicated as the site of the defect by linkage disequilibrium haplotype analysis (MacDonald et al., 1992) is shown as a closed bar. Below the schematic map, the region from D4S180 to D4S182 is expanded to show the cosmid contig (averaging 40 kb per cosmid). The genomic coverage and, where known, the transcriptional orientation (arrows, 5' to 3') of the IT15, IT11, IT10C3, and ADDA genes is also shown. Locus names above the map denote selected polymorphic markers that have been used in HD families. The positions of D4S127 and D4S95, which form the core of haplotype in the region of maximum disequilibrium, are also shown in the cosmid contig. Restriction sites are given for NotI (N), MluI (M),

et al., 1983). Since that time, we have pursued a location cloning approach to isolating and characterizing the HD gene based on progressively refining its localization (Gusella, 1989, 1991). Among other work, this has involved the generation of new genetic markers in the region by a number of techniques (Pohl et al., 1988; Whaley et al., 1991; MacDonald et al., 1989a), the establishment of genetic (MacDonald et al., 1989b; Allitto et al., 1991) and physical maps of the implicated regions (Bucan et al., 1990; Bates et al., 1991; Doucette-Stamm et al., 1991; Altherr et al., 1992), the cloning of the 4p telomere of an HD chromosome in a yeast artificial chromosome clone (Bates et al., 1990; Youngman et al., 1992), the establishment of yeast artificial chromosome (Bates et al., 1992) and cosmid (S. B. et al., unpublished data) contigs of the candidate region, as well as the analysis and characterization of a number of candidate genes from the region (Thompson et al., 1991; Taylor et al., 1992; Ambrose et al., 1992; M. P. D. et al., submitted). Analysis of recombination events in HD kindreds has identified a candidate region of 2.2 Mb, between D4S10 and D4S98 in 4p16.3, as the most likely position of the HD gene (MacDonald et al., 1989b; Bates et al., 1991; Snell et al., 1992). Investigations of linkage disequilibrium between HD and DNA markers in 4p16.3 (Snell et al., 1989; Theilman et al., 1989) have suggested that multiple mutations have occurred to cause the disorder (MacDonald et al., 1991). However, haplotype analysis using multiallele markers has indicated that at least one-third of HD chromosomes are ancestrally related (MacDonald et al., 1992). The haplotype shared by these HD chromosomes indicates that a 500 kb segment between D4S180 and D4S182 is the most likely site of the genetic defect.

Targeting this 500 kb region for saturation with gene

transcripts, we have used exon amplification as a rapid method for obtaining candidate coding sequences (Buckler et al., 1991). This strategy has previously identified three genes: the α -adducin gene (ADDA) (Taylor et al., 1992) and a putative novel transporter gene (IT10C3) in the distal portion of this segment (M. P. D. et al., submitted), and a novel G protein-coupled receptor kinase gene (IT11) in the central portion (Ambrose et al., 1992). However, no defects implicating any of these genes as the HD locus have been found. We have now applied the exon amplification approach to the proximal portion of the 500 kb segment. We have identified a large gene, IT15, spanning ~210 kb, that encodes a previously undescribed protein of ~348 kd. The IT15 reading frame contains a polymorphic (CAG)_n trinucleotide repeat with at least 17 alleles in the normal population, varying from 11 to 34 CAG copies. On HD chromosomes, the length of the trinucleotide repeat is substantially increased, to a range of 42 to over 66 copies, and shows an apparent correlation with age of onset, the longest segments being detected in juvenile HD cases. The instability in the length of the repeat is reminiscent of similar trinucleotide repeats in the fragile X syndrome and in myotonic dystrophy (Suthers et al., 1992). The presence of an unstable, expandable trinucleotide repeat on HD chromosomes in the region of strongest linkage disequilibrium with the disorder suggests that this alteration underlies the dominant phenotype of HD and that IT15 encodes the HD gene.

Results

Application of Exon Amplification to Obtain Trapped, Cloned Exons

The HD candidate region defined by discrete recombina-

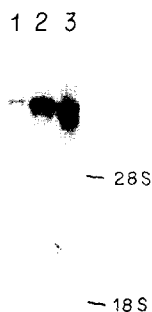


Figure 2. Northern Blot Analysis of the IT15 Transcript

Results of the hybridization of IT15A to a Northern blot of RNA from normal (lane 1) and *HD* homozygous (lanes 2 and 3) lymphoblasts are shown. A single RNA of ~11 kb was detected in all three samples, with slight apparent variations being due to unequal RNA concentrations. The *HD* homozygotes are independent, deriving from an American family (lane 2) and the large Venezuelan family (lane 3), respectively. The Venezuelan *HD* chromosome has a 4p16.3 haplotype of 5 2 2 defined by a (GT)_n polymorphism at *D4S127* and variable number tandem repeat and TaqI restriction fragment length polymorphisms at *D4S95*. The American homozygote carries the most common 4p16.3 haplotype found on *HD* chromosomes: 2 11 1 (MacDonald et al., 1992).

tion events in well-characterized families spans 2.2 Mb between *D4S10* and *D4S98* as shown in Figure 1. The 500 kb segment between *D4S180* and *D4S182* displays the strongest linkage disequilibrium with *HD*, with about one-third of disease chromosomes sharing a common haplotype, anchored by multiallele polymorphisms at *D4S127* and *D4S95* (MacDonald et al., 1992). We have isolated 64 overlapping cosmids, spanning ~480 kb from *D4S180* to a location between *D4S95* and *D4S182*, by a combination of information from yeast artificial chromosome (Baxendale et al., 1991) and cosmid probe hybridiza-

tion to high density filter grids of a chromosome 4-specific library, as well as of additional libraries covering this region. Sixteen of these cosmids providing the complete contig are shown in Figure 1. We have previously used exon amplification to identify *ADDA*, *IT10C3* (a novel putative transporter gene), and *IT11* (a novel G protein-coupled receptor kinase gene) in the region distal to *D4S127* (Figure 1).

We have now applied the exon amplification technique to cosmids from the region of the contig proximal to *D4S127*. This procedure produces trapped exon clones, which can represent single exons or multiple exons spliced together, and is an efficient method for obtaining probes for screening cDNA libraries. Individual cosmids were processed, yielding nine exon clones in the region from cosmids L134B9 to L181B10.

Identification of the IT15 Gene

Two nonoverlapping cDNAs were initially isolated using exon probes. *IT15A* was obtained by screening a transformed adult retinal cell cDNA library with exon clone DL118F5-U. *IT16A* was isolated by screening an adult frontal cortex cDNA library with a pool of three exon clones, DL83D3-8, DL83D3-1, and DL228B6-3. By Northern blot analysis, we discovered that *IT15A* and *IT16A* both detected an ~10–11 kb transcript, suggesting that they derive from the same mRNA. Figure 2 shows an example of a Northern blot containing RNA from lymphoblastoid cell lines representing a normal individual and two independent homozygotes for *HD* chromosomes of different haplotypes. The same ~10–11 kb transcript was also detected in RNA from a variety of human tissues (liver, spleen, kidney, muscle, and various regions of adult brain).

IT15A and *IT16A* were used to “walk” in a number of human tissue cDNA libraries in order to obtain the full-length transcript. Figure 3 shows a representation of five cDNA clones that define the *IT15* transcript, under a sche-

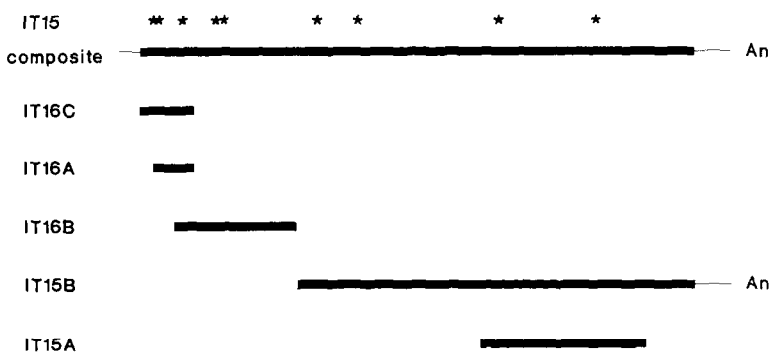


Figure 3. Schematic of cDNA Clones Defining the *IT15* Transcript

Five cDNAs are represented under a schematic of the composite *IT15* sequence. The thin line corresponds to untranslated regions. The thick line corresponds to coding sequence, assuming initiation of translation at the first Met codon in the open reading frame. Stars mark the positions of the following exon clones 5' to 3': DL83D3-8, DL83D3-1, DL228B6-3, DL228B6-5, DL228B6-13, DL69F7-3, DL178H4-6, DL118F5-U, and DL134B9-U4. The composite sequence was derived as follows. From 22 bases 3' to the putative initiator Met ATG, the sequence was compiled from the cDNA clones and exons

shown. There are 9 bases of sequence intervening between the 3' end of *IT16B* and the 5' end of *IT15B*. These were identified by PCR amplification of first-strand cDNA and sequencing of the PCR product. At the 5' end of the composite sequence, the cDNA clone *IT16C* terminates 27 bases upstream of the (CAG)_n. However, when *IT16C* was identified, we had already generated genomic sequence surrounding the (CAG)_n in an attempt to generate new polymorphisms. This sequence matched the *IT16C* sequence and extended it 337 bases upstream, including the apparent Met initiation codon.

matic of the composite sequence derived as described in the figure legend. Figure 3 also displays the locations on the composite sequence of the nine trapped exon clones.

The composite sequence of IT15, containing the entire predicted coding sequence, spans 10,366 bases, including a tail of 18 A's as shown in Figure 4. An open reading frame of 9432 bases begins with a potential initiator methionine codon at base 316, located in the context of an optimal translation initiation sequence. An in-frame stop codon is located 240 bases upstream of this site. The protein product of IT15 is predicted to be a 348 kd protein containing 3144 amino acids. Although we have chosen the first Met codon in the long open reading frame as the probable initiator codon, we cannot exclude the possibility that translation does not actually begin at a more 3' Met codon, producing a smaller protein.

Polymorphic Variation of the (CAG)_n Trinucleotide Repeat

Near its 5' end, the IT15 sequence contains 21 copies of the triplet CAG, encoding glutamine (Figure 5). When this sequence was compared with our collection of genomic sequences surrounding simple sequence repeats in 4p16.3, we found that normal cosmid L191F1 had 18 copies of the triplet, indicating that the (CAG)_n repeat is polymorphic (Figure 5). We chose primers from the genomic sequence flanking the repeat to establish a polymerase chain reaction (PCR) assay for this variation. In the normal population, this simple sequence repeat polymorphism displays at least 17 discrete alleles, ranging from about 11 to 34 repeat units (Table 1). Ninety-eight percent of the 173 normal chromosomes tested contained repeat lengths between 11 and 24 repeats. Two chromosomes were detected in the 25–30 repeat range and 2 normal chromosomes had 33 and 34 repeats, respectively. The overall heterozygosity on normal chromosomes was 80%. We presume, based on sequence analysis of three clones, that the variation is based entirely on the (CAG)_n, but we cannot exclude the potential for variation of the smaller downstream (CCG)₇, which is also included in the PCR product.

Instability of the Trinucleotide Repeat on HD Chromosomes

Sequence analysis of cosmid GUS72-2130, derived from a chromosome with the major HD haplotype (see below), revealed 48 copies of the trinucleotide repeat, far more than the number of copies in the largest normal allele (Figure 5). When the PCR assay was applied to HD chromosomes, a pattern strikingly different from the normal variation was observed. HD heterozygotes contained one discrete allelic product in the normal size range and one PCR product of much larger size, suggesting that the (CAG)_n repeat on HD chromosomes is expanded relative to normal chromosomes.

Figure 6 shows the patterns observed when we performed the PCR assay on lymphoblast DNA from a selected nuclear family in a large Venezuelan HD kindred. In this family, DNA marker analysis has shown previously that the HD chromosome was transmitted from the father

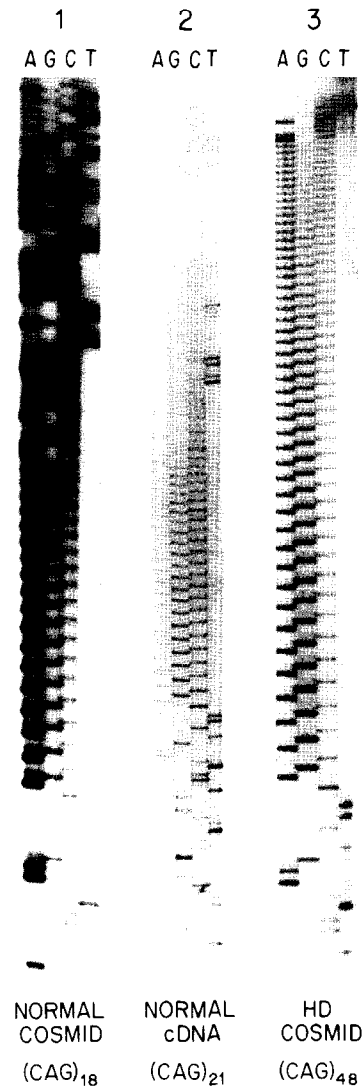


Figure 5. DNA Sequence Analysis of the (CAG)_n Repeat
DNA sequence shown in panels 1, 2, and 3 demonstrates the variation in the (CAG)_n repeat detected in normal cosmid L191F1 (1), cDNA IT16C (2), and HD cosmid GUS72-2130. Panels 1 and 3 were generated by direct sequencing of cosmid subclones using the primer 5'-GGCGGGAGACCGCCATGGCG-3'. Panel 2 was generated using the pBSKII T7 primer 5'-AATACGACTCACTATAG-3'.

Table 1. Comparison of HD and Normal Repeat Length

Range of Allele Sizes (Number of Repeats)	Normal Chromosomes		HD Chromosomes	
	Number	Frequency	Number	Frequency
≥48	0	0	44	0.59
42–47	0	0	30	0.41
30–41	2	0.01	0	0
25–30	2	0.01	0	0
≤24	169	0.98	0	0
Total	173	1.00	74	1.0

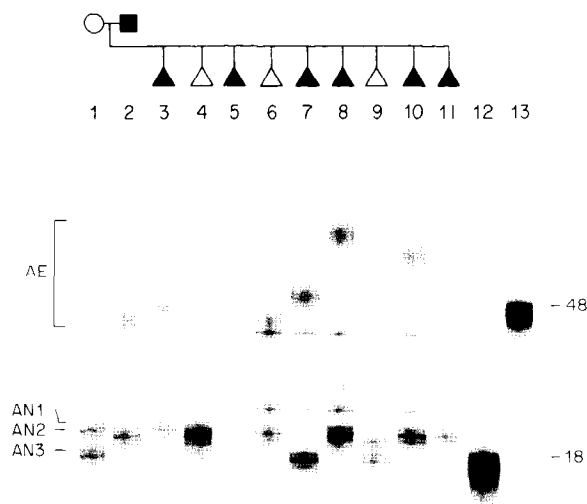


Figure 6. PCR Analysis of the (CAG)_n Repeat in a Venezuelan HD Sibship with Some Offspring Displaying Juvenile Onset

Results of PCR analysis of a sibship in the Venezuelan HD pedigree are shown. Affected individuals are represented by closed symbols. Progeny are shown as triangles, and the birth order of some individuals has been changed for confidentiality. AN1, AN2, and AN3 mark the positions of the allelic products from normal chromosomes. AE marks the range of PCR products from the *HD* chromosome. The intensity of background constant bands, which represent a useful reference for comparison of the above PCR products, varies with slight differences in PCR conditions. The PCR products from cosmids L191F1 and GUS72-2130 are loaded in lanes 12 and 13 and have 18 and 48 CAG repeats, respectively.

(lane 2) to seven children (lanes 3, 5, 6, 7, 8, 10, and 11). The three normal chromosomes present in this mating yielded a PCR product in the normal size range (AN1, AN2, and AN3) that was inherited in a Mendelian fashion. The *HD* chromosome in the father yielded a diffuse, fuzzy PCR product slightly smaller than the 48 repeat product of our non-Venezuelan *HD* cosmid. Except for the DNA in lane 5, which did not PCR amplify, and in lane 11, which displayed only a single normal allele, each of the affected children's DNAs yielded a PCR product of a different size (AE), indicating instability of the *HD* chromosome (CAG)_n repeat. Lane 6 contained an *HD*-specific product slightly smaller than or equal to that of the father's DNA. Lanes 3, 7, 10, and 8, respectively, contained *HD*-specific PCR products of progressively larger size. The absence of an *HD*-specific PCR product in lane 11 suggested that this child's DNA possessed a (CAG)_n repeat that was too long to amplify efficiently. This was verified by Southern blot analysis in which the expanded *HD* allele was easily detected and estimated to contain up to 100 copies of the repeat. Notably, this child had juvenile onset of HD at the very early age of 2 years. The onset of HD in the father was when he was in his early 40s, typical of most adult HD patients in this population. The onset ages of the children represented by lanes 3, 7, 10, and 8 were 26, 25, 14, and 11 years, respectively, suggesting a rough correlation between age at onset of HD and the length of the (CAG)_n repeat on the *HD* chromosome. In keeping with this trend, the offspring represented in lane 6 with the fewest repeats has reached adulthood without showing symptoms of the disorder.

Figure 7 shows PCR analysis for a second sibship from

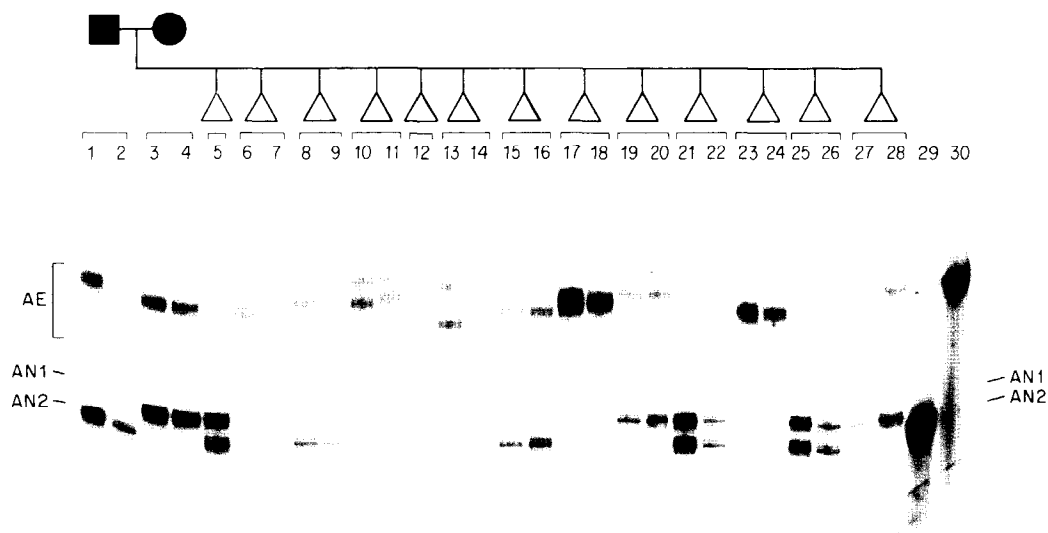


Figure 7. PCR Analysis of the (CAG)_n Repeat in a Venezuelan HD Sibship with Offspring Homozygous for the Same *HD* Haplotype

Results of PCR analysis of a sibship from the Venezuelan HD pedigree in which both parents are affected by HD are shown. Progeny are shown as triangles and birth order has been altered for confidentiality. No HD diagnostic information is given to preserve the blind status of investigators in the Venezuelan Collaborative Group. AN1 and AN2 mark the positions of the allelic products from normal parental chromosomes. AE marks the range of PCR products from the *HD* chromosome. The PCR products from cosmids L191F1 and GUS72-2130 are loaded in lanes 29 and 30 and have 18 and 48 CAG repeats, respectively.

the Venezuelan pedigree, in which both parents are *HD* heterozygotes carrying the same *HD* chromosome based on DNA marker studies. Several of the offspring are *HD* homozygotes (lanes 6 and 7, 10 and 11, 13 and 14, 17 and 18, 23 and 24) as reported previously (Wexler et al., 1987). Each parent's DNA contained 1 allele in the normal range (AN1 and AN2), which was transmitted in a Mendelian fashion. The *HD*-specific products (AE) from the DNA of both parents and children were all much larger than the normal allelic products and also showed extensive variation in mean size. We have not provided a neurologic diagnosis for the offspring in this pedigree to maintain the blind status of investigators involved in the ongoing Venezuela *HD* Project, although age of onset again appears to parallel repeat length. Paired samples under many of the individual symbols represent independent lymphoblast lines initiated at least 1 year apart. The variance between paired samples was not as great as between the different individuals, suggesting that the major differences in size of the PCR products resulted from meiotic transmission. Of special note is the result obtained in lanes 13 and 14. This *HD* homozygote's DNA yielded one PCR product larger and one smaller than the *HD*-specific PCR products of both parents.

To date, we have tested 75 independent *HD* families, representing all different haplotypes reported by MacDonald et al. (1992) and a wide range of ethnic backgrounds. In all 75 cases, a PCR product larger than the normal size range was produced from the *HD* chromosome. The sizes of the *HD*-specific products ranged from 42 repeat copies to more than 66 copies, with a few individuals failing to yield a product because of the extreme length of the repeat. In these cases, Southern blot analysis revealed an increase in the length of an *Eco*RI fragment, with the largest allele approximating 100 copies of the repeat. Figure 8 shows the variation detected in members of an American family of Irish ancestry in which the major *HD* haplotype is segregating. Cosmid GUS72-2130 was cloned from the *HD* homozygous individual whose DNA was amplified in lane 2. As was observed in the Venezuelan *HD* pedigree (Figures 6 and 7), which segregates the disorder with a different 4p16.3 haplotype, the *HD*-specific PCR products for this family display considerable size variation.

New Mutations to *HD*?

The mutation rate in *HD* has been reported to be very low. To test whether the expansion of the (CAG)_n repeat is the mechanism by which new *HD* mutations occur, we have examined two pedigrees with sporadic cases of *HD* in which intensive searching failed to reveal a family history of the disorder. In these cases, we gathered pedigree information sufficient to identify the same chromosomes in both the affected individual and unaffected relatives. Figure 9 shows the results of PCR analysis of the (CAG)_n repeat in these families. The chromosomes in each family were assigned an arbitrary number based on typing for a large number of restriction fragment length polymorphism and simple sequence repeat markers in 4p16.3 defining distinct haplotypes; the presumed *HD* chromosome is starred.

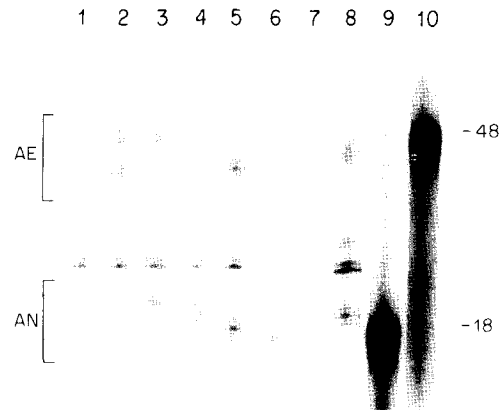


Figure 8. PCR Analysis of the (CAG)_n Repeat in Members of an American Family with an Individual Homozygous for the Major *HD* Haplotype. Results of PCR analysis of members of an American family segregating the major *HD* haplotype. AN marks the range of normal alleles; AE marks the range of *HD* alleles. Lanes 1, 3, 4, 5, 7, and 8 represent PCR products from related *HD* heterozygotes. Lane 2 contains the PCR products from a member of the family homozygous for the same *HD* chromosome. Lane 6 contains PCR products from a normal individual. Pedigree relationships and affected status are not presented to preserve confidentiality. The PCR products from cosmids L191F1 and GUS72-2130 (which was derived from the individual represented in lane 2) are loaded in lanes 9 and 10 and have 18 and 48 CAG repeats, respectively.

In family 1, *HD* first appeared in individual II-3, who transmitted the disorder, along with chromosome 3*, to III-1. This same chromosome was present in II-2, an elderly unaffected individual. PCR analysis revealed that chromosome 3* from II-2 produced a PCR product at the extreme high end of the normal range (~36 CAG copies). However, the (CAG)_n repeat on the same chromosome in II-3 and III-1 had undergone sequential expansions to ~44 and ~46 copies, respectively. A similar result was obtained in family 2, where the presumed new *HD* mutant III-2 had a considerably expanded repeat relative to the same chromosome in II-1 and III-1 (~49 versus ~33 CAG copies). In both families 1 and 2, the ultimate *HD* chromosome displays the marker haplotype characteristic of one-third of all *HD* chromosomes, suggesting that this haplotype may be predisposed to undergoing repeat expansion.

Discussion

The discovery of an expanded, unstable trinucleotide repeat on *HD* chromosomes suggests that the long-sought *HD* gene has at last been uncovered and that the disorder constitutes an example of a mutational mechanism that may prove quite common in human genetic disease. Elongation of a trinucleotide repeat sequence has been implicated previously as the cause of three quite different human disorders, the fragile X syndrome, myotonic dystrophy, and spino-bulbar muscular atrophy. Our initial observations of repeat expansion in *HD* indicate that this phenomenon shares features with each of these disorders.

In the fragile X syndrome, expression of a constellation of symptoms, including mental retardation and a fragile

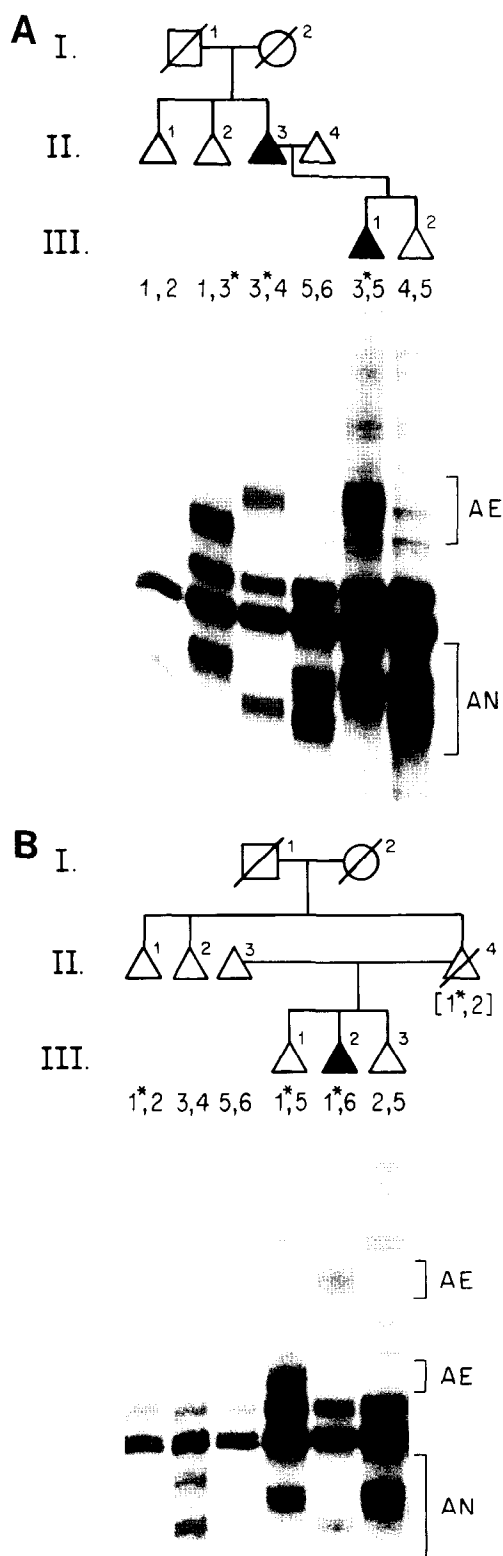


Figure 9. PCR Analysis of the (CAG)_n Repeat in Two Families with a Supposed New Mutation Causing HD

Results of PCR analysis of two families in which sporadic HD cases representing putative new mutants are shown. Individuals in each pedigree are numbered by generation (roman numerals) and order in the

site at Xq27.3, is associated with expansion of a (CGG)_n repeat thought to be in the 5' untranslated region of the *FMR1* gene (Fu et al., 1991; Kremer et al., 1991; Verkerk et al., 1991). In myotonic dystrophy, a dominant disorder involving muscle weakness with myotonia that typically presents in early adulthood, the unstable trinucleotide repeat, (CTG)_n, is located in the 3' untranslated region of the myotonin protein kinase gene (Aslanidis et al., 1992; Brook et al., 1992; Buxton et al., 1992; Fu et al., 1992; Harley et al., 1992a; Mahadevan et al., 1992). The unstable (CAG)_n repeat in *HD* may be within the coding sequence of the IT15 gene, a feature shared with spino-bulbar muscular atrophy, an X-linked recessive adult-onset disorder of the motor neurons caused by expansion of a (CAG)_n repeat in the coding sequence of the androgen receptor gene (LaSpada et al., 1991). The repeat length in both the fragile X syndrome and myotonic dystrophy tends to increase in successive generations, sometimes quite dramatically. Occasionally, decreases in the average repeat length are observed (Fu et al., 1991; Yu et al., 1992; Bruner et al., 1993). The *HD* trinucleotide repeat is also unstable, usually expanding when transmitted to the next generation, but contracting on occasion. In *HD*, as in the other disorders, change in copy number occurs in the absence of recombination. Compared with the fragile X syndrome, myotonic dystrophy, and *HD*, the instability of the disease allele in spino-bulbar muscular atrophy is more limited, and dramatic expansions of repeat length have not been seen (Biancalana et al., 1992).

Expansion of the repeat length in myotonic dystrophy is associated with a particular chromosomal haplotype, suggesting the existence of a primordial predisposing mutation (Harley et al., 1991, 1992a; Ashizawa, and Epstein, 1991). In the fragile X syndrome, there may be a limited number of ancestral mutations that predispose increases in trinucleotide repeat number (Richards et al., 1992; Oudet et al., 1993). The linkage disequilibrium analysis used to home in on IT15 indicates that there are several haplotypes associated with *HD*, but that at least one-third of *HD* chromosomes are ancestrally related (MacDonald et al., 1992). These data, combined with the reported low rate of new mutation to *HD* (Harper, 1992), suggest that expansion of the trinucleotide repeat may only occur on select chromosomes. Our analysis of two families in which new mutation was supposed to have occurred is consistent with the view that there may be particular normal chromosomes that have the capacity to undergo expansion of the repeat into the *HD* range. In each of these families, a chromosome with a (CAG)_n repeat length in the upper end of the normal range was segregating on a chromosome

pedigree. Triangles are used to protect confidentiality. Closed symbols indicate symptomatic individuals. The different chromosomes segregating in the pedigree have been distinguished by extensive typing with polymorphic markers in 4p16.3 and have been assigned arbitrary numbers shown above the gel lanes. The starred chromosomes (chromosome 3 in [A] and 1 in [B]) represent the presumed *HD* chromosome. AN denotes the range of normal alleles; AE denotes the range of alleles present in affected individuals and in their unaffected relatives bearing the same chromosome.

whose 4p16.3 haplotype matched the most common haplotype seen on *HD* chromosomes, and the clinical appearance of HD in these two cases was associated with expansion of the trinucleotide repeat.

The recent application of haplotype analysis to explore the linkage disequilibrium on *HD* chromosomes pointed to a portion of a 2.2 Mb candidate region defined by the majority of recombination events described in HD pedigrees (MacDonald et al., 1992). Previously, the search for the gene was confounded by three matings in which the genetic inheritance pattern was inconsistent with the remainder of the family (MacDonald et al., 1989b; Pritchard et al., 1992). These matings produced apparently affected HD individuals despite the inheritance of only normal alleles for markers throughout 4p16.3, effectively excluding inheritance of the *HD* chromosome present in the rest of the pedigree. Using our PCR assay, we have tested each of these families and find that, like other HD kindreds, an expanded allele generally segregated with *HD* in affected individuals of all three pedigrees. However, an expanded allele was not present in those specific individuals with the inconsistent 4p16.3 genotypes. Instead, these individuals displayed the normal alleles expected, based on analysis of other markers in 4p16.3. It is conceivable that these inconsistent individuals do not, in fact, have HD, but some other disorder. Alternatively, they might represent genetic mosaics in which the *HD* allele is more heavily represented and/or more expanded in brain tissue than in the lymphoblast DNA used for genotyping.

It can be expected that the capacity to monitor directly the size of the trinucleotide repeat in individuals "at risk" for HD will revolutionize preclinical testing for the disorder, eliminating the need for complicated linkage analyses, facilitating genetic counseling, and extending the applicability of presymptomatic and prenatal diagnosis to at risk individuals with no living affected relatives. We consider it of the utmost importance that the current internationally accepted guidelines and counseling protocols for testing those at risk continue to be observed, and that samples from unaffected relatives should not be tested inadvertently or without full consent. In our limited initial series of patients, there is an apparent correlation between repeat length and age of onset of the disease, reminiscent of that reported in myotonic dystrophy (Harley et al., 1992b; Tsilfidis et al., 1992). The largest *HD* trinucleotide repeat segments were found in juvenile onset cases, where there is a known preponderance of male transmission (Merritt et al., 1969). More detailed studies will be required to establish whether expansion of the repeat occurs preferentially in transmission from males. It will also be essential to perform a careful analysis of the extent, if any, of overlap between the range of repeat lengths in normal and HD individuals, to evaluate fully the relationship between age of onset and repeat length, and to examine the possibility of somatic variation in repeat length due to mitotic instability. These studies must be completed before the $(CAG)_n$ size is used to provide prognostic information to at risk HD individuals.

The expression of fragile X syndrome is associated with direct inactivation of the *FMR1* gene (Pieretti et al., 1991;

DeBouille et al., 1993). The recessive inheritance pattern of spino-bulbar muscular atrophy suggests that in this disorder an inactive gene product is produced. In myotonic dystrophy, the manner in which repeat expansion leads to the dominant disease phenotype is unknown. There are numerous possibilities for the mechanism of pathogenesis of the expanded trinucleotide repeat in *HD*. Since Wolf-Hirschhorn patients hemizygous for 4p16.3 do not display features of HD and IT15 mRNA is present in *HD* homozygotes, the expanded trinucleotide repeat does not cause simple inactivation of the gene containing it. The observation that the phenotype of *HD* is completely dominant, since homozygotes for the disease allele do not differ clinically from heterozygotes, has suggested that HD results from a gain-of-function mutation, in which either the mRNA product or the protein product of the disease allele would have some new property or would be expressed inappropriately (Wexler et al., 1987; Myers et al., 1989). If the expanded trinucleotide repeat were translated, the consequences on the protein product would be dramatic, increasing the length of the poly-glutamine stretch near the N-terminus. It is possible, however, that despite the presence of an upstream Met codon, the normal translational start occurs 3' to the $(CAG)_n$ repeat and there is no poly-glutamine stretch in the protein product. In this case, the repeat would be in the 5' untranslated region and might be expected to have its dominant effect at the mRNA level. The presence of an expanded repeat might directly alter regulation, localization, stability, or translatability of the mRNA containing it, and could indirectly affect its counterpart from the normal allele in *HD* heterozygotes. Other conceivable scenarios are that the presence of an expanded repeat might alter the effective translation start site for the *HD* transcript, thereby truncating the protein, or alter the transcription start site for the IT15 gene, disrupting control of mRNA expression. Finally, although the repeat is located within the IT15 transcript, the possibility that it leads to HD by virtue of an action on the expression of an adjacent gene cannot be excluded.

Despite this final caveat, we believe it most likely that the trinucleotide repeat expansion causes HD by its effect, either at the mRNA or protein level, on the expression and/or structure of the protein product of the IT15 gene, which we have named huntingtin. Outside of the region of the triplet repeat, the IT15 DNA sequence detected no significant similarity to any previously reported gene in the GenBank data base. Except for the stretches of glutamine and proline near the N-terminus, the amino acid sequence displayed no similarity to known proteins, providing no conspicuous clues to huntingtin's function. The poly-glutamine and poly-proline regions near the N-terminus indicate similarity to a large number of proteins that also contain long stretches of these amino acids. It is difficult to assess the significance of such similarities, although it is notable that many of these similarities are to DNA-binding proteins and that huntingtin does have a single leucine zipper motif at residue 1443. Huntingtin appears to be widely expressed, yet cell death in HD is confined to specific neurons in particular regions of the brain. Thus, with the mystery of the genetic basis of HD apparently solved,

defining the normal function of the huntingtin protein and delineating the mechanism whereby increased trinucleotide repeat length leads to the characteristic neuropathology of HD represent the next challenges in the effort to understand and to treat this devastating disorder.

Experimental Procedures

HD Cell Lines

Lymphoblast cell lines from HD families of varied ethnic backgrounds used for genetic linkage and disequilibrium studies (Conneally et al., 1989; MacDonald et al., 1992) have been established (Anderson and Gusella, 1984) in the Molecular Neurogenetics Unit, Massachusetts General Hospital, over the past 13 years. The Venezuelan HD pedigree is an extended kindred of over 12,000 members, in which all affected individuals have inherited the *HD* gene from a common founder (Gusella et al., 1983, 1984; Wexler et al., 1987).

DNA and RNA Blotting

DNA was prepared from cultured cells, and DNA blots were prepared and hybridized as described (Gusella et al., 1979, 1983). RNA was prepared and Northern blotting was performed as described by Taylor et al. (1992).

Construction of Cosmid Contig

The initial construction of the cosmid contig was by chromosome walking from cosmids L19 and BJ56 (Allitto et al., 1991; Lin et al., 1991). Two libraries were employed, a collection of Alu-positive cosmids from the reduced cell hybrid H39-8C10 (Whaley et al., 1991) and an arrayed flow-sorted chromosome 4 cosmid library (NM87545) provided by the Los Alamos National Laboratory. Walking was accomplished by hybridization of whole cosmid DNA, using suppression of repetitive and vector sequences, to robot-generated high density filter grids (Nizetic et al., 1991; Lehrach et al., 1990). Cosmids L1C2, L69F7, L228B6, and L83D3 were first identified by hybridization of yeast artificial chromosome clone YGA2 to the same arrayed library (Bates et al., 1992; Baxendale et al., 1991). *HD* cosmid GUS72-2130 was isolated by standard screening of a GUS72 cosmid library using a single-copy probe. Cosmid overlaps were confirmed by a combination of clone to clone and clone to genomic hybridizations, single-copy probe hybridizations, and restriction mapping.

cDNA Isolation and Characterization

Exon probes were isolated and cloned as described (Buckler et al., 1991). Exon probes and cDNAs were used to screen human λ ZAPII cDNA libraries constructed from adult frontal cortex, fetal brain, adenovirus-transformed retinal cell line RCA, and liver RNA. cDNA clones, PCR products, and trapped exons were sequenced as described (Sanger et al., 1977). Direct cosmid sequencing was performed as described (McClatchey et al., 1992). Data base searches were performed using the BLAST network service of the National Center for Biotechnology Information (Altschul et al., 1990).

PCR Assay of the (CAG)_n Repeat

Genomic primers flanking the (CAG)_n repeat are 5'-ATGAAGG-CCTTCGAGTCCCTCAAGTCTTC-3' and 5'-AAACTCACGGTCCGGT-GCAGCGGCTCCTCAG-3'. PCR amplification was performed in a reaction volume of 25 μ l using 50 ng of genomic DNA, 5 μ g of each primer, 10 mM Tris (pH 8.3), 5 mM KCl, 2 mM MgCl₂, 200 μ M (each) dNTPs, 10% dimethylsulfoxide, 0.1 U of Perfectmatch (Stratagene), 2.5 μ Ci of [³²P]dCTP (Amersham), and 1.25 U of Taq polymerase (Boehringer Mannheim). After heating to 94°C for 1.5 min, the reaction mix was cycled according to the following program: 40 cycles of 1 min at 94°C, 1 min at 60°C, 2 min at 72°C. Five microliters of each PCR was diluted with an equal volume of 95% formamide loading dye and heat denatured for 2 min at 95°C. The products were resolved on 5% denaturing polyacrylamide gels. The PCR product from this reaction using cosmid L191F1 (CAG₁₈) as the template was 247 bp. Allele sizes were estimated relative to a DNA sequencing ladder, the PCR products from sequenced cosmids, and the invariant background bands often present on the gel. Estimates of allelic variation were obtained by typing unrelated individuals of largely Western European ancestry,

who were normal parents of affected HD individuals from various pedigrees.

Acknowledgments

We thank the many investigators who have supplied blood samples from HD pedigrees, particularly the Venezuela Huntington's Disease Collaborative Group, and the many investigators who have over the years participated in or contributed to this project. We are also extremely grateful to the HD families themselves for supporting and participating in this research effort. We thank P. M. Conneally, G. Evans, M. Frangione, C. Gilliam, R. Horvitz, L. Moyzis, R. Mulligan, A. Novelletto, A. Tobin, and L. Zipursky for helpful discussions. This work was supported by National Institutes of Health grants NS16367 (Huntington's Disease Center Without Walls), NS22031, and NS25631 and by grants from Bristol-Myers Squibb, Inc., the Hereditary Disease Foundation Collaborative Research Agreement, the Joan and William A. Schreyer/Merrill Lynch Fund to Cure Huntington's Disease of the Hereditary Disease Foundation, the Huntington's Disease Society of America, the Foundation for the Care and Cure of Huntington's Disease, the W. M. Keck Foundation, the William J. Matheson Foundation, the Bay Foundation, The Charles A. Dana Foundation, the Medical Research Council (England), the Welsh Office, and the Wellcome Trust. Fellowship support was provided to C. M. A. by the Andrew B. Cogan Fellowship of the Hereditary Disease Foundation. Other postdoctoral fellowships to various investigators were generously provided by the Hereditary Disease Foundation, the Huntington's Disease Society of America, and the Huntington's Society of Canada.

Received February 26, 1993; revised March 4, 1993.

References

- Allitto, B. A., MacDonald, M. E., Bucan, M., Richards, J., Romano, D., Whaley, W. L., Falcone, B., Ianazzi, J., Wexler, N. S., Wasmuth, J. J., Collins, F. S., Lehrach, H., Haines, J. L., and Gusella, J. F. (1991). Increased recombination adjacent to the Huntington's disease-linked *D4S10* marker. *Genomics* 9, 104-112.
- Altherr, M. R., Plummer, S., Bates, G., MacDonald, M., Taylor, S., Lehrach, T. H., Frischauf, A. M., Gusella, J. F., Boehnke, M., and Wasmuth, J. J. (1992). Radiation hybrid map spanning the Huntington disease gene region of chromosome 4. *Genomics* 13, 1040-1046.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.
- Ambrose, C., James, M., Barnes, G., Lin, C., Bates, G., Altherr, M., Duyao, M., Groot, N., Church, D., Wasmuth, J. J., Lehrach, H., Housman, D., Buckler, A., Gusella, J. F., and MacDonald, M. E. (1992). A novel G protein-coupled receptor kinase cloned from 4p16.3. *Hum. Mol. Genet.* 1, 697-703.
- Anderson, M. A., and Gusella, J. F. (1984). Use of cyclosporin A in establishing Epstein-Barr virus-transformed human lymphoblastoid cell lines. *In Vitro* 20, 856-858.
- Ashizawa, T., and Epstein, H. F. (1991). Ethnic distribution of myotonic dystrophy gene. *Lancet* 338, 642-643.
- Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevan, M., Tsiflidis, C., Chen, C., Alleman, J., Wormskamp, N. G. M., Vooijs, M., Buxton, J., Johnson, K., Smeets, H. J. M., Lennon, G. G., Carrano, A. V., Korneluk, R. G., Wieringa, B., and de Jong, P. J. (1992). Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature* 355, 548-551.
- Bates, G. P., MacDonald, M. E., Baxendale, S., Sedlacek, Z., Youngman, S., Romano, D., Whaley, W. L., Allitto, B. A., Poustka, A., Gusella, J. F., and Lehrach, H. (1990). A YAC telomere clone spanning a possible location of the Huntington's disease gene. *Am. J. Hum. Genet.* 46, 762-775.
- Bates, G. P., MacDonald, M. E., Baxendale, S., Youngman, S., Lin, C., Whaley, W. L., Wasmuth, J. J., Gusella, J. F., and Lehrach, H. (1991). Defined physical limits of the Huntington disease gene candidate region. *Am. J. Hum. Genet.* 49, 7-16.
- Bates, G. P., Valdes, J., Hummerich, H., Baxendale, S., Le Paslier, D. L., Monaco, A. P., Tagle, D., MacDonald, M. E., Altherr, M., Ross,

- M., Brownstein, B. H., Bentley, D., Wasmuth, J. J., Gusella, J. F., Cohen, D., Collins, F., and Lehrach, H. (1992). Characterization of a yeast artificial chromosome contig spanning the Huntington disease gene candidate region. *Nature Genet.* *1*, 180-187.
- Baxendale, S., Bates, G. P., MacDonald, M. E., Gusella, J. F., and Lehrach, H. (1991). The direct screening of cosmid libraries with yeast artificial chromosomes. *Nucl. Acids Res.* *19*, 6651.
- Biancalana, V., Serville, F., Pommier, J., Julien, J., Hanauer, A., and Mandel, J. L. (1992). Moderate instability of the trinucleotide repeat in spino-bulbar muscular atrophy. *Hum. Mol. Genet.* *1*, 255-258.
- Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J.-P., Hudson, T., Sohn, R., Zemelman, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P. S., Shaw, D. J., and Housman, D. E. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* *68*, 799-808.
- Bruner, H. G., Jansen, G., Nillesen, W., Nelen, M. R., DeDie, C. E. M., Howeller, C. J., van Oost, B. A., Wieringa, B., Ropers, H. H., and Smeets, H. J. M. (1993). Reverse mutation in myotonic dystrophy. *N. Engl. J. Med.* *328*, 476-480.
- Bucan, M., Zimmer, M., Whaley, W. L., Poustka, A., Youngman, S., Allitto, B. A., Ormondroyd, E., Smith, B., Pohl, T. M., MacDonald, M., Bates, G., Richards, J., Volinia, S., Gilliam, T. C., Sedlacek, Z., Collins, F. S., Wasmuth, J. J., Shaw, D. J., Gusella, J. F., Frischauf, A. M., and Lehrach, H. (1990). Physical maps of 4p16.3, the area expected to contain the Huntington's disease mutation. *Genomics* *6*, 1-15.
- Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. (1991). Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* *88*, 4005-4009.
- Buxton, J., Shelbourne, P., Davies, J., Jones, C., Van Tongeren, T., Aslanidis, C., de Jong, P., Jansen, G., Anvret, M., Riley, B., Williamson, R., and Johnson, K. (1992). Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* *355*, 547-548.
- Conneally, P. M., Haines, J. L., Tanzi, R. E., Wexler, N. S., Penchaszadeh, G. K., Harper, P. S., Folstein, S. E., Cassiman, J. J., Myers, R. H., Young, A. B., Hayden, M. R., Falek, A., Tolosa, E. S., Crespi, S., Di Maio, L., Holmgren, G., Anvret, M., Kanazawa, I., and Gusella, J. F. (1989). Huntington disease: no evidence for locus heterogeneity. *Genomics* *5*, 304-308.
- DeBouille, K., Verkerk, A. J. M. H., Reyniers, E., Vits, L., Hendrickx, J., VanRoy, B., VanDenBos, F., deGraaff, E., Oostra, B. A., and Willems, P. J. (1993). A point mutation in the *FMR-1* gene associated with fragile X mental retardation. *Nature Genet.* *3*, 31-35.
- Doucette-Stamm, L. A., Riba, L., Handelin, B., Difilippantonio, M., Ward, D. C., Wasmuth, J. J., Gusella, J. F., and Housman, D. E. (1991). Generation and characterization of Goss-Harris hybrids of human chromosome 4. *Somat. Cell Mol. Genet.* *17*, 471-480.
- Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J. M. H., Holden, J. J. A., Fenwick, R. G., Jr., Warren, S. T., Oostra, B. A., Nelson, D. L., and Caskey, C. T. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* *67*, 1047-1058.
- Fu, Y.-H., Pizzuti, A., Fenwick, R. G., King, J., Jr., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., DeJong, P., Wieringa, B., Korneluk, R., Peryman, M. B., Epstein, H. F., and Caskey, C. T. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* *255*, 1256-1259.
- Gusella, J. F. (1989). Location cloning strategy for characterizing genetic defects in Huntington's disease and Alzheimer's disease. *FASEB J.* *3*, 2036-2041.
- Gusella, J. F. (1991). Huntington's disease. *Adv. Hum. Genet.* *20*, 125-151.
- Gusella, J. F., Varsanyi-Breiner, A., Kao, F. T., Jones, C., Puck, T. T., Keys, C., Orkin, S., and Housman, D. E. (1979). Precise localization of the human β -globin gene complex on chromosome 11. *Proc. Natl. Acad. Sci. USA* *76*, 5239-5243.
- Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y., Young, A. B., Shoulson, I., Bonilla, E., and Martin, J. B. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* *306*, 234-238.
- Gusella, J. F., Tanzi, R. E., Anderson, M. A., Hobbs, W., Gibbons, K., Raschtchian, R., Gilliam, T. C., Wallace, M. R., Wexler, N. S., and Conneally, P. M. (1984). DNA markers for nervous system diseases. *Science* *225*, 1320-1326.
- Harley, H. G., Brook, J. D., Floyd, J., Rundle, S. A., Crow, S., Walsh, K. V., Thibault, M. C., Harper, P. S., and Shaw, D. J. (1991). Detection of linkage disequilibrium between the myotonic dystrophy locus and a new polymorphic DNA marker. *Am. J. Hum. Genet.* *49*, 68-75.
- Harley, H. G., Brook, J. D., Rundle, S. A., Crow, S., Reardon, W., Buckler, A. J., Harper, P. S., and Housman, D. E. (1992a). Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* *355*, 545-546.
- Harley, H. G., Rundle, S. A., Reardon, W., Myring, J., Crow, S., Brook, J. D., Harper, P. S., and Shaw, D. J. (1992b). Unstable DNA sequence in myotonic dystrophy. *Lancet* *339*, 1125-1128.
- Harper, P. S. (1992). The epidemiology of Huntington's disease. *J. Med. Genet.* *89*, 365-376.
- Harper, P. S., Morris, M. J., Quarrell, O., Shaw, D. J., Tyler, A., and Youngman, S. (1991). Huntington's Disease (Philadelphia: W. B. Saunders).
- Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S. T., Schlessinger, D., Sutherland, G. R., and Richards, R. I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)_n. *Science* *252*, 1711-1714.
- LaSpada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., and Fishbeck, H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* *352*, 77-79.
- Lehrach, H., Drmanac, R., Hoheisel, J., Larin, Z., Lennon, G., Nizetic, D., Monaco, A., Zehetner, G., and Poustka, A. (1990). Hybridisation fingerprinting in genome mapping and sequencing. In *Genome Analysis: Genetic and Physical Mapping*, Volume 1, K. E. Davies and S. M. Tighman, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 39-81.
- Lin, C. S., Altherr, M., Bates, G., Whaley, W. L., Read, A. P., Harris, R., Lehrach, H., Wasmuth, J. J., Gusella, J. F., and MacDonald, M. E. (1991). New DNA markers in the Huntington's disease gene candidate region. *Somat. Cell Mol. Genet.* *17*, 481-488.
- MacDonald, M. E., Cheng, S. V., Zimmer, M., Haines, J. L., Poustka, A. M., Allitto, B. A., Smith, B., Whaley, W. L., Romano, D., Jagadeesh, J., Lehrach, H., Wasmuth, J. J., Frischauf, A. M., and Gusella, J. F. (1989a). Clustering of multi-allele DNA markers near the Huntington's disease gene. *J. Clin. Invest.* *84*, 1013-1016.
- MacDonald, M. E., Haines, J. L., Zimmer, M., Cheng, S. V., Youngman, S., Whaley, W. L., Bucan, W. L., Allitto, B. A., Smith, B., Leavitt, J., Poustka, A. M., Harper, P., Lehrach, H., Wasmuth, J. J., Frischauf, A. M., and Gusella, J. F. (1989b). Recombination events suggest possible locations for the Huntington's disease gene. *Neuron* *3*, 183-190.
- MacDonald, M. E., Lin, C., Srinidhi, L., Bates, G., Altherr, M., Whaley, W. L., Lehrach, H., Wasmuth, J., and Gusella, J. F. (1991). Complex patterns of linkage disequilibrium in the Huntington disease region. *Am. J. Hum. Genet.* *49*, 723-734.
- MacDonald, M. E., Novelletto, A., Lin, C., Tagle, D., Barnes, G., Bates, G., Taylor, S., Allitto, B., Altherr, M., Myers, R., Lehrach, H., Collins, F. S., Wasmuth, J. J., Frontali, M., and Gusella, J. F. (1992). The Huntington's disease candidate region exhibits many different haplotypes. *Nature Genet.* *1*, 99-103.
- Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., Leblond, S., Earle-MacDonald, J., DeJong, P. J., Wieringa, B., and Korneluk, G. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* *255*, 1253-1255.
- Martin, J. B., and Gusella, J. F. (1986). Huntington's disease: pathogenesis and management. *N. Engl. J. Med.* *315*, 1267-1276.
- McClatchey, A. I., Lin, C. S., Wang, J., Hoffman, E. P., Rojas, C., and

- Gusella, J. F. (1992). The genomic structure of the human skeletal muscle sodium channel gene. *Hum. Mol. Genet.* 1, 521–527.
- Merritt, A. D., Conneally, P. M., Rahman, N. F., and Drew, A. L. (1969). Juvenile Huntington's chorea. In *Progress in Neurogenetics*, A. Barbeau and J. R. Brunette, eds. (Amsterdam: Excerpta Medica), pp. 645–650.
- Myers, R. H., Leavitt, J., Farrer, L. A., Jagadeesh, J., McFarlane, H., Mark, R. J., and Gusella, J. F. (1989). Homozygote for Huntington's disease. *Am. J. Hum. Genet.* 45, 615–618.
- Nizetic, D., Zehetner, G., Monaco, A., Gellen, L., Young, B. D., and Lehrach, H. (1991). Construction, arraying, and high density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. *Proc. Natl. Acad. Sci. USA* 88, 3233–3237.
- Oudet, C., Mornet, E., Serre, J. L., Thomas, F., Lentes-Zengerling, S., Kretz, C., Deluchat, C., Tejada, I., Boue, J., Boue, A., and Mandel, J. L. (1993). Linkage disequilibrium between the fragile X mutation and two closely linked CA repeats suggests that fragile X chromosomes are derived from a small number of founder chromosomes. *Am. J. Hum. Genet.* 52, 297–304.
- Pieretti, M., Zhang, F., Fu, Y.-H., Warren, S. T., Oostra, B. A., Caskey, C. T., and Nelson, D. L. (1991). Absence of expression of the *FMR-1* gene in fragile X syndrome. *Cell* 66, 817–822.
- Pohl, T. M., Zimmer, M., MacDonald, M. E., Smith, B., Bucan, M., Poustka, A., Volinia, S., Searle, S., Zehetner, G., Wasmuth, J. J., Gusella, J., Lehrach, H., and Frischauf, A. M. (1988). Construction of a NotI linking library and isolation of new markers close to the Huntington's disease gene. *Nucl. Acids Res.* 16, 9185–9198.
- Pritchard, C., Zhu, N., Zuo, J., Bull, L., Pericak-Vance, M. A., Roses, A. D., Milatovich, A., Francke, U., Cox, D. R., and Myers, R. M. (1992). Recombination of 4p16 DNA markers in an unusual family with Huntington disease. *Am. J. Hum. Genet.* 50, 1218–1230.
- Richards, R. I., Holman, K., Friend, K., Kremer, E., Hillen, D., Staples, A., Brown, W. T., Goonewardena, P., Tarleton, J., Schwartz, C., and Sutherland, G. R. (1992). Fragile X syndrome: evidence of founder chromosomes. *Nature Genet.* 1, 257–260.
- Sanger, T., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- Snell, R. G., Lazarou, L., Youngman, S., Quarrell, O. W. J., Wasmuth, J. J., Shaw, D. J., and Harper, P. S. (1989). Linkage disequilibrium in Huntington's disease: an improved localization for the gene. *J. Med. Genet.* 26, 673–675.
- Snell, R. G., Thompson, L. M., Tagle, D. A., Holloway, T. L., Barnes, G., Harley, H. G., Sandkuijl, L. A., MacDonald, M. E., Collins, F. S., Gusella, J. F., Harper, P. S., and Shaw, D. J. (1992). A recombination event that redefines the Huntington disease region. *Am. J. Hum. Genet.* 51, 357–362.
- Suthers, G. K., Huson, S. M., and Davies, K. E. (1992). Instability versus predictability: the molecular diagnosis of myotonic dystrophy. *J. Med. Genet.* 29, 761–765.
- Taylor, S. A. M., Snell, R. G., Buckler, A., Ambrose, C., Duyao, M., Church, D., Lin, C. S., Altherr, M., Bates, G. P., Groot, N., Barnes, G., Shaw, D. J., Lehrach, H., Wasmuth, J. J., Harper, P. S., Housman, D. E., MacDonald, M. E., and Gusella, J. F. (1992). Cloning of the α -adducin gene from the Huntington's disease candidate region of chromosome 4 by exon amplification. *Nature Genet.* 2, 223–227.
- Theilman, J., Kanani, S., Shiang, R., Robbins, C., Quarrell, O., Huggins, M., Hedrick, A., and Hayden, M. R. (1989). Non-random association between alleles detected at *D4S95* and *D4S98* and the Huntington's disease gene. *J. Med. Genet.* 26, 676–681.
- Thompson, L. M., Plummer, S., Schalling, M., Altherr, M. R., Gusella, J. F., Housman, D. E., and Wasmuth, J. J. (1991). A gene encoding a fibroblast growth factor receptor isolated from the Huntington disease gene region of human chromosome 4. *Genomics* 11, 1133–1142.
- Tsilfidis, C., McKenzie, A. E., Mettler, G., Barcelo, J., and Korneluk, R. G. (1992). Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. *Nature Genet.* 1, 192–195.
- Verkerk, A. J. M. H., Pieretti, M., Sutcliffe, J. S., Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, R., Eussen, B. E., van Ommen, G.-J. B., Blonden, L. A. J., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A., and Warren, S. T. (1991). Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65, 905–914.
- Wexler, N. S., Young, A. B., Tanzi, R. E., Travers, H., Starosta-Rubenstein, S., Penney, J. B., Snodgrass, S. R., Shoulson, I., Gomez, F., Ramos-Arroyo, M. A., Penchaszadeh, G., Moreno, R., Gibbons, K., Faryniarz, A., Hobbs, W., Anderson, M. A., Bonilla, E., Conneally, P. M., and Gusella, J. F. (1987). Homozygotes for Huntington's disease. *Nature* 326, 194–197.
- Whaley, W. L., Bates, G. P., Novelleto, A., Sedlacek, Z., Cheng, S., Romano, D., Ormondroyd, E., Allitto, B. A., Lin, C., Youngman, S., Baxendale, S., Bucan, M., Altherr, M., Wasmuth, J., Wexler, N. S., Frontali, M., Frischauf, A. M., Lehrach, H., MacDonald, M. E., and Gusella, J. F. (1991). Mapping of cosmid clones in the Huntington's disease region of chromosome 4. *Somat. Cell Mol. Genet.* 17, 83–91.
- Youngman, S., Bates, G. P., Williams, S., McClatchey, A. I., Baxendale, S., Sedlacek, Z., Altherr, M., Wasmuth, J. J., MacDonald, M. E., Gusella, J. F., and Lehrach, H. (1992). The telomeric 60 kb of chromosome arm 4p is homologous to telomeric regions on 13p, 15p, 21p, and 22p. *Genomics* 14, 350–356.
- Yu, S., Mulley, J., Loesch, D., Turner, G., Donnelly, A., Gedeon, A., Hillen, D., Kremer, E., Lynch, M., Pritchard, M., Sutherland, G. R., and Richards, R. I. (1992). Fragile-X syndrome: unique genetics of the heritable unstable element. *Am. J. Hum. Genet.* 50, 968–980.

GenBank Accession Number

The accession number for the sequence reported in this paper is L12392.