

Mapping and Characterization of Novel (CAG)_n Repeat cDNAs from Adult Human Brain Derived by the Oligo Capture Method

P. H. Reddy,* Elaine Stockburger,* Patrick Gillevet,† and Danilo A. Tagle*¹

*Genetics and Molecular Biology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892-4442; and †Institute of Bioscience, Bioinformatics and Biotechnology, George Mason University, Manassas, Virginia 22030

Received July 31, 1997; accepted September 17, 1997

The expansion of a (CAG)_n trinucleotide repeat has been associated with at least eight neurological disorders in which the repeats code for polyglutamine in the protein. To identify additional genes that possess (CAG)_n repeats, single-stranded cDNA clones derived from adult human brain were screened using biotinylated oligonucleotide (CAG)₈, and the hybridizing complexes were isolated with streptavidin-coated paramagnetic beads. A total of 119 cDNA clones were isolated and initially characterized by end sequencing. BLAST homology searches were used to reduce redundancies with overlapping clones and to eliminate those that show sequence identity with previously published cDNAs with triplet repeats. Only cDNA clones with more than five CAG repeats were pursued for analysis. A total of 19 novel cDNAs were further characterized by determining chromosomal assignments using the Stanford G3 and Genebridge radiation-reduced hybrid mapping panels. Transcript sizes and tissue expression patterns were determined by Northern blot analysis. Two of 19 clones showed specific or high expression in brain. These cDNAs are ideal candidate genes for other neurodegenerative disorders, such as spinocerebellar ataxia types 5 and 7, and may also be implicated in psychiatric diseases such as bipolar affected disorder and schizophrenia. © 1997 Academic Press

Key Words: trinucleotide repeats; CAG expansion; EST; neurodegeneration; brain expression.

INTRODUCTION

Dynamic mutations in trinucleotide repeat tracts have been implicated in at least 13 inherited diseases in humans. These dynamic mutations can be effectively categorized into expansions within noncoding and coding sequences. The former includes (CGG)_n expansions resulting in chromosome fragile sites (FRAXA, FRAXE,

and FRA11B) associated with mental retardation (Jones *et al.*, 1994; Knight *et al.*, 1993; Verkerk *et al.*, 1991); (CTG)_n expansions found in myotonic dystrophy (Brook *et al.*, 1992; Fu *et al.*, 1992; Mahadevan *et al.*, 1992); and (GAA)_n expansions seen in Friedreich ataxia (Campuzano *et al.*, 1996). Two additional fragile sites (FRAXF and FRA16A) are associated with (CGG)_n triplet repeat expansions but have not been implicated with any human diseases (Nancarrow *et al.*, 1994; Parish *et al.*, 1994). Dynamic mutations in coding regions involve (CAG)_n expansions observed in patients with spinal bulbar muscular atrophy (SBMA; La Spada *et al.*, 1991), Huntington disease (HD; The Huntington's Disease Collaborative Research Group, 1993), spinocerebellar ataxia 1 (SCA1; Orr *et al.*, 1993), dentatorubropallidolusian atrophy (DRPLA; Koide *et al.*, 1994; Nagafuchi *et al.*, 1994; which is allelic to Haw River syndrome; Burke *et al.*, 1994; Potter *et al.*, 1996), Machado–Joseph disease (which is allelic to SCA3; Kawaguchi *et al.*, 1994), and, more recently, SCA2 (Imbert *et al.*, 1996) and SCA6 (Zhuchenko *et al.*, 1997).

Most of the diseases caused by (CAG)_n repeat expansions share certain clinical features such as neuronal degeneration, dominant inheritance patterns, and genetic anticipation, resulting in earlier age of onset in successive generations largely in part due to the tendency of the affected parent's allele to expand in the germ line, particularly when transmitted from the paternal line (Albin and Tagle, 1995; Hummerich and Lehrach, 1995; Warren, 1996; Zoghbi, 1996). A number of strategies have been employed in the past to identify genes or cDNAs with triple repeats that may be associated with disorders of the brain. These include screening of brain cDNA libraries (Jiang *et al.*, 1995; Li *et al.*, 1993; Riggins *et al.*, 1992; Bulle *et al.*, 1997; Margolis *et al.*, 1997), screening of genomic DNA from normal (Gastier *et al.*, 1996) or patient DNA (Sanpei *et al.*, 1996; Schalling *et al.*, 1993), detection of polyglutamine-expanded proteins from patient samples (Imbert *et al.*, 1996; Stevanin *et al.*, 1996; Trottier *et al.*, 1995), searches in the EST database (Néri *et al.*, 1996; Bulle *et al.*, 1997), and FISH (Haaf *et al.*, 1996). These ap-

¹ To whom correspondence should be addressed at the Genetics and Molecular Biology Branch, NHGRI, NIH, Building 49, Room 3A14, 49 Convent Drive MSC 4442, Bethesda, MD 20892-4442. Telephone: (301) 402-2530. Fax: (301) 402-4929.

proaches have led, in some cases, toward the identification of the genes for SBMA, DRPLA, SCA2, and SCA6 (Imbert *et al.*, 1996; Koide *et al.*, 1994; La Spada *et al.*, 1991; Nagafuchi *et al.*, 1994; Sanpei *et al.*, 1996; Zhuchenko *et al.*, 1997). In other cases (CAG)_n repeat expansions have been indirectly implicated after analysis of genomic DNA from patients with schizophrenia and bipolar disorder (Morris *et al.*, 1995; O'Donovan *et al.*, 1995) and for glutamine expansions in the protein lysates from patients with autosomal dominant cerebellar ataxias (Stevanin *et al.*, 1996).

In this study, we have identified and characterized additional (CAG)_n clones that were isolated from adult human brain by (1) hybridization of single-stranded cDNA to biotinylated (CAG)₈ oligonucleotide and capture of the annealed complexes using streptavidin-coated paramagnetic beads, (2) sequencing of the isolated clones, (3) BLAST searches in the nucleotide and protein databases to identify clone homologies and redundancies, (4) determination of transcript size and tissue-specific expression patterns by Northern blot analysis, (5) generation of gene-based STSs that flank the (CAG)_n repeats and allow amplification directly from genomic DNA, (6) chromosomal localization of the clones, using the Stanford G8 and Genebridge radiation-reduced hybrid panels, (7) determining the degree of variability of the (CAG)_n repeat in these clones, using at least 40 parents from the CEPH reference pedigree, and (8) correlation of chromosomal assignment of the clones with known loci for a number of autosomal dominant neurodegenerative diseases.

MATERIALS AND METHODS

Screening for (CAG)_n repeats. A phagemid cDNA library from adult human brain (Life Technologies, Inc., Gaithersburg, MD) was used as a source of DNA in this study. The adult human brain library was screened using the GeneTrapper cDNA positive selection system. Oligonucleotide (CAG)₈ (Midland Certified Reagent Co., Midland, TX) was 5' biotinylated with biotin-14-dUTP. Approximately 5 μg of double-stranded phagemid cDNA inserts was converted to single-stranded DNA using Gene II (Life Technologies, Inc.) and *Escherichia coli* Exonuclease III enzymes. The biotinylated oligonucleotide was hybridized to the single-stranded cDNA at 37°C for 1 h, and the hybrids were captured with streptavidin-coated paramagnetic beads (Life Technologies, Inc.) and precipitated using a magnetic particle concentrator (DynaL, Inc., Lake Success, NY). The captured cDNAs were eluted and converted to double-stranded DNA using T7 DNA polymerase and nonbiotinylated (CAG)₈ oligonucleotide. The DNA was electroporated into DH10B cells (Life Technologies, Inc.) and plated on LB-ampicillin (100 mg/liter) plates.

Sequencing of (CAG)_n-enriched clones. Plasmid DNA was prepared from randomly picked cDNA clones and sequenced using the ABI Dye Deoxy Terminator cycle sequencing chemistry (Applied Biosystems, Inc., Foster City, CA). Vector primers SP6 and T7 were initially used for sequencing. Custom primers for each clone were made from the initial sequences until a CAG repeat tract was encountered. Reactions consisted of 1 μg of plasmid DNA and 5 to 7 pmol of primer. Reactions were purified on Centrisep columns and electrophoresed on an ABI Model 373A semiautomated DNA sequencer (Applied Biosystems, Inc.). Sequences were assembled and edited using Sequencher Version 3.0 (Gene Codes, Inc., Ann Arbor, MI).

BLAST search analysis and data management. The sequences were submitted for BLASTn and BLASTx searches (Altschul *et al.*,

1990) using "Multiblast" shellsript in GDE (Smith *et al.*, 1994). The output from the BLAST searches was sorted by probability using another shellsript ("Sortblast"), which condensed the BLAST report to the description line and alignments. A Microsoft Excel table of raw data of the clones showing the number of repeats, type of repeat, sequencing status, mapping status, and BLAST identities was maintained.

Mapping of (CAG)_n clones. PCR primers immediately flanking the (CAG)_n repeats were designed for each clone. Each primer set was tested on a 9600 PE PCR machine (Perkin-Elmer Corp., Norwalk, CT) for specific amplification of the cDNA clone and human genomic DNA but not on mouse and hamster genomic DNAs. PCR conditions consisted of initial denaturation at 94°C for 4 min, followed by 35 cycles consisting of 94°C for 30 s, annealing at 55–65°C (see Table 2) for 30 s, and 72°C extension for either 1 or 2 min, followed by a 10-min 72°C final extension period. Polymerase chain reactions were performed in 10-μl volume consisting of 10 ng genomic DNA, 20 pM each primer, and a 1× concentration of PCR SuperMIX (Life Technologies, Inc.). Chromosomal mapping was performed by PCR for (CAG)_n clones with five or more tandem repeats using GeneBridge 4 and Stanford G3 radiation hybrid panels. The Coriell NIGMS Cell Repositories (Camden, NJ) or BIOS Laboratories, Inc. (New Haven, CT) monochromosomal hybrid panels were used for additional confirmation of the chromosomal assignments.

Clone DT1P1B10 gave 141- and 118-bp PCR products in genomic DNA but only the higher molecular weight product when cDNA was used as target. The 118-bp product was subcloned into plasmid using the TA cloning system (Invitrogen Corp., San Diego, CA) and sequenced as described above.

Northern blot analysis. Northern blot analysis was done for (CAG)_n clones to determine the mRNA transcript size, levels of expression, and tissue-specific expression patterns. mRNA blots were obtained from Clontech Laboratories, Inc. (Palo Alto, CA). About 100 ng of the cDNA insert was radiolabeled by random-priming using [α -³²P]dCTP. Membrane blots were prehybridized with 50% deionized formamide buffer for 3 h at 42°C and hybridized overnight with radiolabeled probe. Membranes were washed at 55°C with wash solutions, and blots were exposed to X-ray film for 30 min to overnight at -80°C.

Test for polymorphism. Primer sets optimized for each clone were tested on genomic DNA from at least 40 parents of the CEPH reference pedigrees (families 12, 21, 23, 28, 134, 142, 1332, 1333, 1334, 1341, 1344, 1345, 1346, 1349, 1362, 1375, 1377, 1408, 1413, 1416, 1418, 1420, 1421, and 1423). PCR amplifications were performed as described above except [α -³²P]dCTP was incorporated in the reaction (1.5 μM dCTP and 125 μM each dATP, dTTP, and dGTP), and the PCR samples were resolved on 6% denaturing Sequagel 6 acrylamide gel (National Diagnostics, Atlanta, GA). A sequencing ladder generated by sequencing control plasmid using -40 forward primer from the Sequenase kit (United States Biochemical, Cleveland, OH) was used as size marker to estimate the allele sizes.

RESULTS

Captured (CAG)_n Clones

Figure 1 shows the strategy used in oligo capture of (CAG)_n cDNA clones derived from adult human brain. A number of clones were obtained using this strategy, of which 119 clones were randomly sampled for sequencing. The clones were examined for repeat size and for similarities to each other and in the sequence database. In myotonic dystrophy, a repeat size of five was the most common normal allele and yet can expand to a pathogenic range (Imbert *et al.*, 1993). Thus only cDNA clones that had five tandem repeats or more and those that did not show homologies to previously pub-

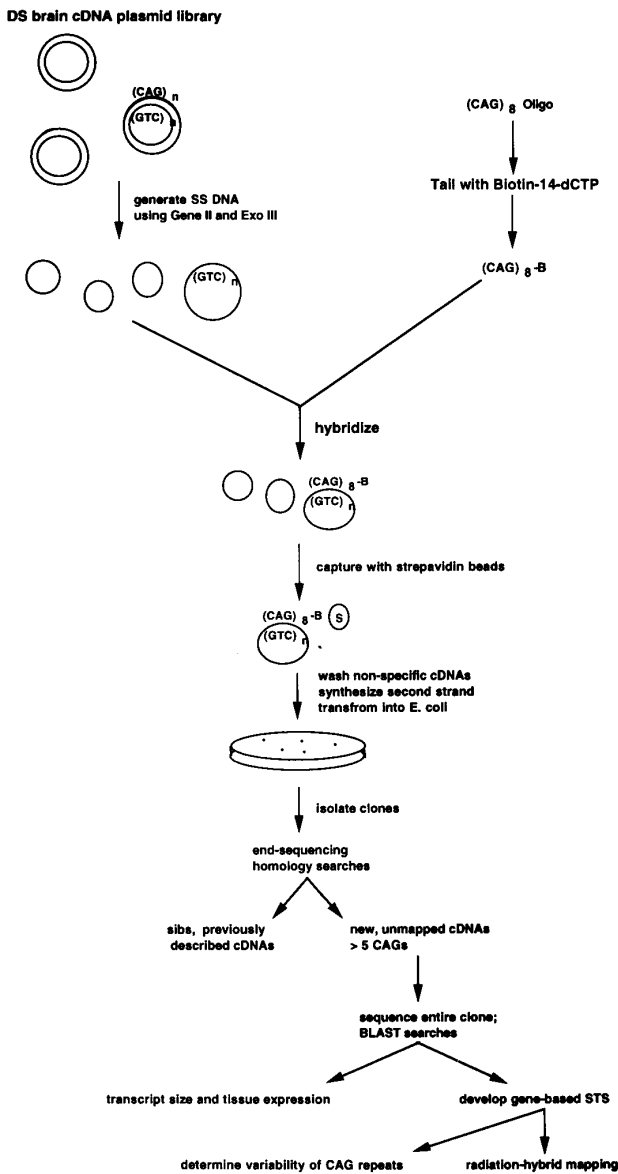


FIG. 1. Overview of the strategy employed to capture and characterize the CAG repeat-containing cDNAs from an adult human brain library.

lished triplet repeat genes or represented sib clones were analyzed further. These included cDNA clones that showed homology to the gene for SCA3. Of the 119 clones, 19 had five or more CAG repeats and represented unique, nonredundant clones (Table 1).

BLAST homology searches on the 19 clones are shown in Table 1. Ten of 19 clones showed greater than 90% sequence identity to known proteins or genes in the databases. Three of 19 clones showed significant matches to known ESTs that have known $(CAG)_n$ repeats. However, in these cases the chromosomal locations have not been determined for these ESTs, and they were thus included in this study.

Mapping of $(CAG)_n$ Clones

The sequences were used to turn the cDNA clones into gene-based STSs by designing PCR primers imme-

diately flanking the repeat sequences for the 19 cDNA clones. Primers were positioned on the cDNA sequence to give a short amplification product (approximately 100 bases on the average) to avoid amplifying across any introns when genomic DNA is used as a target. In two clones (DT1P1B12 and DT2P1F10) mapping primers were made 3' of the CAG repeat tract because of failure to amplify across the repeat sequence. This is likely due to the repeat tract lying very close to an exon-intron junction, since the expected PCR product size was obtained from their respective cDNA clones but no amplification was obtained from total human genomic DNA. Using the gene-based STS primers on the Genebridge 4 and/or Stanford G3 radiation-reduced hybrid panels, linkage was obtained for 18/19 clones to previously assigned chromosomal markers (Table 3). Mapping information was not obtained for DT2P1A11 since all the possible primers tested for this clone gave PCR signals in both human and rodent samples, which is likely due to the high sequence conservation of this gene in these species.

The radiation hybrid mapping system showed mapping assignment to at least two chromosomes for three clones: DT1P2C12 was mapped to chromosomes 5 and 6, DT2P1F10 was mapped to chromosomes 6 and 9, and DT1P1B10 was mapped to chromosomes 3 and 12. Clone DT1P1B10 consistently gave the expected 141-bp PCR product plus an additional 118-bp product. When only the 141-bp product was scored by radiation hybrid mapping, a chromosomal assignment to human chromosome 3 was obtained, whereas the lower band gave a map location to human chromosome 12. The faster migrating PCR product was subcloned, and sequence analysis confirmed the presence of a related gene on chromosome 12 with 8 CAG repeats compared to the locus on chromosome 3, which has 15 repeats. The two sequences showed 98% similarity and most likely represent members of the same gene family. To resolve the exact chromosomal location for clones DT1P2C12 and DT2P1F10, we hybridized the cDNAs to a somatic cell hybrid panel (Oncor, Inc., Gaithersburg, MD). However, Southern blot analysis showed strong cross-hybridizing signals from the rodent background for clones DT1P2C12 and DT2P1F10 (data not shown); thus, precise chromosomal assignment was not obtained for these clones.

Expression Patterns of $(CAG)_n$ Clones

The mRNA sizes for each of the 19 clones were determined by Northern blot hybridization, and the results are shown in Table 4. Transcript sizes ranged from 0.7 to 12.0 kb with an average size of 3.7 kb. In three of these genes, the transcript is represented in its entirety in the plasmid clones DT1P1A10, DT1P2C4, and DT2P1G7 (Table 4). The majority of the clones showed widespread expression in all the tissues tested and the hybridization patterns of DT1P1A2, DT1P1A4, DT1P1B12, and DT1P2C6 are shown as examples in

TABLE 1

Blast Search Results for the cDNA Clones with Five or More CAG Repeats

Name of clone	Number of repeats	Genbank accession no. ^a	Homologies on BLAST search ^b	Genbank accession no. ^c
DT1P1A2	(CAG) ₆	U92978	Human nucleobindin gene (91%)	M96824
DT1P1A4	(CTG) ₆	U92979	Glial fibrillary acidic protein (93%)	J04569
DT1P2A7	(CTG) ₁₀	U92985	Human clone mcag 11 (84%)	U23859
DT1P1A10	(CAG) ₅	U92980	None	—
DT2P1A11	CAA(CAG) ₅ (CGG) ₄ (CAG) ₅	U92992	None	—
DT1P1B6	(CAG) ₆	U92981	None	—
DT1P1B10	(CTG) ₁₅	U92982	Human mRNA/CTG repeat (100%) Ribosomal protein (99%)	D87735
DT1P1B11	(CAG) ₄ CAA(CAG) ₆ (CAA) ₂ CAG	U92983	mRNA for PQ-rich protein (99%) Mouse TDAG51 mRNA (84%)	Z50194 U44088
DT1P1B12	(CAG) ₂ (CGG) ₆ (CAG) ₆	U92984	<i>Homo sapiens</i> cytoplasmic antiprotein (CAP2) (97%)	L403771
DT2P1C4 ^d	CAGCAA(CAG) ₃ CAACAGCAA (CAG) ₇	U92994	Human MAR/SAR DNA binding protein (99%) Nuclear matrix attachment DNA binding protein (88%)	M97287
DT1P2C6	(CTG) ₇	U92986	None	—
DT1P2C12	(CAG) ₂ CTACAG CGA(CAG) ₆ CGA (GAG) ₂ (CAG) ₈ CGA GAGCGG(CAG) ₄ GAG(CAG) ₂ (CAA) ₂ (CAG) ₂	U92987	None	—
DT1P2D1	(CAG) ₄ CAACGG TAG(CAG) ₆ TAG (CAG) ₂	U93216	Human autoantigen mRNA (96%)	U17474
DT1P2D2 ^d	(CAG) ₈ CACCTG CCG(CAG) ₂ CAA (CAG) ₂ (CAACAG) ₄ CAA(CAG) ₂	U92988	mRNA for signal recognition subunit 14 (99%)	X73459
DT1P2E11	(CAG) ₇	U92989	None	—
DT2P1F5	(CAG) ₆	U92994	Human IFN-responsive transcription factor (91%)	M875031
DT2P1F10	(CGG) ₄ (CAG) ₇ CAA	U92990	<i>H. sapiens</i> splicing factor	L10910
DT2P1G7	(CAG) ₅	U92995	None	—
DT1P2H10	(CAG) ₆	U92991	Human mRNA nucleosome assembly (100%)	D50370

^a Accession no. for clones in this study.

^b Numbers in parentheses indicate percentage of sequence identity over the region compared.

^c Accession no. of BLAST sequence homologies.

^d Identified to be homologous to clones a04500s and g04502m, respectively, in Bulle *et al.*, 1997, after submission of this article.

Figs. 2A, 2B, 2C, and 2D, respectively. One of the clones (DT1P1A4) also showed brain-specific expression (Fig. 2B). Dot blot hybridization to RNA derived from 14 subregions of the brain and the spinal cord indicated that DT1P1A4 expression is highest in the medulla oblongata, substantia nigra, and spinal cord (data not shown). Expression in fetal brain was also detected for this clone. DT1P2H10 also showed relatively high expression levels in testes and brain, particularly in the substantia nigra and spinal cord (data not shown).

Polymorphism of (CAG)_n Clones

The degree of polymorphism for each of the (CAG)_n repeats was assessed on at least 80 chromosomes from the CEPH reference families. Table 4 presents a summary of these results, indicating that 8/17 clones are polymorphic. The heterozygosity index for the polymorphic clones ranged from 0.12 to 0.73. Five of 8 clones

have a heterozygosity index greater than 0.5. The number of alleles for the clones with polymorphic CAG repeats ranged from 2 to 6 (data not shown).

DISCUSSION

In this study we have identified cDNA clones from adult human brain that contain variable-length (CAG)_n repeats, using the oligo capture method. Some of the clones identified using our approach were distinct from clones obtained by large-scale sequencing of fetal brain and infant brain cDNA libraries in the EST database, which is consistent with previous observations that the genes expressed in the human brain show variable developmental expression patterns (Adams *et al.*, 1995; Néri *et al.*, 1996). Our use of the adult human brain cDNA library thus complements previous efforts to catalog (CAG)_n clones (Jiang *et al.*, 1995; Li *et al.*, 1993; Riggins *et al.*, 1992) from fetal brain.

TABLE 2
PCR Primers and Conditions Used for CAG Repeat Amplification

Name of clone	Primer sequences	PCR product in bp	Annealing temperature
DT1P1A2	SN 5'-cacatggagcagcggaagcagcag-3' ASN 5'-tgcccctcagggtgggcagc-3'	73	Two step
DT1P1A4	SN 5'-cagtgccctgaagattagcagcag-3' ASN 5'-gtccttccctccactccttct-3'	63	55
DT1P2A7	SN 5'-aggtcagcttggtgctgagctct-3' ASN 5'-aaagggaaaagtgagaggtctctgctg-3'	105	65
DT1P1A10	SN 5'-ctagatgccggtggacagtgaggca-3' ASN 5'-cttcacttgcatgttcatctcac-3'	120	55
DT1P1B6	SN 5'-acatgtgctcctggatcctcccctg-3' ASN 5'-aatgagccctgctgtgggcagcc-3'	109	55
DT1P1B10	SN 5'-agetctcccaaaaaagcactgg-3' ASN 5'-aggaaccttctgggctggagc-3'	141	65
DT1P1B11	SN 5'-cgaggaagggtctgcttatecc-3' ASN 5'-cggatcggccggccctgcccggg-3'	108	65
DT2P1C4	SN 5'-ccgcccacattatccatgttcca-3' ASN 5'-gtggctgctgctggtgtggag-3'	127	65
DT1P2C6	SN 5'-tagaagccactgttcttataaacaca-3' ASN 5'-gtccattgtgaaattatcccagca-3'	118	65
DT1P2C12	SN 5'-agatgctgctgttgtgctgctgc-3' ASN 5'-tggagaagggtcggtgcagcagct-3'	129	65
DT1P2D1	SN 5'-gccttcatgacacctagcgcctcc-3' ASN 5'-gctgtaaggagctgctgctgctg-3'	107	65
DT1P2D2	SN 5'-agaacaaaactaagaagaccaaagca-3' ASN 5'-ttgctgctgctggcaggtgctgctg-3'	66	65
DT1P2E11	SN 5'-cgtcttcaggcagtgacagtggg-3' ASN 5'-gccctctggcctgtccgggctat-3'	81	60
DT2P1F5	SN 5'-cattcagacattgggagcagcagc-3' ASN 5'-tgtctgaatcttctgtggctcag-3'	60	60
DT1P2F10	SN 5'-agctgccgaagccgctctctgct-3' ASN 5'-cccgaatatcgggttccaagcagc 3	104	65
DT2P1G7	SN 5'-gaagaatcgacataggtccagacc-3' ASN 5'-tattagcgtaggatgctgctgctg-3'	76	65
DT1P2H10	SN 5'-gtgattctggggaagaatctgacagc-3' ASN 5'-ccgctgccgctgctgctccacta-3'	118	65

TABLE 3
Chromosomal Map Location for CAG Repeat-Containing cDNA Clones

Name of clone	Chromosome location	Between marker 1	and marker 2
DT1P1A2	19q13.32	WI-9028	NIB1805
DT1P1A4	17q21.1-q25.1	D17S1970	D17S2047
DT1P2A7	1q12	WI-9711	D1S305
DT1P1A10	Xp11.22	DXS1013E	
DT1P1B6	2p23.1	GCT11G10	CHLC.ATC5A0
DT1P1B10-1	3p21.33	WI-4073	RP_SA_1
DT1P1B10-2	12q21.1	CHLC.GATA7A108	AFMa066xB9
DT1P1B11	12q21.2	CHLC.GATA26D02	WI-5371
DT1P1B12	18q22.1-q22.2	D18S68	D18S477
DT2P1C4	3p24.3	RP_L15_1	D3S1583
DT1P2C6	1p33	WI-11021	WI-6608
DT1P2C12	5, 6	?	?
DT1P2D1	16	D16S3211	
DT1P2D2	15q15.1-q21.1	D15S144	NIB1832
DT1P2E11	17q24.1	D17S2104	D17S1352
DT2P1F5	14q11.2	D14S264	D14S275
DT2P1F10	6, 9	?	?
DT2P1G7	11q11	WI-8652	CHLC.GATA46
DT1P2H10	Xq21.33	DXS990	

Our strategy allows relatively rapid screening of approximately 10^6 to 10^7 clones in a few days compared to several weeks for conventional radioactive screening of cDNA libraries and can be adapted to identify members of gene families that share certain motif sequences. BLAST analysis of sequences from 119 randomly picked clones indicated cross-homologies with one other, as well as with a number of previously identified genes, such as the gene for SCA3 or the Machado-Joseph disease gene. However, cDNAs for SBMA, HD, SCA1, DRPLA, and SCA2 were not identified in our search. In these diseases, the (CAG)_n repeat is positioned very close to the 5' end of their respective transcripts, unlike that for SCA3 of which the repeat is near the 3' end of the cDNA. This suggests that the adult human brain cDNA library used in this study is enriched for clones for the 3' end of genes. This is further corroborated by the observation of relatively short insert clones in this library. Further oligo capture using random-primed or large-insert cDNA libraries will most likely yield additional unique (CAG)_n clones.

The selected cDNAs described here showed a wide range of expression in several different human tissues. Only the clone DT1P1A4, which showed homology to

glial fibrillary acidic protein, has a brain-specific expression.

It has been suggested that triple repeats of eight or more in noncoding sequences are likely to be polymorphic (Gastier *et al.*, 1996). Though repeat size has been shown to be a strong indicator of polymorphic length variability for (CA)_n dinucleotide repeats (Weber, 1990), it is possible that the repeats found in coding regions are under strong selection constraints for either expansion or contraction. A recent example has been described for SCA2, in which the CAG repeat size of 22 is the frequent allele in the general population, representing about 90% of the 506 alleles tested (Zoghbi, 1996). Despite the low heterozygosity index, the repeat in this gene is capable of expanding to a pathogenic size in a subset of spinocerebellar ataxia patients. In the cases of SCA1 and SCA2, interruptions within the (CAG)_n repeat tract stabilize the repeat but such interruptions are lost in the expanded alleles. Thus in searching for the potential expansions in cDNAs, those with interruptions and those that show low polymorphic information content should not be excluded as potential candidate genes.

Recently the loci for SCA5 on chromosome 11 (Ranum *et al.*, 1994), SCA7 on chromosome 3 (David *et al.*, 1996; Holmberg *et al.*, 1995), and bipolar disorder on chromosome 18 (Freimer *et al.*, 1996; McInnes *et al.*, 1996) have been mapped by linkage analysis. While the identification of the genes responsible for these disorders are being actively pursued in these laboratories,

TABLE 4

Transcript Size and CAG Repeat Polymorphism of cDNA Clones

Clone	Transcript size (in kb)	Available insert size in pcMVSPORT (in kb)	Heterozygosity
DT1P1A2	2.4	1.4	0
DT1P1A4	3.5	0.8	0
DT1P2A7	1.5	1.1	0.66
DT1P1A10	1.5	1.4	0.50
DT1P1B6	9.6	1.4	0
DT1P1B10	1.0	0.6	0.73
Chromosome 3			
Chromosome 12	1.0	0.12 ^a	0.18
DT1P1B11	6.6	1.5	0
DT1P1B12	6.0	1.2	ND ^b
DT2P1C4	0.7	0.7	0.50
DT1P2C6	5.4	0.8	0.12
DT1P2C12	4.6	1.5	0.50
DT1P2D1	12.0	0.4	0
DT1P2D2	4.7	1.0	0.16
DT1P1E11	2.6	0.8	0
DT2P1F5	4.4	0.9	0
DT2P1F10	4.9	2.2	ND
DT2P1G7	1.3	1.3	0
DT1P2H10	2.9	1.2	0

^a PCR product cloned into TA vector.

^b ND, not determined.

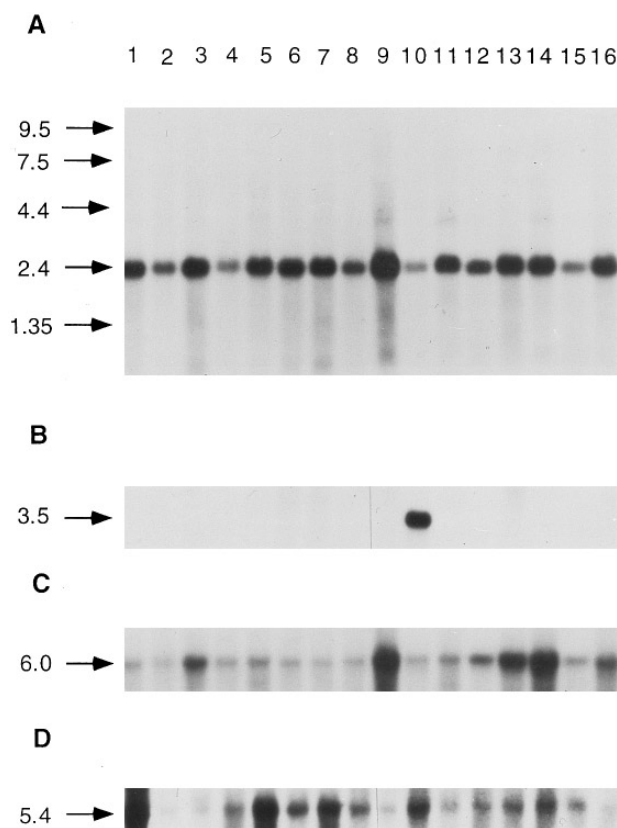


FIG. 2. Northern blot analysis of clone (A) DT1P1A2, (B) DT1P1A4, (C) DT1P1B12, and (D) DT1P2C6. Lanes for RNA sources are (1) spleen, (2) thymus, (3) prostate, (4) testis, (5) ovary, (6) small intestine, (7) colon, (8) leukocyte, (9) heart, (10) brain, (11) placenta, (12) lung, (13) liver, (14) skeletal muscle, (15) kidney, and (16) pancreas.

it is possible that some of the cDNAs described in this study (Table 3) may be suitable candidate genes for these diseases. Of the cDNA clones we have isolated, 18/19 have been placed on the chromosome map. The chromosomal locations for some of these cDNAs coincide with the critical regions for a number of neurodegenerative disorders. For example, the cDNA clone DT2P1G7 was positioned in the region that shows linkage for SCA5. This position was also confirmed further by somatic cell hybrid mapping of chromosome 11 derivatives (data not shown). The cDNA clones DT1P1B10 and DT2P1C4 have been mapped to the region on the short arm for chromosome 3 to which SCA7 or autosomal dominant cerebellar ataxia type II has been mapped. Likewise DT1P1B12 has been placed in the region of chromosome 18 to which bipolar disorder has been mapped (Detera-Wadleigh *et al.*, 1994; Freimer *et al.*, 1996; McInnes *et al.*, 1996). We may also have identified additional candidate genes for the disorders familial dementia on chromosome 3 (Brown *et al.*, 1995) and schizophrenia on chromosome 18 (Moises *et al.*, 1995; Schwab *et al.*, 1995; Straub *et al.*, 1995; Wang *et al.*, 1995).

In other neurodegenerative diseases where the frequency of the disorder is low, thus precluding linkage

analysis due to the lack of informative pedigrees, a large collection of these CAG-containing cDNAs can be used to screen a collection of affected individuals for evidence of expansions. Such a strategy will substantially reduce efforts to clone the responsible genes through conventional positional cloning methods, but provides the opportunity to go directly from mapped candidate genes to the disease locus. Diseases that show genetic heterogeneity where some of the genes involved may show (CAG)_n repeat expansions, such as bipolar disorder and schizophrenia (Morris *et al.*, 1995; O'Donovan *et al.*, 1995), can also be readily tested with these (CAG)_n cDNAs. In addition, families with autosomal neurodegenerative disorders that show anticipation can be tested for expansions with these cDNAs, precluding the chromosomal linkage analysis of these families. These diseases may include Cowden syndrome (Hanssen *et al.*, 1993), spastic paraplegia (Burger *et al.*, 1996; Hazan *et al.*, 1993), familial Parkinson's disease (Bonifati *et al.*, 1994, 1995), and zonal cataract (Marnier, 1949; Eiberg *et al.*, 1988). Thus the (CAG)_n clones described provide a gene-based STS as well as a rich resource for candidate genes on which other inherited neurological disorders can be tested for repeat expansions.

ACKNOWLEDGMENTS

We thank Lawrence Brody and William Pavan for comments and for critically reading the manuscript and Settara Chandrasekarappa for providing somatic cell hybrid lines for fine mapping of DT2P1G7 in the SCA5 locus. We also thank Jeff Wilderson and Jane Ellison for their help in sequencing the clones and Elvin Hsu for database management.

REFERENCES

- Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult, C. J., Lee, N. H., Kirkness, E. F., Weinstock, K. G., Gocayne, J. D., White, O., Sutton, G., Blake, J. A., Brandon, R. C., Chiu, M.-L., Clayton, R. A., Cline, R. T., Cotton, M. D., Earle-Hughes, J., Fine, L. D., Fitzgerald, L. M., Fitzhugh, W. M., Fritchman, J. L., Geogaghan, N. S. M., Glodek, A., Gnehm, C. L., Hanna, M. C., Hedblom, E., Hinkle, P. S., Jr., Kelley, J. M., Klimek, K. M., Kelley, J. C., Liu, L., Marmoros, S. M., Merrick, J. M., Moreno-Palauques, R. F., McDonald, L. A., Nguyen, D. T., Pellegrino, S. M., Philillips, C. A., Ryder, S. E., Scott, J. L., Saudek, D. M., Shirley, R., Small, K. V., Spriggs, T. A., Utterback, T. R., Weidman, J. F., Li, Y., Barthlow, R., Bednarik, D. P., Cao, L., Cepeda, M. A., Coleman, T. A., Collins, E.-J., Dimke, D., Feng, P., Ferrie, A., Fischer, C., Hastings, G. A., He, W.-W., Hu, J.-S., Huddleston, K. A., Greene, J. M., Gruber, J., Hudson, P., Kim, A., Kozak, L., Kunsch, C., Ji, H., Li, H., Meissner, P. S., Olsen, H., Raymond, L., Wei, Y. F., Wing, J., Xu, C., Yu, G. L., Ruben, S. M., Dillon, P. J., Fannon, M. R., Rosen, C. A., Haseltine, W. A., Fields, C., Fraser, C. M., and Venter, J. C. (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377**: 3–174.
- Albin, R. L., and Tagle, D. A. (1995). Genetics and molecular biology of Huntington's disease. *Trends Neurosci.* **18**: 11–14.
- 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.
- Bonifati, V., Fabrizio, E., Vanacore, N., De Mari, M., and Meco, G. (1995). Familial Parkinson's disease: A clinical genetic analysis. *Can. J. Neurol. Sci.* **22**: 272–279.
- Bonifati, V., Vanacore, N., and Meco, G. (1994). Anticipation of onset age in familial Parkinson's disease. *Neurology* **44**: 1978–1979.
- 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., Zemelmann, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourn, 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.
- Brown, J., Ashworth, A., Gydesen, S., Sorensen, A., Rossor, M., Hardy, J., and Collinge, J. (1995). Familial non-specific dementia maps to chromosome 3. *Hum. Mol. Genet.* **4**: 1625–1628.
- Bulle, F., Chiannikulchai, N., Pawlak, A., Weissenbach, J., Gyapay, G., and Guellaen, G. (1997). Identification and chromosomal location of human genes containing CAG/CTG repeats expressed in testis and brain. *Genome Res.* **7**: 705–715.
- Burger, J., Metzke, H., Paternotte, C., Schilling, F., Hazan, J., and Reis, A. (1996). Autosomal dominant spastic paraplegia with anticipation maps to a 4-cM interval on chromosome 2p21–p24 in a large German family. *Hum. Genet.* **98**: 371–375.
- Burke, J. R., Wingfield, M. S., Lewis, K. E., Roses, A. D., Lee, J. E., Hulette, C., Pericak-Vance, M. A., and Vance, J. M. (1994). The Haw River syndrome: Dentatorubropallidolusian atrophy (DRPLA) in an African-American family. *Nature Genet.* **7**: 521–524.
- Campuzano, V., Montermini, L., Molto, M. D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Canizares, J., Koutnikova, H., Bidichandani, S. I., Gellera, C., Brice, A., Trouillas, P., Michele, G. D., Filla, A., Frutos, R. D., Palau, F., Patel, P. I., Donato, S. D., Mandel, J. L., Corozza, S., Koenig, M., and Pandolfi, M. (1996). Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**: 1423–1427.
- David, G., Giunti, P., Abbas, N., Coullin, P., Stevanin, G., Horta, W., Gemmill, J., Weissenbach, R., Wood, N., Cunha, S., Drabkin, H., Harding, A. E., Agid, Y., and Brice, A. (1996). The gene for autosomal dominant cerebellar ataxia type II is located in a 5-cM region in 3p12–p13: Genetic and physical mapping of the SCA7 locus. *Am. J. Hum. Genet.* **59**: 1328–1336.
- Detera-Wadleigh, S. D., Hsieh, W. T., Berrettini, W. H., Goldin, L. R., Rollins, D. Y., Muniec, D., Grewal, R., Guroff, J. J., Turner, G., Coffman, D., Barrick, J., Mills, K., Murray, J., Donohue, S. J., Klein, D. C., Sanders, J., Nurenberger, J. I., Jr., and Gershan, E. S. (1994). Genetic linkage mapping for a susceptibility locus to bipolar illness: Chromosomes 2, 3, 4, 7, 9, 10p, 11p, 22, and Xpter. *Am. J. Med. Genet.* **54**: 206–218.
- Eiberg, H., Marnier, E., Rosenberg, T., and Mohr, J. (1988). Marnier's cataract (CAM) assigned to chromosome 16: Linkage to haptoglobin. *Clin. Genet.* **34**: 272–275.
- Freimer, N. B., Reus, V. I., Escamilla, M. A., McInnes, L. A., Spesny, M., Leon, P., Service, S. K., Smith, L. B., Silva, S., Rojas, E., Gallegos, A., Meza, L., Fournier, E., Baharloo, S., Blankenship, K., Tyler, D. J., Batki, S., Vinogradov, S., Weissenbach, J., Barondes, S. H., and Sandkuijl, L. A. (1996). Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22–q23. *Nature Genet.* **12**: 436–441.
- Fu, Y.-H., Pizzuti, A., Fenwick, R. G., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., De Jong, P., Wieringa, B., Korneluk, R., Perryman, M. B., Epstein, H. F., and Caskey, C. T. (1992). An unstable triplet repeat in a gene related to myotonic dystrophy. *Science* **255**: 1256–1258.
- Gastier, J. M., Brody, T., Pulido, J. C., Businga, T., Sundén, S., Hu, X., Maitra, S., Buetow, K. H., Murray, J. C., Sheffield, V. C., Boguski, M., Duyk, G. M., and Hudson, T. J. (1996). Development of a screening set for new (CAG/CTG)_n dynamic mutations. *Genomics* **32**: 75–85.

- Haaf, T., Sirugo, G., Kidd, K. K., and Ward, D. C. (1996). Chromosomal localization of long trinucleotide repeats in the human genome by fluorescence in situ hybridization. *Nature Genet.* **12**: 183–185.
- Hanssen, A. M., Werquin, H., Suys, E., and Fryns, E. P. (1993). Cowden syndrome: Report of a large family with macrocephaly and increased severity of signs in subsequent generations. *Clin. Genet.* **44**: 281–286.
- Hazan, J., Lamy, C., Melki, J., Munnich, A., de Recondo, J., and Weissenbach, J. (1993). Autosomal dominant familial spastic paraplegia is genetically heterogeneous and one locus maps to chromosome 14q. *Nature Genet.* **5**: 163–167.
- Holmberg, M., Johansson, J., Forsgren, L., Heijbel, J., Sandgren, O., and Holmgren, G. (1995). Localization of autosomal dominant cerebellar ataxia associated with retinal degeneration and anticipation to chromosome 3p12–p21.1. *Hum. Mol. Genet.* **4**: 1441–1445.
- Hummerich, H., and Lehrach, H. (1995). Trinucleotide repeat expansion and human disease. *Electrophoresis* **16**: 1698–1704.
- Imbert, G., Kretz, C., Johnson, K., and Mandel, J. L. (1993). Origin of the expansion mutation in myotonic dystrophy. *Nature Genet.* **4**: 72–76.
- Imbert, G., Saudou, F., Yvert, G., Devys, D., Trottier, Y., Garnier, J. M., Weber, C., Mandel, J. L., Cancel, G., Abbas, N., Dürr, A., Didierjean, O., Stevanin, G., Agid, Y., and Brice, A. (1996). Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nature Genet.* **14**: 285–291.
- Jiang, J. X., Deprez, R. H., Zwarthoff, E. C., and Riegman, P. H. (1995). Characterization of four novel CAG repeat-containing cDNAs. *Genomics* **30**: 91–93.
- Jones, C., Slijepcevic, P., Marsh, S., Baker, E., Langdon, W. Y., Richards, R. I., and Tunnacliffe, A. (1994). Physical linkage of the fragile site FRA11B and a Jacobsen syndrome chromosome deletion breakpoint in 11q23.3. *Hum. Mol. Genet.* **3**: 2123–2130.
- Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Aki-guchi, I., Kimura, J., Narumiya, S., and Kakizuka, A. (1994). CAG expansions in a novel gene for Machado–Joseph disease at chromosome 14q32.1. *Nature Genet.* **8**: 221–228.
- Kendler, K. S., Straub, R. E., MacLean, C. J., and Walsh, D. (1996). Reflections on the evidence for a vulnerability locus for schizophrenia on chromosome 6p24–22. *Am. J. Med. Genet.* **67**: 124–126.
- Knight, S. J., Flannery, A. V., Hirst, M. C., Campbell, L., Christodoulou, Z., Phelps, S. R., Pointon, J., Middleton-Price, H. R., Barnicoat, A., Pembrey, M. E., Holland, J., Oostra, B. A., Bobrow, M., and Davies, K. E. (1993). Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74**: 127–134.
- Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., Saito, M., Tomoda, A., Miike, T., Naito, H., Ikuta, F., and Tsujii, S. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral–pallidolusian atrophy (DRPLA). *Nature Genet.* **6**: 9–13.
- La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding A. E., and Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**: 77–79.
- Li, S. H., McInnis, M. G., Margolis, R. L., Antonarakis, S. E., and Ross, C. A. (1993). Novel triplet repeat containing genes in human brain: Cloning, expression, and length polymorphisms. *Genomics* **16**: 572–579.
- Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, C., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., Leblond, S., Earle-MacDonald, J., De Jong, P. J., Wieringa, B., and Korneluk, R. G. (1992). Myotonic dystrophy mutation: An unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**: 1253–1255.
- Margolis, R. L., Abraham, M. R., Gatchel, S. B., Li, S. H., Kidwai, A. S., Breschel, T. S., Stine, O. C., Callahan, C., McInnis, M. G., and Ross, C. A. (1997). cDNAs with long CAG trinucleotide repeats from human brain. *Hum. Genet.* **100**: 114–122.
- Marner, E. (1949). A family with eight generations of hereditary cataract. *Acta. Ophthalmol.* **27**: 537–551.
- McInnes, L. A., Escamilla, M. A., Service, S. K., Reus, V. I., Leon, P., Silva, S., Rojas, E., Spesny, M., Baharloo, S., Blankenship, K., Peterson, A., Tyler, D., Shimayoshi, N., Tobey, C., Batki, S., Vinograd, S., Meza, L., Gallegos, A., Fournier, E., Smith, L. B., Baron-des, S. H., Sandkuij, L. A., and Freimer, N. B. (1996). A complete genome screen for genes predisposing to severe bipolar disorder in two Costa Rican pedigrees. *Proc. Natl. Acad. Sci. USA* **93**: 13060–13065.
- Moises, H. W., Yang, L., Kristbjarnarson, H., Wiese, C., Byerley, W., Macciardi, F., Arold, V., Blackwood, D., Liu, X., Sjogren, B., Aschauer, H. N., Hwu, H. G., Jang, K., Livesley, W. J., Kennedy, J. L., Zoega, T., Ivarsson, O., Bui, M. T., Yu, M. H., Havsteen, B., Commenges, D., Weissenbach, J., Scwinger, E., Gottesman, I. I., Pakistis, A. J., Wetterberg, L., Kidd, K. K., and Helgason, T. (1995). An international two-stage genome wide search for schizophrenia susceptibility genes. *Nature Genet.* **11**: 321–324.
- Morris, A. G., Gaitonde, E., McKenna, P. J., Mollon, J. D., and Hunt, D. M. (1995). CAG repeat expansions and schizophrenia: Association with disease in females and with early age-at-onset. *Hum. Mol. Genet.* **4**: 1957–1961.
- Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I., Murayama, N., Tanaka, Y., Kukishima, H., Umino, K., Furukawa, T., Nihei, K., Inoue, T., Sano, A., Komure, O., Takahashi, M., Yoshizawa, T., Kanazawa, I., and Yamada, M. (1994). Dentatorubral and pallidolusian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nature Genet.* **6**: 14–18.
- Nancarrow, J. K., Kremer, E., Holman, K., Eyre, H., Doggett, N. A., Le Paslier, D., Callen, D. F., Sutherland, G. R., and Richards, R. I. (1994). Implications of FRA16A structure for the mechanism of chromosomal fragile site genesis. *Science* **264**: 1938–1941.
- Néri, C., Albanèse, V., Lebre, A. S., Holbert, S., Saada, C., Bougueret, L., Meier-Ewert, S., Le Gall, I., Millasseau, P., Bui, H., Giudicelli, C., Massart, C., Guillou, S., Gervy, P., Poullier, E., Rigault, P., Weissenbach, J., Lennon, G., Chumakov, I., Dausset, J., Lehrach, H., Cohen, D., and Cann, H. M. (1996). Survey of CAG/CTG repeats in human cDNAs representing new genes: Candidates for inherited neurological disorders. *Hum. Mol. Genet.* **5**: 1001–1009.
- O'Donovan, M. C., Guy, C., Craddock, N., Murphy, K. C., Cardno, A. G., Jones, L. A., Owen, M. J., and McGuffin, P. (1995). Expanded CAG repeats in schizophrenia and bipolar disorder. *Nature Genet.* **10**: 380–381.
- Orr, H. T., Chung, M. Y., Banfi, S., Kwiatkowski, T. J., Servadio, A., Jr., Beaudet, A. L., McCall, M. E., Duvick, L. A., Ranum, L. P., and Zoghbi, H. Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genet.* **4**: 221–226.
- Parrish, J. E., Oostra, B. A., Verkerk, A. J., Richards, C. S., Reynolds, J., Spikes, A. S., Shaffer, L. G., and Nelson, D. L. (1994). Isolation of a GCC repeat showing expansion in FRAXF, a fragile site distal to FRAXA and FRAXE. *Nature Genet.* **8**: 229–235.
- Potter, N. T., Yanagisawa, H., and Yamada, M. (1996). Different origins of expanded repeats for Haw River syndrome and dentatorubral–pallidolusian atrophy. *Lancet* **347**: 1271.
- Ranum, L. P., Schut, L. J., Lundgren, J. K., Orr, H. T., and Livingston, D. M. (1994). Spinocerebellar ataxia type 5 in a family descended from the grandparents of President Lincoln maps to chromosome 11. *Nature Genet.* **8**: 280–284.
- Riggins, G. J., Lokey, L. K., Chastain, J. L., Leiner, H. L., Sherman, S. L., Wilkinson, K. D., and Warren, S. T. (1992). Human genes containing polymorphic trinucleotide repeats. *Nature Genet.* **2**: 186–191.
- Sanpei, K., Takano, H., Igarashi, S., Sato, T., Oyake, M., Sasaki, H., Wakisaka, A., Tashiro, K., Ishida, Y., Ikeuchi, T., Koide, R., Saito,

- M., Sato, A., Tanaka, T., Hanyu, S., Takiyama, Y., Nishizawa, M., Shimizu, N., Nomura, Y., Segawa, M., Iwabuchi, K., Eguchi, I., Tanaka, H., Takahashi, H., and Tsuji, S. (1996). Identification of the spinocerebellar ataxia type 2 gene using a direct identification of a repeat expansion and cloning technique, DIRECT. *Nature Genet.* **14**: 277–284.
- Schalling, M., Hudson, T. J., Buetow, H. K., and Housman, D. E. (1993). Direct detection of novel expanded trinucleotide repeats in the human genome. *Nature Genet.* **4**: 135–139.
- Schwab, S. G., Albus, M., Hallmayer, J., Honig, S., Borrmann, M., Lichtermann, D., Ebstein, R. P., Ackenheil, M., Lerer, B., Risch, N., Maier, W., and Wildnauer, D. B. (1995). Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. *Nature Genet.* **11**: 325–327.
- Smith, S., Overbeek, R., Woese, C. R., Gilbert, W., and Gillet, P. (1994). The genetic data environment: An expandable GUI for multiple sequence analysis. *CABIOS* **10**: 671–675.
- Stevanin, G., Trottier, Y., Cancel, G., Durr, A., David, G., Didierjean, O., Burk, K., Imbert, G., Saudou, F., Abada-Bendib, M., Gourfinkel-An, I., Benomar, A., Abbas, N., Klockgether, T., Grid, D., Agid, Y., Mandel, J. L., and Brice, A. (1996). Screening for proteins with polyglutamine expansions in autosomal dominant cerebellar ataxias. *Hum. Mol. Genet.* **5**: 1887–1892.
- Straub, R. E., MacLean, C. J., O'Neill, F. A., Burke, J., Murphy, B., Duke, F., Shinkwin, R., Webb, B. T., Zhang, J., Walsh, D., and Kenneth, K. S. (1995). A potential vulnerability locus for schizophrenia on chromosome 6p24–22: Evidence for genetic heterogeneity. *Nature Genet.* **11**: 287–293.
- The Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**: 971–983.
- Trottier, Y., Lutz, Y., Stevanin, G., Imbert, G., Devys, D., Cancel, G., Saudou, F., Weber, C., David, G., Tora, L., Agid, V., Brice, A., and Mendel, J-L. (1995). Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. *Nature* **378**: 403–405.
- Verkerk, A. J. M. H., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F. P., Eussen, B. E., 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 CCG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**: 905–914.
- Wang, S., Sun, C. E., Walczak, C. A., Ziegler, J. S., Kipps, B. R., Goldin, L. R., and Diehl, S. R. (1995). Evidence for a susceptibility locus for schizophrenia on chromosome 6pter–p22. *Nature Genet.* **10**: 41–46.
- Warren, S. T. (1996). The expanding world of trinucleotide repeats. *Science* **271**: 1374–1375.
- Weber, J. L. (1990). Informativeness of human (dC–dA)_n·(dG–dT)_n polymorphisms. *Genomics* **7**: 524–530.
- Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D. W., Amos, C., Dobyns, W. B., Subramony, S. H., Zoghbi, H. Y., and Lee, C. C. (1997). Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nature Genet.* **15**: 62–69.
- Zoghbi, H. Y. (1996). The expanding world of ataxins. *Nature Genet.* **14**: 237–238.