PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68 — A2

(11) International Publication Number: **WO 00/24938**

(43) International Publication Date: 4 May 2000 (04.05.00)

(21) International Application Number: PCT/US99/25119

(22) International Filing Date: 27 October 1999 (27.10.99)

(30) Priority Data:

60/105,885	27 October 1998 (27.10.98)	US
Not furnished	26 October 1999 (26.10.99)	US

(71) Applicant *(for all designated States except US)*: THE JOHNS HOPKINS UNIVERSITY [US/US]; Suite 906, 111 Market Place, Baltimore, MD 21201 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: MARGOLIS, Russell [US/US]; The Johns Hopkins University, Suite 906, 111 Market Place, Baltimore, MD 21201 (US). ROSS, Christopher [US/US]; The Johns Hopkins University, Suite 906, 111 Market Place, Baltimore, MD 21201 (US). NISSON, Paul, B. [US/US]; The Johns Hopkins University, Suite 906, 111 Market Place, Baltimore, MD 21201 (US). LI, Wu, B. [US/US]; The Johns Hopkins University, Suite 906, 111 Market Place, Baltimore, MD 21201 (US).

(74) Agents: KAGAN, Sarah, A. et al.; Banner & Witcoff, Ltd., 11th floor, 1001 G Street, N.W., Washington, DC 20001–4597 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: CCG REPEATS IN cDNAs FROM HUMAN BRAIN

(57) Abstract

Expansion mutations of trinucleotide repeats and other units of unstable DNA have been proposed to account for at least some of the genetic susceptibility to a number of neuropsychiatric disorders, including bipolar affective disorder, schizophrenia, autism, and panic disorder. To generate additional candidate genes for these and other disorders, cDNA libraries from humain brain were probed at high stringency for clones containing CCG, CGC, GCC, CGG, GCG, and GGC repeats (referred to collectively as CCG repeats). 18 cDNAs containing previously unpublished or uncharacterized repeats were characterized for chromosomal locus, repeat length polymorphism, and similarity to genes of known function. The cDNAs were also compared to the 37 human genes in GenBank with eight or more consecutive CCG triplets. The repeats were mapped to a number of loci, including 1p34, 2p11.2, 2q30–32, 3p21, 3p22, 4q35, 6q22, 7qter, 13p13, 17q24, 18p11, 19p13.3, 20q12, 20q13.3 and 22q12. Length polymorphism was detected in 50 % of the repects. The newly cloned cDNAs include a complete transcript of human neurexin 1B, a portion of BCNG–1 (a newly described brain–specific ion channel), a previously unreported polymorphic repeat located in the 5' UTR region of the guanine nucleotide–binding protein (G–protein)β2 subunit, and a human version of the mouse proline rich protein 7.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

CCG Repeats in cDNAs From Human Brain

This application claims the benefit of copending provisional application no. 60/105,885 filed October 27, 1998, which is expressly incorporated by reference in its entirety herein.

This invention was made using funds supplied by the United States government. Under the terms of grants from the National Institutes of Health, MH01275 and MH50763, the U.S. government may retain certain rights in the invention.

TECHNICAL FIELD OF THE INVENTION

This invention is related to diseases which result from expansion of microsatellite repeats. Such diseases are primarily of the central nervous system.

BACKGROUND OF THE INVENTION

Seventeen human disorders caused by trinucleotide repeat expansion or related mutations of repetitive DNA have been identified over the past seven years. Eight genes contain an expansion of a CAG repeat encoding glutamine, resulting in a group of eight dominant neurodegenerative disorders (Ross, 1997; Ross et al. 1998). Inheritance of two alleles with an expanded intronic GAA expansion in the frataxin gene causes most cases of Friedreich's ataxia (Campuzano et al. 1996), and a 3' untranslated CTG repeat in the DMPK gene causes myotonic dystropy (Brook et al. 1992; Mahadevan et al. 1992).

Six different disorders are associated with mutations in the length of CCG, CGC, GCC, CGG, GCG, and GGC repeats (here collectively referred to as CCG repeats). Large expansions of 5' untranslated CCG repeats cause fragile X A (FMR1 gene) (Yu et al. 1992) and fragile X E (FMR2 gene) (Knight et al. 1993). These disorders are characterized phenotypically by cognitive and psychiatric abnormalities, and genetically by the phenomenon of anticipation. Large expansions of a 5'-UTR CCG repeat in the CBL2 protooncogene also result in some cases of Jacobsen's syndrome, an 11q deletion

syndrome (Jones et al. 1995). Relatively small changes in the number of (GCN)n triplets encoding alanine cause several different disorders. A form of the developmental disorder cleidocranial dysplasia results from insertion of (GCN)10 (encoding 10 additional alanines) into a GCG repeat in core binding factor alpha1 subunit A (CBFA1) (Mundlos et al. 1997). In HOXD13, an increase from (GCN)14 (encoding polyalanine) to (GCN)21-28 results in synpolydactyly (Muragaki et al. 1996). Pedigrees with longer alanine expansions tend to have a more severe phenotype, suggesting that the insertion leads to a gain-of-function that alters the transcriptional regulatory function of this gene (Goodman et al. 1997). Most recently, an increase from (GCG)6 to (GCG)8-13 in PABP2 was found to cause the autosomal dominant form of oculopharyngeal muscular dystrophy (OMPD) (Brais et al. 1998). Homozygotes with (GCG)7--a single triplet longer than normal--develop an autosomal recessive form of the disease.

Two other disorders also reflect mutations in C-G rich sequences. An expansion of a dodecamer repeat (CCCCGCCCCGCG)in the 5' flanking region of the cystatin B gene is one of the mutations that can lead to the recessive disease progressive episodic myoclonic epilepsy (EPM1) (Lalioti et al. 1997; LaFreniere et al. 1997). The semidominant mouse mutant hypodactyly, characterized by a deficit in digital arch formation (Mortlock et al. 1996), is caused by a 50 base pair deletion in which the sequence ...CGGCGGCGGC N_{40} CGGCGGCGGC... is reduced to CGGCGGCGGC, presumably from a recombination or misalignment of the two identical trinucleotide repeat-containing regions during replication.

These discoveries have led to the hypothesis that variation in CCG repeat length may cause other diseases characterized by developmental abnormalities and/or anticipation, including neuropsychiatric disorders such as autism, schizophrenia, and bipolar affective disorder (Ross et al. 1993; McInnisã and Margolis, 1998; McInnis, 1996). Previous efforts to find candidate genes with repeats have primarily focused on CAG repeats (Li et al. 1993; Margolis et al. 1997; Reddy et al. 1997; Riggins et al. 1992; Neri et al. 1996; Bulle et al. 1997; Jiang et al. 1995; Albanese et al. 1998; Gastier et al. 1996; Breschel et al. 1997), with some attention to AAT (Margolis et al. 1995b) and CCA (Margolis et al. 1995a) repeats. Identification of candidate CAG repeats has led to the discovery of expansion mutations causing four neurodegenerative diseases:

dentatorubral pallidoluysian atrophy (DRPLA) (Koide et al. 1994; Nagafuchi et al. 1994), spinocerebellar ataxia type 2 (Imbert et al. 1996), spinocerebellar ataxia type 3 (Machado-Joseph disease, Kawaguchi et al. 1994), and spinocerebellar ataxia type 6 (Zhuchenko et al. 1997). Despite the success of using library screens to identify CAG repeats associated with disease, less attention has been paid to the identification of CCG repeats (Li et al. 1993; Riggins et al. 1992; Albanese et al. 1998), even though seven human diseases arise from expansion of CCG (and related) repeats. There is a need in the art for identification of additional microsatellite markers which are involved in diseases of the central nervous system.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide a polynucleotide for detecting a microsatellite marker.

It is another object of the present invention to provide a method of determining a change in number of trinucleotide repeats in a microsatellite marker.

It is an object of the present invention to provide a pair of primers for amplifying a microsatellite marker.

These and other objects of the invention are achieved by providing by one or more of the embodiments described below. In one embodiment a polynucleotide is provided for detecting a microsatellite marker selected from the group consisting of: P12A7, P12E1, P32B10, P32D9, P32H12, P42A5, P42F11, P55G12, P62D12, P72D4, P95B10, CCG43, CCG82, CCG98, CCGFB48, CCGFB60, CCGFB64, and CCGFB84. The polynucleotide comprises at least 12 nucleotides complementary to contiguous nucleotides within 500 nucleotides of a trinucleotide repeat in the microsatellite marker in the human genome.

According to another aspect of the invention a method is provided for determining a change in number of trinucleotide repeats in a microsatellite marker. A polynucleotide is hybridized to a nucleic acid sample of a patient to form a hybridized polynucleotide. The polynucleotide comprises at least 12 nucleotides complementary to contiguous nucleotides within 500 nucleotides of a trinucleotide repeat in a microsatellite marker in the human genome. The size of the hybridized polynucleotide is determined. An increase in the size of the hybridized polynucleotide relative to size

of the polynucleotide hybridized to a nucleic acid sample of a normal human indicates a change in the number of trinucleotide repeats.

According to still another embodiment of the invention a pair of primers is provided for amplifying a microsatellite marker selected from the group consisting of P12A7, P12E1, P32B10, P32D9, P32H12, P42A5, P42F11, P55G12, P62D12, P72D4, P95B10, CCG43, CCG82, CCG98, CCGFB48, CCGFB60, CCGFB64, and CCGFB84. Each primer is complementary to at least 12 contiguous nucleotides which are within 500 nucleotides of a trinucleotide repeat in the microsatellite marker in the human genome. Each primer of the pair is complementary to opposite strands of the microsatellite marker.

In yet another aspect of the invention a method is provided for determining a change in number of trinucleotide repeats in a microsatellite marker comprising. A pair of primers is used to amplify a template comprising a nucleic acid sample of a patient. The pair of primers amplify a microsatellite marker selected from the group consisting of P12A7, P12E1, P32B10, P32D9, P32H12, P42A5, P42F11, P55G12, P62D12, P72D4, P95B10, CCG43, CCG82, CCG98, CCGFB48, CCGFB60, CCGFB64, and CCGFB84. Each primer is complementary to at least 12 contiguous nucleotides which are within 500 nucleotides of a trinucleotide repeat in the microsatellite marker in the human genome. Each primer of the pair is complementary to opposite strands of the microsatellite marker. The size of the microsatellite marker amplified is determined. An increase in size of the amplified microsatellite marker relative to the size of a microsatellite marker amplified using the pair of primers and a template comprising a nucleic acid sample of a normal human indicates a change in the number of trinucleotide repeats.

Thus the present invention provides the art with additional tools for detecting the presence and in some cases the severity of disease-causing trinucleotide repeat expansion mutations.

DETAILED DESCRIPTION

We have found a series of cDNAs derived from human brain that contain previously undescribed or uncharacterized CCG repeats. These cDNAs are useful in the diagnosis and evaluation of neuropsychiatric diseases.

Rearrangements of microsatellite markers can be detected by Southern blotting, PCR amplification, or any other technique known in the art for observing particular segments of DNA. Rearrangements typically involve an increase or decrease in the copy number of the repeated sequence, more typically an increase. For analysis of size of such markers one typically generates fragments of defined length. This can be done using restriction endonucleases, or PCR amplification, for example. Any other types of reactions which generate fragments of defined length as are known in the art can also be used.

Oligonucleotide probes and primers for detecting such microsatellite markers are preferably complementary, or mostly complementary to contiguous nucleotides which are adjacent to the actual CCG trinucleotide repeat. Typically the contiguous nucleotides are within 2 kb of the trinucleotide repeat, more typically within 1 kb, preferably within 500 bp, and more preferably within 250 bp. Additional features of the probes or primers may be present, such as linker sites comprising particular restriction endonuclease sites or other sites for specific interactions with particular proteins.

The hybridization probes of the present invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, fluorescent label, biotin-avidin label, chemiluminescence, and the like. After hybridization, the probes may be detected using known methods. The nucleic acid probes of the present invention include RNA, as well as DNA probes, suchprobes being generated using techniques known in the art.

A typical method for measuring the size of a microsatellite marker is by its electrophoretic mobility on a polyacrylamide gel. Other size measurements of polynucleotides may also be employed as are known in the art.

Samples for testing can be derived from patient samples or from immortalized cell lines. For example, blood cells can be used as a source of DNA which can be tested for the changes in the size of the trinucleotide repeat markers of the present invention. Any

cells of the human body from which nuclear DNA can be extracted can be used as the sample source. Alternatively, cells can be immortalized and then tested, such as immortalized lymphoblastoid cells.

37 genes previously entered in GenBank have eight or more consecutive CCG repeats. A large proportion of these genes (38%) encode some form of transcription factor. As a rough comparison, 3.6% (229 of 41018 entries) of the human unigene database, 2.6% (2643 of 101225) of the Entrez GenBank human entries, and 2.6% (164 of 6245 entries) of the Swissprot database human entries contain the word "transcription". Since some transcription factors are not labeled as such, these figures underestimate the true number of transcription factor entries in these databanks. Nonetheless, the implication is that genes with CCG repeats are more likely to encode transcription factors than other genes. In addition, expansion mutations in six of these 37 genes are known to cause human disease.

We have now cloned and partially characterized 18 cDNAs from human brain containing CCG repeats. 15 of these repeats have not been previously described. Our detection of a relatively large number of unreported repeats probably reflects the difficulty of sequencing ESTs that include extremely C-G rich domains, and also a bias against the presence of extreme 5' UTR sequence in known genes, particularly when such sequence is C-G rich. This suggests that as large scale sequencing of human genomic DNA continues over the next several years, specific efforts to clone C-G rich expressed sequences may remain a valuable tool in identifying transcribed C-G rich repeats.

Five of the repeats reported here have heterozygosities of at least 40%. The recent descriptions of disease-associated repeats that are only minimally polymorphic in the general population yet expand to cause disease (the repeats in SCA2, CBL2, CBFA1, HOXD13, and PABP2) demonstrate that marked polymorphism is not necessary for expansion. Nonetheless, a disproportionate number of expansion mutations arise from polymorphic repeats, and such repeats are therefore of particular interest.

Analysis of microsatellite sequences by polymerase chain reaction, for example, may follow the methods of Weber and May, Abundant Class of Human DNA

Polymorphisms Which Can Be Typed Using the Polymerase Chain Reaction, Am. I. Hum. Genet. 44:388-96 (1989). Primer pairs are selected to hybridize to the DNA flanking the interspersed repetitive sequence loci at selected chromosomal locations.

A variety of gene amplification techniques may be used for analysis of individual loci of interspersed repetitive sequences. Such methods may include, without limitation, Polymerase Chain Reaction (PCR), Saiki et al., Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for the Diagnosis of Sickle Cell Anemia, Science 230:1350-54 (1985); Ligase Chain Reaction (LCR), Wu and Wallace, The Ligation Amplification Reaction (LAR)-Amplification of Specific DNA Sequences Using Sequential Rounds of Template-Dependent Ligation, Genomics 4:560-69 (1989) and Landegren et al., A Ligase Mediated Gene Detection Technique, Science 241:1077-80 (1988); Q-beta-Replicase Template Amplification (Q-beta), Lomeli et al., Quantitative Assays Based on the Use of Replicatable Hybridization Probes, Clin. Chem. 35:1826-31 (1989); and Strand Displacement Activation (SDA), Walder et al., Isothermal in vitro Amplification of DNA by a Restriction Enzyme/DNA Polymerase System, Proc. Nat'l Acad. Sci. USA 89:392-96 (1992). RNA-based amplification methods may be used for those interspersed repetitive sequences that are expressed as RNA. An example of an RNA-based amplification method is Self-Sustained Sequence Replication (3SR), Guatelli et al., Isothermal In Vitro Amplification of Nucleic Acids by a Multi-enzyme Reaction Modified After Retroviral Replication, Proc. Nat'l Acad. Sci. USA 87:1874 (1990).

Of the eighteen cDNAs with CCG repeats that we cloned, at least eight are present in 5'-UTR, consistent with the high rate (59%) of 5'-UTR repeats in known genes with CCG repeats. The function of 5'-UTR C-G rich regions in general remains unclear, though alterations of the G-C content in the promotor regions of various expression vectors influences the expression of reporter genes (Amirhaeri et al. 1995; Krajewski, 1996). It remains unknown whether the high rate of CCG trinucleotide repeats in 5'-UTRs reflects a specific function of the repeat or is a stochastic epiphenomenon of the high G-C content of 5'-UTRs.

Four of the cDNAs are of particular interest because of their homology at the amino acid level to proteins with known neuronal functions. CCGFB60 contains the

entire human coding sequence for neurexin-1β, one of a family of brain-specific cell surface proteins that may have a role in mediating cell recognition (Missler and Sudhof, 1998). P42F11 contains part of human BCNG-1, a recently described brain-specific gene that appears to encode a new form of ion channel with pacemaker properties (Santoro et al. 1997, Santoro et al. 1998). P72D4 corresponds to the guanine nucleotide-binding protein (G-protein) β2 subunit, a ubiquitous component of signal transduction pathways (Sprang, 1997). The clone includes a CCG repeat located 5' to the sequence entry in GenBank. The repeat is polymorphic (heterozygosity of 40%), reaching a length of at least 18 consecutive triplets. CCGFB84 is the human version of the mouse proline rich protein 7, which interacts with the neuronal protein FE65 that in turn interacts with the β-amyloid precursor protein (Ermekova et al. 1997).

Finally, assignment of these repeats to chromosomal loci facilitates testing for associated expansions in diseases linked to these regions. For instance, P12A7 is near a region (18p11) linked to bipolar affective disorder (Berrettini et al. 1994), P42A5 is near a linkage site for schizophrenia (22q12) (Gill et al. 1996) and CCG98 is near a linkage site for oculodentodigitial dysplasia (6q22) (Galdwin et al. 1997).

The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific examples which are provided herein for purposes of illustration only, and are not intended to limit the scope of the invention.

EXAMPLE 1

Methods

Approximately 1 million plaques each from human frontal cortex and fetal brain cDNA phagemid libraries (Stratagene) were screened for cDNA inserts containing CCG repeats, as previously described (Margolis et al. 1997). The final wash was performed at 74-75° and filters were rescreened to exclude inserts of 28S RNA. Clones CCG43, CCG82, and CCG98 were obtained from the frontal cortex library, and clones CCGFB48, CCGFB60, CCGFB64, CCGFB84 were obtained from the fetal brain library. In addition, a human fetal brain cDNA library in pCMV-Sport (Life Technologies) was similarly screened using a modified version of the GeneTrapper™ protocol (Wang et al. 1996).

Inserts were sequenced (ABI) and compared to GenBank entries using BLASTN (Altshul et al. 1997) to exclude cDNAs derived from genes with previously characterized repeats. Double stranded sequence was obtained for at least 500 base pairs in the region of the repeat of each novel insert. Long open reading frames were examined for motifs at the amino acid level (Henikoff and Henikoff, 1994; Bairoch et al. 1997; Worley et al. 1995).

Length polymorphism was assessed in cDNAs containing at least five consecutive triplets by amplification across the repeat using a radiolabelled PCR primer as previously described (Margolis et al. 1997). The typical PCR protocol involved denaturation at 96° for 5 minutes, then 33 cycles of 95° for 1 minute, annealing (see Table 3) for 1 minute, and 72° for 1 minute, followed by a final extension of 72° for 7 minutes. Buffer J (Epicentre) improved product specificity, with the addition of 5% DMSO (CCGFB84) or 2.5% DMSO (P62D12). Template consisted of 40-80 ng of genomic DNA from a set of unrelated individuals from the Centre d'Etude du Polymorphisme Humaine (CEPH) collection (Dausset et al. 1990). Southern blots of each PCR product were probed with a radiolabelled (CGG)10 oligonucleotide to establish the presence of the repeat.

Most cDNAs were assigned to a specific locus using the Genebridge4 radiation hybrid panel (Walter et al. 1996). When possible, PCR was performed with the same primer pair used for analysis of length polymorphism. Primer pairs amplifying a region of cDNA adjacent to the repeat were used for radiation hybrid mapping of clones CCGFB48 (TGGCCTGCTGCTGGAG, ATGCCACTTGGTGCTCGTAT), CCGFB64 (CACCGGAGGCAGTGAGG, CCAGCACCAGCCAATAAAGC), P12E1 (GCGGGCAGGGTCATCAAG, TACGCGGTCGAGTCCAGGTA), P62D12 (GCACGCTGTCTCAATGTG, CATCATATTCTTGGCGATTT). Clones CCGFB60 and P32H12 were assigned to a locus by sequence identity to a mapped STS (Schuler, 1996). Clone CCGFB64 was assigned to chromosome 2 or 10 and P12E1 was assigned to chromosome 3 with the NIGMS monochromosomal human-rodent hybrid cell line panel 2 (Dubois and Naylor, 1993).

EXAMPLE 2

Search of GenBank for CCG repeats

To determine the general characteristics of known genes containing CCG repeats in the human genome, the nonredundant GenBank database was searched for perfect matches with a (CGG)8 sequence using the BLASTN algorithm (Altshul et al. 1997). Only human genes with 8 or more CGG repeats, in any orientation, were included. The resulting list (Table 1) contains 37 genes (and 3 other repeats of interest). 14 of the 37 (38%) encode some form of transcription factor.

Of the 37 genes with CCG repeats, 3 are in untranscribed regions, 22 (59%) are located in 5' untranslated regions (UTRs), while only 1 (3%) is in a 3' UTR. 11 repeats are in coding sequence; 6 encode polyalanine, 4 encode polyglycine, and 1 encodes polyproline. At least 12 of the 37 repeats are polymorphic in length, though information on many of the others is not available. Expansion mutation in the CGG repeats of 3 of the 37 are known to cause disease (FMR1, FMR2 and CBL2). CAG repeats in three other genes from this group expand to cause disease (HD, CACNA1A, and AR).

EXAMPLE 3

Screening cDNA libraries for CCG repeats

Details of the cDNAs isolated by screening cDNA libraries are described in Table 2. The number of consecutive triplets ranges from five to 13. Many repeats are flanked by regions containing exclusively C-G base pairs, and five of the cDNAs contain two adjacent regions of perfect repeats. In 10 cDNAs, the coding status of the repeat could be definitively ascertained. Eight of these repeats are in 5' UTRs. Of the remaining two, one encodes alanine and the other encodes proline. 16 of the 18 cDNAs could be assigned to a chromosome by somatic hybrid mapping or a match to an STS.

Polymorphism was generally assessed in 20 chromosomes, providing a rough estimate of the extent of heterozygosity and the range of common alleles (Table 3). Nine of the 18 repeats are polymorphic in length. The mean heterozygosity of the polymorphic repeats is 34%. Longer repeats tend to be more polymorphic: heterozygosity is 10% for loci with modal repeat lengths of 5-9 triplets (N=13) and 36%

for loci with modal repeat lengths of 10-13 triplets (N=5) ($t = .099$, $df = 16$, $p = .01$, one-tailed). Differences in allele length in all cases appear to be in units of three base pairs (Wells and Warren, 1998).

References

Albanese V, Holbert S, Saada C, Meier-Ewert S, Lebre A-S, Moriniere S, Bougueleret L, Le Gall I, Weissenbach J, Lennon G, Lehrach H, Cohen D, Cann HM, Neri C (1998) CAG/CTG and CGG/GCC repeats in human brain reference cDNA: outcome in searching for new dynamic mutations. Genom 47:414-418

Altshul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 25:3389-3402

Amirhaeri S, Wohlrab F, Wells RD (1995) Differential effects of simple repeating DNA sequences on gene expression from the SV40 early promotor. J Biol Chem 270:3313-3319

Bairoch A, Bucher P, Hofmann K (1997) The PROSITE database, its status in 1997. Nucl Acids Res 25:217-221

Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger JI J, Gershon ES (1994) Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. Proc Natl Acad Sci U S A 91:5918-5921

Brais B, Bourchard JP, Xie YG, Rochefort DL, Chretien N, Tome FM, LaFreniere RG, Rommens JM, Uyama E, Nohira O, Blumen S, Korcyn AD, Heutink P, Mathieu J, Duranceau A, Codere F, Fardeau M, Rouleau GA (1998) Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nature Genet 18:164-167

Breschel T, McInnis MG, Margolis RL, Sirugo G, Corneliuseen B, Simpson SG, McMahon FJ, MacKinnon DF, Xu JF, Pleasant N, Huo Y, Ashworth RG, Grundstrom C, Grundstrom T, Kidd KK, DePaulo JR, Ross CA (1997) A novel, heritable, expanding CTG repeat in an intron of SEF2-1 localizes to human chromosome 18q21.1. Hum Mol Genet 6:1855-1864

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, Johnson K, Harper PS, Shaw DJ, Housman DE (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

Bulle F, Chiannilkulchai N, Pawlak A, Weissenbach J, Gyapay G, Guellaen G (1997) Identification and chromosomal localization of human genes containing CAG/CTG repeats expressed in testis and brain. Genome Res 7:705-715

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Canizares J, Koutnikova H, Bidichandani S, Gellera C, Brice A, Trouillas P, DeMichele G, Filla A, DeFrutos R, Palau F, Patel PI, Pragna I, DiDonato S, Mandel J-L, Cocozza S, Koenig M, Pandolfo M (1996) Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271:1423-1427

Dausset J, Cann H, Cohen D, Lathrop m, Lalouel JM, White R (1990) Centre d'Etude du Polymorphisme humaine (CEPH): Collaborative genetic mapping of the human genome. . Genom 6:575-577

Dubois BL, Naylor SL (1993) Characterization of NIGMS human/rodent somatic cell hybrid mapping panel 2 by PCR. Genom 16:315-319

Ermekova KS, Zambrano N, Linn H, Minopoli G, Gertler F, Russo T, Sudol M (1997) The WW domain of neural protein FE65 interacts with proline-rich motifs in Mena, the mammalian homolog of Drosophila enabled. J Biol Chem 272:32869-32877

Galdwin A, Donnai D, Metcalfe K, Schrander-Stumpel C, Brueton L, Verloes A, Aylsworth A, Toriello H, Winter R, Dixon M (1997) Localization of a gene for oculodentodigital syndrome to human chromosome 6q22-q24. Hum Mol Genet 6:123-127

Gastier JM, Brody T, Pulido JC, Businga T, Sunden S, Hu X, Maitra S, Buetow KH, Murray JC, Sheffield VC, Boguski M, Duyk GM, Hudson TJ (1996) Development of a screening set for new (CAG/CTG)n dynamic mutations. Genom 32:75-85

Gill M, Vallada H, Collier D, Sham P, Holmans P, Murray R, McGuffin P, Nanko S, Owen M, Antonarakis S, Housman D, Kazazian H, Nestadt G, Pulver AE, Straub RE, MacLean CJ, Walsh D, Kendler KS, DeLisi L, Polymeropoulos M, Coon H, Byerley W, Lofthouse R, Gershon E, et al. (1996) A combined analysis of D22S278 marker alleles in affected sib-pairs: support for a susceptibility locus for schizophrenia at chromosome 22q12. Schizophrenia Collaborative Linkage Group (Chromosome 22). Am J Med Genet 67:40-45

Goodman FR, Mundlos C, Muragaki Y, Donnai D, Giovannucci-Uzielli ML, Lapi E, Majewski F, McGaughran J, McKeown C, Reardon W, Upton J, Winter RM, Olsen BR, Scambler PJ (1997) Synpolydactyly phenotypes correlate with size of expansions in HOXD13 polyalanine tract. Proc Natl Acad Sci USA 94:7458-7463

Henikoff S, Henikoff JG (1994) Protein family classification based on searching a database of blocks Genom 19:97-107

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

Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier J-M, Weber C, Mandel J-L (Group 1), Cancel G, Abbas N, Durr A, Didierjean O, Stevanin , Agid Y, Brice A (Group 2) (1996) Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. Nature Genet 14:295-296

Jiang JX, Deprez RH, Zwarthoff EC, Riegman PH (1995) Characterization of four novel CAG repeat-containing cDNAs. Genom 30:91-93

Jones C, Penny L, Mattina T, Yu S, Baker E, Voullaire L, Langdon WY, Sutherland GR, Richards RI, Tunnacliffe A (1995) Association of a chromosome deletion syndrome with a fragile site with the proto-oncogene CBL2. Nature 376:145-149

Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I, Kimura J, Narumiya S, Kakizuka A (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nature Genet 8:221-228

Knight SJL, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, Pointon J, Middletonprice HR, Barnicoat A, Pembrey ME, Holland J, Oostra BA, Bobrow M, Davies KE (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell 74:127-134

Koide O, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, Kondo R, Ishikawa A, Hayashi T, Tomoda A, Miike T, Naito H, Ikuta F, Tsuji S (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nature Genet 6:9-13

Krajewski WA (1996) Enhancement of transcription by short alternating C.G tracts incorporated within a Rous sarcoma virus-based chimeric promotor: in vivo studies. Mol Gen Genet 252:249-254

La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352:77-79

LaFreniere RG, Rochefort DL, Chretien N, Rommens JM, Cochius JI, Kalviainen R, Nousiainen U, Patry G, Farrell K, Soderfeldt B, Federico A, Hale BR, Cossio OH, Sorensen T, Pouliot MA, Kmiec T, Uldall P, Janszky J, Pranzatelli MR, Andermann F, Andermann E, Rouleau GA (1997) Unstable insertion in the 5' flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1. Nat Genet 15:298-302

Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, Malafosse A, Antonarakis SE (1997) Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy [see comments]. Nature 386:847-851

Li S-H, McInnis MG, Margolis RL, Antonarakis SE, Ross CA (1993) Novel triplet repeat containing genes in human brain: Cloning, expression, and length polymorphisms. Genom 16:572-279

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Nelville C, Narang M, Barcelo J, O'Hoy K, Leblond S, Earle-MacDonald J, DeJong PJ, Wieringa B, Korneluk RG (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science 255:1253-1255

Margolis RL, Breschel TS, Li S-H, Kidwai AS, Antonarakis SE, McInnis MG, Ross CA (1995a) Characterization of cDNA clones containing CCA trinucleotide repeats derived from human brain. Somat Cell Mol Genet 21:279-284

Margolis RL, Breschel TS, Li S-H, Kidwai AS, McInnis MG, Ross CA (1995b) Polymorphic (AAT)n trinucleotide repeats derived from a human brain cDNA library. Hum Genet 96:495-496

Margolis RL, Abraham MA, Gatchell SB, Li S-H, Kidwai AS, Breschel TS, Stine OC, Callahan C, McInnis MG, Ross CA (1997) cDNAs with long CAG trinucleotide repeats from human brain. Hum Genet 100:114-122

McInnis MG (1996) Anticipation: an old idea in new genes. Am J Hum Genet 59:973-979

McInnisã MG, Margolis RL (1998) Anticipation and psychiatric disease. In: Wells RD, Warren ST (eds) Genetic Instabilities and Hereditary Neurological Diseases. Academic Press, Inc., San Diego pp. 401-411

Missler M, Sudhof TC (1998) Neurexins: three genes and 1001 products. TIG 14:20-26

Mortlock DP, Post LC, Innis JW (1996) The molecular basis of hypodactyly (Hd): a deletion in Hoxa 13 leads to arrest of digital arch formation. Nature Genet 13:284-289

Mundlos S, Otto F, Mundlos C, Mulliken JB, Aylsworth AS, Albright S, Lindhout D, Cole WG, Henn W, Knoll JHM, Owen MJ, Mertelsmann R, Zabel BU, Olsen BR (1997) Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. Cell 89:773-779

Muragaki Y, Mundlos S, Upton J, Olsen BR (1996) Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. Science 272:548-551

Nagafuchi S, Yanagisawa H, Soto K, Shirayama T, Ohsaki E, Bundo M, Takeda T, Tadokoro K, Kondo I, Murayama N, Tanaka Y, Kikushima H, Umino K, Kurosawa H, Furukawa T, Nihei K, Inoue T, Sano A, Komure O, Takahashi M, Yoshizawa T, Kanazawa I, Yamada M (1994) Dentatorubral and pallidoluysian atrophy is caused by an expansion of an unstable CAG trinucleotide repeat on chromosome 12p. Nature Genet 6:14-18

Neri C, Albanese V, Lebre AS, Holbert S, Saada C, Bougueleret 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, Cann HM (1996) Survey of CAG/CTG repeats in human cDNAs representing new genes: candidates for inherited neurological disorders. Hum Mol Genet 5:1001-1009

Reddy PH, Stockburger E, Gillevet P, Tagle DA (1997) Mapping and characterization of novel (CAG)n repeat cDNAs from

adult human brain derived by the oligo capture method. Genom 46:174-182

Riggins GJ, Lokey LK, Chastain JL, Leiner HA, Sherman SL, Wilkinson KD, Warren ST (1992) Human genes containing polymorphic trinucleotide repeats. Nat Genet 2:186-190

Ross CA, McInnis MG, Margolis RL, Li S-H (1993) Genes with triplet repeats: Candidate mediators of neuropsychiatric disorders. Trends Neurosci 16:254-260

Ross CA (1997) Intranuclear neuronal inclusions: A common pathogenic mechanism for glutamine-repeat neurodegenerative diseases? Neur0n 19:1147-1150

Ross CA, Margolis RL, Becher MW, Wood JD, Engelender S, Sharp AH (1998) Pathogenesis of polyglutamine neurodegenerative diseases: towards a unifying mechanism. In: Wells RD, Warren ST (eds) Genetic instabilities and hereditary neurological diseases. Academic Press, Inc., San Diego:

Santoro B, Grant SG, Bartsch D, Kandel ER (1997) Interactive cloning with the SH3 domain of N-src identifies a new brain specific ion channel protein, with homology to eag and cyclic nucleotide-gated channels. Proc Natl Acad Sci USA 94:14815-14820

Santoro B, Liu DT, Yao H, Bartsch D, Kandel ER, Siegelbaum SA, Tibbs GR (1998) Identification of a gene encoding a hyperpolarization-activated pacemaker channel of brain. Cell 93:717-729

Schuler GD (1996) A gene map of the human genome. Science 274:540-546

Sprang SR (1997) G proteins, effectors and GAPs: structure and mechanism. Curr Opin Struct Bio 7:849-856

Walter M, Spillett D, Thomas P, Weissenbach J, Goodfellow P (1996) A method for constructing radiation hybrid maps of whole genomes. Nature Genet 7:22-28

Wang Q, Curran ME, Splawski I, Burn TC, Millholland JM, VanRaay TJ, Shen J, Timothy KW, Vincent GM, de Jager T, Schwartz PJ, Toubin JA, Moss AJ, Atkinson DL, Landes GM, Connors TD, Keating MT (1996) Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. Nat Genet 12:17-23

Wells RD, Warren ST (1998) Genetic Instabilities and Hereditary Neurological Diseases. Academic Press, Inc., San Diego

Worley KC, Wiese BA, Smith RF (1995) BEAUTY: An enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. Gen Res 5:173-184

Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, Baker E, Holman K, Mulley JC, Warren ST, Schlessinger D, Sutherland GR, Richards RI (1992) Fragile X genotype characterized by an unstable region of DNA. Science 252:1179-1981

Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, Dobyns WB, Subramony SH, Zoghbi HY (1997) Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha$_{1A}$-voltage-dependent calcium channel. Nature Genet 15:62-69

Table 1. Human genes with eight or more consecutive CCG triplets.

Gene	Longest Repeat *=polymorphic	Repeat encoding	Locus
CACNA1A (a1A voltage dependent calcium channel)	(GCG)9	5' UTR	19p13
ALK-5 (TGF beta receptor type 1)	(GCG)9	Ala	9q33-q34
AR (androgen receptor)	(GGC)18*	Gly	Xq11.2-Xq12
AE3 (anion exchanger 3)	(CGG)10	intronic	2q32-33
ATBF1 (α-fetoprotein enhancer-binding protein)	(CCG)10	5'UTR	16q22.3-23.1
BARK1 (β-adrenergic receptor kinase type 1)	(CGG)8	5' UTR	11q13
Bleo (bleomycin hydrolase)	(CGG)10*	5' flanking	17q11.2
Brain-1 (a class III POU gene)	(CGG)9	3' UTR	---
Brn-3b (a POU domain-containing gene)	(GGC)12	Gly	4q31.2
c-Ha-ras-1 oncogene	(CGG)8	5' flanking	11p15
Calpain, small subunit (calcium-dependent protease)	(GCC)10	Gly	19
CBL2 protooncogene	(CGG)11*	5'UTR	11q23.3
Finb (zinc finger transcriptional activator)	(GCC)9	5'UTR	---
FKhL15 (forkhead family DNA binding protein)	(GCC)9	Ala	9q22
FMR1 (fragile X mental retardation 1)	(CGG)10*	5' UTR	Xq27.3
FMR2 (fragile X mental retardation 2)	(CCG)15*	5' UTR	Xq27.3-q28
FREAC-1 (forkhead-related activator protein)	(CGG)8	5' UTR	16q24

GST1-Hs (GTP-binding protein)	(CGG)10*	5' UTR	16p13.1
β-HLH DNA binding protein (H-twist)	(GGC)9	Gly	7p21
HAUSP (herpesvirus assoc ubiquitin-specific protease)	(CCG)8	5' UTR	---
HB-9 (homoeobox gene)	(GCC)11	Ala	1q41-q42.1
HFKH4 (Forkhead like 4)	(GCC)9	Ala	---
HHR6B (ubiquitin conjug. enzyme)	(CGG)11	5' UTR	5q23-q31
HD (huntington)	(CCG)8*	Pro	4p16.3
HXC-26	(CCG)9	5' UTR	Xq28
IRS-2 (insulin receptor substrate-2)	(GCC)8	Ala	---
KIAA0359 (brain-derived)	(CGG)12	5' UTR	---
LTG-19 (chromosomal translocation associated gene)	(CGG)8	5' UTR	19p13
MRP (multidrug resistance-associated protein)	(GCC)14	5' UTR	16p13.1
Na,K-ATPase β subunit	(CCG)8*	5' UTR	1q23-q25
NCAD (N-cadherin)	(CCG)8*	5' UTR	18q11.2
PILOT (transcription factor)	(CGG)10	5' UTR	8p21-23
RELN (reelin)	(CGG)11	5' UTR	7q22
TrkC (tyrosine receptor kinase)	(CGG)8*	5' UTR	15q24-q25
VLDL (very low density lipid) receptor	(CGG)8*	5' UTR	9p24
XAP-5	(CCG)9*	5' UTR	Xq28
ZIC3 (zinc-finger cerebellum 3 transcription factor)	(GCC)8	Ala	Xq26.2

Other CCG repeats of interest:			
FRAXF	(GCC)8*	N/A	Xq28
FRA16A	(CCG)7*	N/A	16p13.11
28s rRNA	(CCG)8	N/A	p12 on chrom 13, 14, 15, 21, 22

Table 2. cDNAs with CCG repeats: locus and homologies to known genes

Clone (GenBank #)	Longest consecutive repeat	Repeat translation	Radiation hybrid results	Chromosomal locus	Homology
CCG43 (AF06777)	9	?	83 cR from GATA85A06	**2p11.2**	repeat region and part of 5' flanking region similar to brain-1 (POU3F3, G17183)
CCG82 (AF064839)	6-4	?	3.15 cR from WI-9324	**3p21**	STS G24812
CCG98 (AF064840)	9	Ala	2.74 cR from D6S266	**6q22**	none

CCGFB48 (AF064841)	13	5' UTR or Arg	44 cR from WI-8572	**13p13**	human tolloid-like protein (35% identity/50 % conserved over 110 aa, U91963), also motifs similar to low density lipoprotein receptor and membrane attack complex components
CCGFB60 (AF064842)	3-6	5' UTR	202 .6 cR from top of chromosome	**2q30-32**	cow neurexin 1-B (L27870), STS WI-16372 (G23715)
CCGFB64 (AF064843)	10	Gly or Arg	----	**2 or 10**	none
CCGFB84 (AF064844)	7-4	Pro	----	-----	mouse proline rich protein 7 (AF020311)

P12A7 (AF064846)	10	Gly?	7 cR from WI-5607	**18p11**	decorin (U03394) 47% amino acid identity over 75 amino acids)
P12E1 (AF064845)	3-4-5	5' UTR	----	**3**	zinc finger protein 85 (1731445) 50% identity over 171 amino acids, multiple zinc finger motifs
P32B10 (AF064847)	7	5'UTR	51cR from D20S173	**20q13.3**	HS1 protein (X57346) (does not contain repeat, unmapped)
P32D9 (AF064848)	8	?	3.87 cR from WI-7121	**20q12**	EST H23496 (does not contain repeat)

P32H12	5	5' UTR	----	**4q35**	β-tubulin (P05217) (no polymorphism analysis)
P42A5 (AF064850)	5-10	?	4.71 cR from WI-268	**22q12**	chromosome 22 PAC (Z83733)
P42F11 (AF064851)	6	5' UTR	18.87 cR from 1B1264	**19p13.3**	BCNG-1 (AF064876) (does not contain repeat)
P55G12 (AF064852)	6	5' UTR	11.65 cR from WI-611	**1p34**	EST M62228 (does not contain repeat)
P62D12 (pending)	9	?	9.65 cR from GATA 87B02	**3p22**	EST AA280956 (does not contain repeat)

P72D4 (AF064853)	10	5' UTR	41.65 cR from WI-8540	**7qter**	Guanine nucleotide binding protein β-2 unit (M16514, does not contain repeat)
P95B10 (AF064854)	7	5' UTR	5.13 cR from GATA 41C05	**17q24**	61% identity to carbonic anhydrase-related protein XI (AF050106)

Table 3. Repeat length polymorphism. Repeat length indicates greatest number of consecutive perfect repeats. Hz = heterozygosity.

Clone	Repeat length	PCR primers	Annealing temp °C	Product length (bp)	# chrom tested	Allele length: frequency	Hz
CCG43	9	TCCGCAGCCGCCGTCAGCAC TCTCTGGGTGGAGGGGAGGG	65	145	20	145: 1.00	0%
CCG82	5	CAACGGCATGGAACAGCG GGAGCGGGTCACTTGGTCG	62	100	20	100: 1.00	0%
CCG98	9	CGTGAATGAGAGCAAACC CGCCCTAAACTCCACTACTT	59	160	20	160: 1.00	0%
CCGFB48	13	CCGGTGGCTCGGCGG GGAGCTGGAGGTAGACGACGA	60	68	20	77: .15 68: .75 65: .10	40%
CCGFB60	10	TGCCGGTGACCTGTAGATT CCTGGCCCTGCTTTGGA	56	155	20	155: 1.00	0%
CCGFB64	10	CTCGCGCTCTGCCTCCCTC ACTGCCTCCGGTGGATGATG	60	268	30	280: .05 277: .10 271: .15 268: .45 265: .15 262: .10	60%
CCGFB84	7-4	CGA6GAGGAGGACGACGAC TTCTTCACACCCAATGCTGA	50	190	20	190: 1.00	0%
P12A7	10	TCGCGCTGCTGGGAGGCT GGCCGCCCTCGTGCTTG	65	91	20	91: .55 82: .45	30%

Table 3. Repeat length polymorphism. Repeat length indicates greatest number of consecutive perfect repeats. Hz = heterozygosity.

P12E1	3-4-5	CGAGGTGCCGTCAAAGACAC TGCAGCACACAGCCACAGT	60	164	20	164: 1.00	0%
P32B10	7	CCACCGCCGCCGCCGATT CTTCCTTCCTAAGCCCACTACTC A	62	110	20	119: .25 110: .15 107: .60	40%
P32D9	8	ATCCCCCACCCCCGCACC GGGCGCGAGATGGGCTGC	65	149	40	149: 1.00	0%
P32H12	5	GCCCGCTCTTCTGCTGCTGT CCGATTTGGTTGCCGCACTG	60	114	20	114: 1.00	0%
P42A5	10	CCCCGCCACCAACAACAATAAC AGTATTCTGTAAGCCCTTGAGC	59	148	40	148: .20 145: .10 142: .70	50%
P42F11	6	CATGGCCGCCGCCTGC CGTCAAGTACGACCGCGAGAT	65	116	20	116: .95 110: .05	10%
P55G12	6	CCCGCCTCCGAGTCGCTACTT GCTGCGGCGTCACTGGGGGTCTC	65	130	20	130: 1.00	0%
P62D12	9	CCAACCGCCGCCTCA TTGAGACAGCGTGCTCAGCAT	58	72	20	81: .05 72: .95	10%
P72D4	10	GGGGCCACTGAGGAAATCCAT CGGGCGGCGCTGTCTTCC	65	86	20	110: .05 95: .10 92: .65 86: .20	40%
P95B10	7	AATCGGGGTCTCGTTTTTG CGCTCGCGTGTCTCTTCTCT	56	127	38	133: .80 130: .05 127: .15	25%

CLAIMS

1. A polynucleotide for detecting a microsatellite marker selected from the group consisting of: P12A7, P12E1, P32B10, P32D9, P32H12, P42A5, P42F11, P55G12, P62D12, P72D4, P95B10, CCG43, CCG82, CCG98, CCGFB48, CCGFB60, CCGFB64, and CCGFB84, wherein the polynucleotide comprises at least 12 nucleotides complementary to contiguous nucleotides within 500 nucleotides of a trinucleotide repeat in the microsatellite marker in the human genome.

2. A method of determining a change in number of trinucleotide repeats in a microsatellite marker, comprising:

hybridizing a polynucleotide according to claim 1 to a nucleic acid sample of a patient to form a hybridized polynucleotide;

determining size of the hybridized polynucleotide, wherein an increase in the size of the hybridized polynucleotide relative to size of the polynucleotide hybridized to a nucleic acid sample of a normal human indicates a change in the number of trinucleotide repeats.

3. The method of claim 2 wherein the nucleic acid sample of the patient comprises fragments of known sizes.

4. The method of claim 2 wherein the nucleic acid sample of the patient comprises restiction enzyme digested DNA.

5. The method of claim 2 wherein the patient is suspected of having a neurological disorder.

6. A pair of primers for amplifying a microsatellite marker selected from the group consisting of P12A7, P12E1, P32B10, P32D9, P32H12, P42A5, P42F11, P55G12, P62D12, P72D4, P95B10, CCG43, CCG82, CCG98, CCGFB48, CCGFB60, CCGFB64, and CCGFB84, wherein each primer is complementary to at least 12 consecutive nucleotides which are within 500 nucleotides of a trinucleotide repeat in the microsatellite marker in the human genome, and wherein each primer of the pair is complementary to opposite strands of the microsatellite marker.

7. The pair of primers of claim 6 which is selected from primers 1-36, (SEQ ID NOS: 19-54).

8. A method for determining a change in number of trinucleotide repeats in a microsatellite marker comprising:

amplifying a microsatellite marker using a pair of primers according to claim 6 and a template comprising a nucleic acid sample of a patient;

determining size of the microsatellite marker amplified, wherein an increase in size of the amplified microsatellite marker relative to the size of a microsatellite marker amplified using the pair of primers and a template comprising a nucleic acid sample of a normal human indicates a change in the number of trinucleotide repeats.

9. The method of claim 8 wherein the patient is suspected of having a neurological disorder.

10. The method of claim 2 or 8 wherein the microsatellite marker is P12A7 and the patient is suspected of having bipolar affective disorder.

11. The method of claim 2 or 8 wherein the microsatellite marker is P42A5 and the the patient is suspected of having schizophrenia.

12. The method of claim 2 or 8 wherein the microsatellite marker is CCG98 and the the patient is suspected of having oculodentodigitial dysplasia.