J Hum Genet (2002) 47:275–278

ORIGINAL ARTICLE

**Masaji Tachikawa · Yoshitaka Nagai · Koichiro Nakamura
Kazuhiro Kobayashi · Tsutomu Fujiwara · Hye-Jung Han
Yuko Nakabayashi · Yaeko Ichikawa · Jun Goto
Ichiro Kanazawa · Yusuke Nakamura · Tatsushi Toda**

Identification of CAG repeat-containing genes expressed in human brain as candidate genes for autosomal dominant spinocerebellar ataxias and other neurodegenerative diseases

Received: January 15, 2002 / Accepted: March 4, 2002

Abstract To obtain novel candidate genes for autosomal dominant spinocerebellar ataxia and other neurodegenerative disorders in which gene mutations remain unidentified, we screened a human fetal brain cDNA library using $(CAG)_{10}$ repeat probes. Sixteen cDNAs were isolated and mapped to chromosomes 1, 2, 3, 6, 9, 13, 15, 16, 22, and X. Although we failed to detect abnormal CAG repeat expansion within these genes in Japanese patients with inherited neurodegenerative diseases, these genes remain potential candidate genes for neurodegenerative diseases that feature anticipation.

Key words CAG repeat · Neurodegenerative diseases · Spinocerebellar ataxia · Polyglutamine · Anticipation

Introduction

Recent advances in molecular genetics have revealed a novel class of genetic mutations, called trinucleotide repeat expansion, which are commonly observed in at least 15 inherited neuromuscular disorders (Margolis et al. 1999). Among these, expanded CAG repeats are associated with inherited neurodegenerative diseases, including spinal and bulbar muscular atrophy (SBMA) (La Spada et al. 1991), Huntington's disease (HD) (The Huntington's Disease Collaborative Research Group, 1993), spinocerebellar ataxia types 1, 2, 3, 6, 7, 12, and 17 (SCA1, 2, 3, 6, 7, 12, 17) (Orr et al. 1993; Kawaguchi et al. 1994; Pulst et al. 1996; Sanpei et al. 1996; Imbert et al. 1996; Zhuchenko et al. 1997; David et al. 1997; Holmes et al. 1999; Nakamura et al. 2001), and dentatorubral and pallidoluysian atrophy (DRPLA) (Koide et al. 1994; Nagafuchi et al. 1994). With the exception of SCA12, each is caused by a CAG repeat expansion within an open reading frame, resulting in an expanded glutamine repeat (Zoghbi and Orr 2000). CAG repeat expansion and concomitant polyglutamine expression have been linked to a variety of neurodegenerative conditions. Common features of these disorders include autosomal dominant inheritance (except for SBMA), varying clinical presentation correlated with the length of the mutated trinucleotide repeats, and phenotypic anticipation caused by the intergenerational instability of the trinucleotide repeats.

Several reports, however, have indicated that approximately one quarter to one third of Japanese patients with autosomal dominant spinocerebellar ataxia have no mutational expansion of the known CAG repeats (Matsuyama et al. 1997; Takano et al. 1998; Watanabe et al. 1998; Sasaki et al. 2000). This raises the possibility that novel trinucleotide expansion could be responsible for autosomal dominant spinocerebellar ataxias and other inherited neurodegenerative diseases for which gene mutations remain unidentified. To identify candidate genes for these diseases, we screened a human brain cDNA library for cDNA clones containing CAG repeat sequences and examined the size of the repeats in Japanese patients with autosomal dominant neurodegenerative diseases.

Patients and methods

Screening, sequencing, and mapping

A human fetal brain cDNA library (Stratagene, La Jolla CA, USA) was screened with a radiolabeled $(CAG)_{10}$ oligonucleotide using standard techniques. Approximately

M. Tachikawa · Y. Nagai · K. Kobayashi · Y. Nakabayashi · T. Toda (✉)
Division of Functional Genomics, Department of Post-Genomics and Diseases, Osaka University Graduate School of Medicine, 2-2-B9 Yamadaoka, Suita, Osaka 565-0871, Japan
Tel. +81-6-6879-3380; Fax +81-6-6879-3389
e-mail: toda@clgene.med.osaka-u.ac.jp

K. Nakamura · Y. Ichikawa · J. Goto · I. Kanazawa
Department of Neurology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

T. Fujiwara
Otsuka GEN Research Institute, Otsuka Pharmaceutical Co., Ltd., Tokushima, Japan

H.-J. Han · Y. Nakamura
Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan

1,000,000 plaques were screened. Filters were hybridized at 50°C in 0.9 M sodium chloride and 90 mM sodium citrate (6 × SSC) and washed at 65°C in 6 × SSC/0.1% sodium dodecyl sulfate. After plaque purification, plasmids were excised using an in vivo excision procedure. Inserts were sequenced using the PRISM ReadyReaction BigDye Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) on an ABI Model 377 DNA sequencer. Genes were assigned to a chromosome or specific locus using the following databases: NCBI LocusLink (http://www.ncbi.nlm.nih.gov/LocusLink); GenomeNet Homo sapiens Genes (http://www.genome.ad.jp/dbget-bin/get_htext?H.sapiens); HGREP (http://hgrep.ims.u-tokyo.ac.jp/); and Celera (http://publication.celera.com/).

Patients, PCR, and polymorphism analysis

Two hundred and seventy-eight genetically undiagnosed patients whose clinical phenotypes included hereditary spinocerebellar degeneration, sporadic spinocerebellar degeneration, multiple system atrophy, familial dementia, dystonia/chorea, or spastic paraplegia were selected from the DNA collection maintained by the Department of Neurology, University of Tokyo. Polymerase chain reaction (PCR) was performed using primers flanking the CAG regions to analyze the length of the CAG repeats. The variability of repeat length was determined by size separation of the radiolabeled PCR products on 6% denaturing polyacrylamide gels.

Results

Through a screen for CAG repeat-containing cDNAs, we isolated 62 clones and identified 16 genes containing CAG repeats, as shown in Table 1. Fourteen of the repeats occur in open reading frames, with eight of the repeats coding for polyglutamine, three for polyalanine, and three for polyserine. The remaining two repeats occur in noncoding regions. Among these clones were the androgen receptor (*AR*) gene and the TATA-binding protein (*TBP*) gene, which are known to be responsible for SBMA (La Spada et al. 1991) and SCA17 (Nakamura et al. 2001). For the remaining CAG repeat-containing genes, we examined genomic DNA from 278 patients with neurodegenerative diseases to identify potential CAG repeat expansion. Of these 14 genes, we found polymorphisms in CAG repeat length only in the ribosomal protein L14 (*RPL14*) gene (Tanaka et al. 1998), the EST corresponding to accession no. U80740 (Margolis et al. 1997), and the dachsund (*DACH*) gene, which encodes a putative transcription factor (Kozmik et al. 1999) (Table 2). Except for the three cases, there were no sequence polymorphisms in the isolated clones. The CAG repeat length polymorphisms in the *RPL14* gene and the EST U80740 have been reported previously (Aoki et al. 1996; Margolis et al. 1997), and the distribution of CAG repeat lengths in our study was within the range of those previously reported. We focused on characterizing those in the *DACH* gene, comparing the allele frequency between patients and controls. The frequency of

Table 1. CAG repeat-containing genes isolated in this study

Gene	Accession no.	Number of clones isolated	Repeat sequence	Amino acid	Locus
TBP	M55654	11	$(CAG)_3CAACAACAA(CAG)_9CAA$-$CAGCAA(CAG)_{18}CAACAG$	Q_{38}	6q27
AR	NM–000044	2	$(CAG)_{20}CAA$	Q_{21}	Xq11.2–12
THRAP	AF117755	4	$(CAG)_5CAA(CAG)_3CAACAGCAA(CAG)_5CAACAG$-$CAA(CAG)_7$	Q_{27}	Xq13
EST	AL137275	1	$(CAG)_8CAACAGCAA(CAG)_4CAACAGCAA(CAG)_7$	Q_{25}	22q11.22–12.3
SATB1	M97287	4	$CAGCAA(CAG)_5CAACAGCAA(CAG)_7$	Q_{15}	3p23
EST	AB014542	2	$(CAG)_6CAA(CAG)_2$—93 base inserted—$(CAG)_6CAACACCAG$	Q_9—Q_7HQ	2pter
TNRC9	U80736	4	$(CAG)_6CAACAACAACAGATGCAACAG$-$ATG(CAG)_3CAA$	$Q_9MQ_2MQ_4$	16q12
vitiligo-associated protein VIT-1	AF176706	1	$(CAG)_7$	Q_7	2p16
EST	U80740	1	$(GCA)_{17}$	A_{17}[a]	Xp11.4
RPL14	D87735	8	$(GCA)_{13}$	A_{13}[a]	3q26.2–27
SRP14	X73459	15	$(GCA)_8CCTGCC(GCA)_4ACA(GCA)_2$	$A_8PA_4TA_2$	15q22
MLLT3	L13744	4	$(AGC)_7AGT(AGC)_{10}AGT(AGC)_2AAC(AGC)_3$-$AAC(AGC)_3AGT(AGC)_3AGT(AGC)_6(AGT)_2$	$S_{21}NS_3NS_{16}$	9p22
DACH	AF102546	2	$(AGC)_5AGT(AGC)_7AGTAGT(AGC)_4AGT(AGC)_4$	S_{24}[a]	13q21
BNAP protein	E14365	1	$AGCAGTAGCTCT(AGC)_3ACTAGTGAC(AGC)_6$	S_7TSDS_6	Xq21.3–22
RNA-binding protein	AF093097	1	$(CAG)_7$	3′ noncode	1p34–35
CGI-19 protein	AL355815	1	$(CAG)_7$	3′ noncode	6pter-21.1

TBP, TATA-binding protein; AR, androgen receptor; THRAP, thyroid hormone receptor-associated protein; SATB1, special AT-rich sequence-binding protein 1; TNRC9, trinucleotide repeat containing 9; RPL14, ribosomal protein L14; SRP14, signal recognition particle 14 kD; MLLT3, myeloid/lymphoid or mixed-lineage leukemia human homolog of trithorax (Drosophila); DACH, human homolog of dachshund (Drosophila); BNAP, Brain nucleotide assembly protein; CGI-19, comparative gene identification-19

[a] Polymorphic allele

Table 2. Allele frequency of polymorphic genes

Gene and accession no.	Repeat sequence	Amino acid	Allele frequency This report	Previous reports	
DACH	$(AGC)_5AGT(AGC)_9AGTAGT(AGC)_4AGT(AGC)_4$	S_{26}	0.01		
(AF102546)	$(AGC)_5AGT(AGC)_7AGTAGT(AGC)_4AGT(AGC)_4$	S_{24}	0.99		
EST	$(GCA)_{18}$	A_{18}	0.06		0.08
(U80740)	$(GCA)_{17}$	A_{17}	0.76		0.69
	$(GCA)_{16}$	A_{16}	0.18		—
	$(GCA)_{14}$	A_{14}	0.01		0.23[a]
RPL14	$(GCA)_{18}$	A_{18}	—	0.01	—
(D87735)	$(GCA)_{15}$	A_{15}	0.06	0.01	0.38
	$(GCA)_{14}$	A_{14}	—	0.03	0.09
	$(GCA)_{13}$	A_{13}	0.58	0.09	0.16
	$(GCA)_{12}$	A_{12}	0.36	0.02	0.16
	$(GCA)_{11}$	A_{11}	—	0.16	—
	$(GCA)_{10}$	A_{10}	—	0.06	0.19
	$(GCA)_9$	A_9	—	0.21	0.03
	$(GCA)_8$	A_8	—	0.12	—
	$(GCA)_7$	A_7	—	0.20	—
	$(GCA)_6$	A_6	—	0.04	—
	$(GCA)_5$	A_5	—	0.07[b]	—[c]

[a] Margolis et al. 1997. Ethnic origin was unknown
[b,c] Aoki et al. 1996. Caucasians (b) and African–Americans (c)

the dachshund serine 26 allele was 0.94% in our neurodegenerative disease patients ($n = 212$) and 0.22% in controls ($n = 1348$), giving no significant differences ($\chi^2 = 1.54$, $P = 0.214$). Therefore, we were unable to detect abnormal CAG repeat expansion within these 16 genes in our series of Japanese patients.

Discussion

We report here the isolation of 16 genes containing CAG repeats through the screening of a human fetal brain cDNA library. Most of the repeats were located within open reading frames, although two were found in noncoding regions. Although most CAG repeats responsible for inherited neurodegenerative diseases occur within the coding region and encode polyglutamine, CAG repeats in noncoding regions also can cause neurodegenerative diseases, including SCA8 and SCA12. In addition, abnormal expansion of GCG repeats encoding polyalanine is known to be responsible for oculopharyngeal muscular dystrophy and synpolyductyly (Muragaki et al. 1996; Brais et al. 1998). The CAG repeat-containing genes identified in this study may thus be candidate genes for neurodegenerative diseases without known mutations.

For autosomal dominant spinocerebellar ataxias, several reports have shown considerable differences in mutation frequency among different ethnic origins (Leggo et al. 1997; Moseley et al. 1998; Takano et al. 1998; Tang et al. 2000). For example, although the SCA3 mutation is the most frequent in the majority of ethnic origins studied, DRPLA and SCA6 are much more rare in Caucasians than in Japanese individuals (Takano et al. 1998). It has also been reported that the prevalence of Huntington's disease is markedly different between the Japanese and Caucasian populations (Kremer et al. 1994). Although the genes identified in this study showed no mutational expansion of CAG repeats within our sample of Japanese patients, they remain potential candidate genes for neurodegenerative diseases in other ethnic groups.

Acknowledgments We thank Ms. Helena A. Popiel for editing the manuscript.

References

Aoki M, Koranyi L, Riggs AC, Wasson J, Chiu KC, Vaxillaire M, Froguel P, Gough S, Liu L, Donis-Keller H, Permutt MA (1996) Identification of trinucleotide repeat-containing genes in human pancreatic islets. Diabetes 45:157–164

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

David G, Abbas N, Stevanin G, Durr A, Yvert G, Cancel G, Weber C, Imbert G, Saudou F, Antoniou E, Drabkin H, Gemmill R, Giunti P, Benomar A, Wood N, Ruberg M, Agid Y, Mandel JL, Brice A (1997) Cloning of the *SCA7* gene reveals a highly unstable CAG repeat expansion. Nat Genet 17:65–70

Holmes SE, O'Hearn EE, McInnis MG, Gorelick-Feldman DA, Kleiderlein JJ, Callahan C, Kwak NG, Ingersoll-Ashworth RG, Sherr M, Sumner AJ, Sharp AH, Ananth U, Seltzer WK, Boss MA, Vieria-Saecker AM, Epplen JT, Riess O, Ross CA, Margolis RL (1999) Expansion of a novel CAG trinucleotide repeat in the 5′ region of PPP2R2B is associated with SCA12. Nat Genet 23:391–392

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

Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier JM, Weber C, Mandel JL, Cancel G, Abbas N, Durr A, Didierjean O, Stevanin G, Agid Y, Brice A (1996) Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. Nat Genet 14:285–291

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. Nat Genet 8:221–228

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, Tsuji S (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 6:9–13

Kozmik Z, Pfeffer P, Kralova J, Paces J, Paces V, Kalousova A, Cvekl A (1999) Molecular cloning and expression of the human and mouse homologues of the *Drosophila dachshund* gene. Dev Genes Evol 209:537–545

Kremer B, Goldberg P, Andrew SE, Theilmann J, Telenius H, Zeisler J, Squitieri F, Lin B, Bassett A, Almqvist E, Bird TD, Hayden MR (1994) A worldwide study of the Huntington's disease mutation. The sensitivity and specificity of measuring CAG repeats. N Engl J Med 330:1401–1406

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

Leggo J, Dalton A, Morrison PJ, Dodge A, Connarty M, Kotze MJ, Rubinsztein DC (1997) Analysis of spinocerebellar ataxia types 1, 2, 3, and 6, dentatorubral-pallidoluysian atrophy, and Friedreich's ataxia genes in spinocerebellar ataxia patients in the UK. J Med Genet 34:982–985

Margolis RL, Abraham MR, Gatchell SB, Li SH, 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

Margolis RL, McInnis MG, Rosenblatt A, Ross CA (1999) Trinucleotide repeat expansion and neuropsychiatric disease. Arch Gen Psychiatry 56:1019–1031

Matsuyama Z, Kawakami H, Maruyama H, Izumi Y, Komure O, Udaka F, Kameyama M, Nishio T, Kuroda Y, Nishimura M, Nakamura S (1997) Molecular features of the CAG repeats of spinocerebellar ataxia 6 (SCA6). Hum Mol Genet 6:1283–1287

Moseley ML, Benzow KA, Schut LJ, Bird TD, Gomez CM, Barkhaus PE, Blindauer KA, Labuda M, Pandolfo M, Koob MD, Ranum LP (1998) Incidence of dominant spinocerebellar and Friedreich triplet repeats among 361 ataxia families. Neurology 51:1666–1671

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, Sato 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 expansion of an unstable CAG trinucleotide on chromosome 12p. Nat Genet 6:14–18

Nakamura K, Jeong S-Y, Uchihara T, Anno M, Nagashima K, Nagashima T, Ikeda S, Tsuji S, Kanazawa I (2001) SCA17, a novel autosomal dominant cerebellar ataxia caused by the expanded polyglutamine in TATA-binding protein. Hum Mol Genet 10:1441–1448

Orr HT, Chung MY, Banfi S, Kwiatkowski TJ, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LP, Zoghbi HY (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet 4:221–226

Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunkes A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet 14:269–276

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, Tsuji S (1996) Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nat Genet 14:277–284

Sasaki H, Yabe I, Yamashita I, Tashiro K (2000) Prevalence of triplet repeat expansion in ataxia patients from Hokkaido, the northernmost island of Japan. J Neurol Sci 175:45–51

Takano H, Cancel G, Ikeuchi T, Lorenzetti D, Mawad R, Stevanin G, Didierjean O, Durr A, Oyake M, Shimohata T, Sasaki R, Koide R, Igarashi S, Hayashi S, Takiyama Y, Nishizawa M, Tanaka H, Zoghbi H, Brice A, Tsuji S (1998) Close associations between prevalences of dominantly inherited spinocerebellar ataxias with CAG-repeat expansions and frequencies of large normal CAG alleles in Japanese and Caucasian populations. Am J Hum Genet 63:1060–1066

Tanaka M, Tanaka T, Harata M, Suzuki T, Mitsui Y (1998) Triplet repeat-containing ribosomal protein L14 gene in immortalized human endothelial cell line (t-HUE4). Biochem Biophys Res Commun 243:531–537

Tang B, Liu C, Shen L, Dai H, Pan Q, Jing L, Ouyang S, Xia J (2000) Frequency of SCA1, SCA2, SCA3/MJD, SCA6, SCA7, and DRPLA CAG trinucleotide repeat expansion in patients with hereditary spinocerebellar ataxia from Chinese kindreds. Arch Neurol 57:540–544

Watanabe H, Tanaka F, Matsumoto M, Doyu M, Ando T, Mitsuma T, Sobue G (1998) Frequency analysis of autosomal dominant cerebellar ataxias in Japanese patients and clinical characterization of spinocerebellar ataxia type 6. Clin Genet 53:13–19

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

Zoghbi HY, Orr HT (2000) Glutamine repeats and neurodegeneration. Annu Rev Neurosci 23:217–247