Hum Genet (2002) 110:302–313
DOI 10.1007/s00439-002-0698-2

ORIGINAL INVESTIGATION

Takeya Okada · Yoichi Gondo · Jun Goto
Ichiro Kanazawa · Shinji Hadano · Joh-E Ikeda

Unstable transmission of the RS447 human megasatellite tandem repetitive sequence that contains the *USP17* deubiquitinating enzyme gene

Received: 22 October 2001 / Accepted: 28 January 2002 / Published online: 16 March 2002

Abstract The RS447 megasatellite DNA, which maps to human chromosome 4p16.1, is a highly polymorphic conserved tandem repetitive sequence containing a functional deubiquitinating enzyme gene, *USP17*. To characterize the hypervariability seen in RS447 fully, we have conducted a pedigree analysis of RS447 transmission by high-resolution pulsed-field gel electrophoresis. We have identified 44 distinct alleles in 74 unrelated chromosomes containing 20–103 copies of the 4.7-kb RS447 unit. Five of 60 parent-to-offspring transmissions clearly show changes in copy number, indicating a high frequency (~8.3%) of meiotic instability. Evidence for somatic mosaicism has also been observed. Searches of the database have revealed the presence of minor RS447 sequences mapping to chromosome 8p23, raising the possibility of a rearrangement or transposition of RS447 within the human genome. These results suggest that the unstable nature of RS447 megasatellite DNA gives rise to its hypervariability and may contribute to the structural dynamics of this repetitive DNA in the genome.

T. Okada · Y. Gondo · S. Hadano · J.-E. Ikeda (✉)
Department of Molecular Neuroscience,
Institute of Medical Sciences, Tokai University, Bohseidai,
Isehara, Kanagawa 259-1193, Japan
e-mail: joh-e@nga.med.u-tokai.ac.jp,
Tel.: +81-463-915014, Fax: +81-463-914993

Y. Gondo
RIKEN Genomic Sciences Center, Tsurumi, Yokohama,
Kanagawa 230-0045, Japan

J. Goto · I. Kanazawa
Department of Clinical Neuroscience,
Institute for Brain Research, Faculty of Medicine,
University of Tokyo, Bunkyo-ku, Tokyo 113-0021, Japan

S. Hadano · J.-E. Ikeda
NeuroGenes, International Cooperative Research Project,
Japan Science and Technology Corporation,
Tokai University School of Medicine, Isehara,
Kanagawa 259-1193, Japan

J.-E. Ikeda
Department of Paediatrics, Faculty of Medicine,
University of Ottawa, Ontario K1H 8M5, Canada

Introduction

Repetitive DNA sequences are major components of most eukaryotic genomes. Various types of repetitive sequences have been reported and categorized according to (in addition to the primary sequence) repeat-unit length, the number of repeats, and their mapping in the genome. The head-to-tail tandem repetitive sequences, such as microsatellite (Litt and Luty 1989), minisatellite (Jeffreys et al. 1985), or variable number of tandem repeat (Nakamura et al. 1987), macrosatellite (Kodama et al. 1987; Epstein et al. 1987; Zhu et al. 1990; Giacalone et al. 1992; van Deutekom et al. 1993; Lopez et al. 1994; Takahashi et al. 1994), and megasatellite DNA (Gondo et al. 1998), are highly polymorphic, repetitive DNA sequences postulated as making a substantial contribution to genomic instability because of their high mutation frequencies. Several hypothetical mutation models for repetitive DNA have been proposed following the extensive analyses of micro- or minisatellite DNA sequences. These include a coupling model involving DNA replication and repair process (Levinson and Gutman 1987) and a recombination-based model (Smith 1976; Tartof 1988).

Alteration in microsatellite DNA copy number occurs mainly in somatic cells (Weber and Wong 1993). Replication slippage (Streisinger et al. 1966; Weber 1990; Strand et al. 1993) and/or unequal exchange between sister chromatids (Gacy et al. 1998) are thought to contribute to this process. Mutation frequencies for microsatellite DNA are estimated to be in the range of 0.1%–0.2% in general (Weissenbach et al. 1992; Weber and Wong 1993; Heyer et al. 1997), although a higher mutation rate has also been reported (1.5% at the human *DXS981* locus; Mahtani and Willard 1993). Increased instability in microsatellite and/or minisatellite DNA is observed in many human cancer cells (Armour et al. 1989; Wooster et al. 1994; Eshleman et al. 1995), atherosclerotic plaques (Hatzistamou et al. 1996), and cells derived from irradiated animals (Dubrova et al. 1993; Jeffreys et al. 1997). In contrast, triplet repeat sequences, which are associated with several

inherited neuromuscular degenerative diseases, can undergo significant amplification, particularly in either male or female germinal cells (Richards and Sutherland 1992), despite their comparative stability in somatic cells.

Minisatellite DNAs are comparatively unstable repetitive sequences (Jeffreys et al. 1988), with particularly high mutation rates in germinal cells. The instability is affected both by the gender transmitting these alleles and the size of the repetitive sequences. In particular, the mutation frequencies of the hypervariable minisatellite DNA in the mouse genome are estimated to be approximately 10% (Fan et al. 1995). The highest mutation rate reported is ~15% at the human *CEB1* gene locus (Vergnaud et al. 1991). Several lines of evidence suggest that alteration in the tandem copy number of minisatellite DNA sequences occurs via complex gene conversion-like events with changes that are confined to one end of the tandem arrays with involvement of the flanking DNA sequences (Armour et al. 1993; Jeffreys et al. 1994). Allelic diversity of the mutation rates has also been demonstrated (Monckton et al. 1994; May et al. 1996).

Several tandem repeats composed of units of many kilobases have been identified in the human genome. These are sometimes called macro- or megasatellite DNA (Epstein et al. 1987; Giacalone et al. 1992; van Deutekom et al. 1993; Lopez et al. 1994; Gondo et al. 1998). Giacalone et al. (1992) have described the *DXZ4* locus, which is comprised of 50-100 copies of a basic 3-kb repeat, on chromosome X. The *D4Z4* locus on chromosome 4q35, which is associated with facioscapulohumeral muscular dystrophy (FSHD), consists of a heterochromatin-associated polymorphic 3.3-kb tandem repeat (van Deutekom et al. 1993). Recently, we have identified a novel highly polymorphic head-to-tail 4.7-kb tandem repetitive sequence, the RS447 megasatellite DNA, which maps to human chromosome 4p (major RS447; Kogi et al. 1997; Gondo et al. 1998). We have also mapped minor copies of RS447 DNA to chromosome 8p (minor RS447; Gondo et al. 1998). However, the polymorphic nature and inheritance of these megasatellite DNA sequences and the molecular mechanisms underlying the observed hypervariability/instability remain obscure, mainly because of the large size of these polymorphic loci, which contain a large numbers of those repeats.

The RS447 megasatellite is widely conserved among many mammalian species (Gondo et al. 1998). We have demonstrated that each human RS447 4.7-kb unit contains the *USP17* (ubiquitin-specific protease 17) gene encoding 60 kDa of a functional deubiquitinating enzyme and functional promoter elements (Saitoh et al. 2000). Further, we have shown that antisense RNAs for the *USP17* mRNA are also transcribed from the RS447 alleles, consistent with the existence of an antisense-transcript-mediated gene expression control for *USP17* (Saitoh et al. 2000). Given the potential functional significance for the RS447 allele, it is important to delineate the basic molecular mechanism for RS447 hypervariability, furthering our understanding of the relationship between hypervariability and expression of the intrinsic *USP17* gene and of the influence on chromatin structure and expression of the genes adjacent to the RS447 megasatellite DNA.

We have therefore investigated the meiotic transmission of RS447 DNA by using high-resolution pulsed-field gel electrophoresis (PFGE). We have identified 44 distinct alleles among 74 unrelated chromosomes containing 20–103 copies of the 4.7-kb RS447, revealing a high frequency of meiotic instability. Further, database searches have demonstrated the presence of the less conserved minor RS447 locus on chromosome 8p23. We discuss the potential impact of RS447 megasatellite DNA hypervariability and structural features on chromosomal structure and gene expression within and contiguous to the RS447 DNA.

Materials and methods

Cell lines and blood samples

Forty-six Epstein-Barr-virus-transformed lymphoblastoid cell lines established from 46 individuals belonging to eight Japanese pedigrees were used. Blood samples were obtained from nine volunteers belonging to two Japanese families and from ten unrelated individuals.

DNA preparation

Lymphoblastoid cells were suspended in phosphate-buffered saline at a concentration of 1×10^7 cells/ml. Aliquots of 1 ml peripheral blood samples were diluted with 9 ml ice-cold lysis buffer (20 mM Tris-HCl pH 8.0, 10 mM EDTA) and centrifuged at 1500 rpm for 10 min at 4°C. The pellets were resuspended in 250 µl lysis buffer. Cell suspensions were mixed with an equal volume of melted 1% InCert agarose (FMC), and gel plugs were made. Gel plugs were incubated twice in ESP solution (0.5 M EDTA pH 8.0, 1% sodium lauroylsalcosinate, 1 mg/ml proteinase K) for 24 h at 50°C, washed several times with TE buffer (10 mM TRIS-HCl, 1 mM EDTA, pH 8.0) at room temperature, and treated with 1 mM phenylmethylsulphonyl fluoride (PMSF) in TE buffer to inactivate the proteinase K. DNA gel plugs were washed with TE buffer at room temperature, soaked in an appropriate 1× reaction buffer for 60 min, and digested in 100 µl of fresh 1× reaction buffer containing restriction enzymes overnight according to the supplier's recommendations. Enzyme reactions were stopped by the addition of 10 µl 0.5 M EDTA. Finally, gel plugs were washed in TE buffer and subjected to PFGE.

PFGE protocol

PFGE was performed on the CHEF-DR II system (Bio-Rad). Initially, analyses were conducted on a 14×13-cm standard sized gel, viz., 1% SeaKem Gold agarose (FMC) in 0.25×TBE (1×TBE =0.09 M TRIS-borate, 0.002 M EDTA, pH 8.3) for 30 h at a 30-s pulse time. For high-resolution analyses, PFGE was performed by employing a longer gel (14×21 cm). Two parameters (pulse and running times) were changed depending on the size-range being analyzed (see Results). All experiments were conducted at 11–12°C with a fixed running voltage of 160 V. λ DNA-PFGE markers (Pharmacia Biotech) were used as a size standard.

Southern blotting and hybridization

DNA was stained with ethidium bromide, nicked by a UV exposure of 10 s, transferred onto a GeneScreen nylon membrane (DuPont) under alkaline conditions, and fixed by UV cross-linking

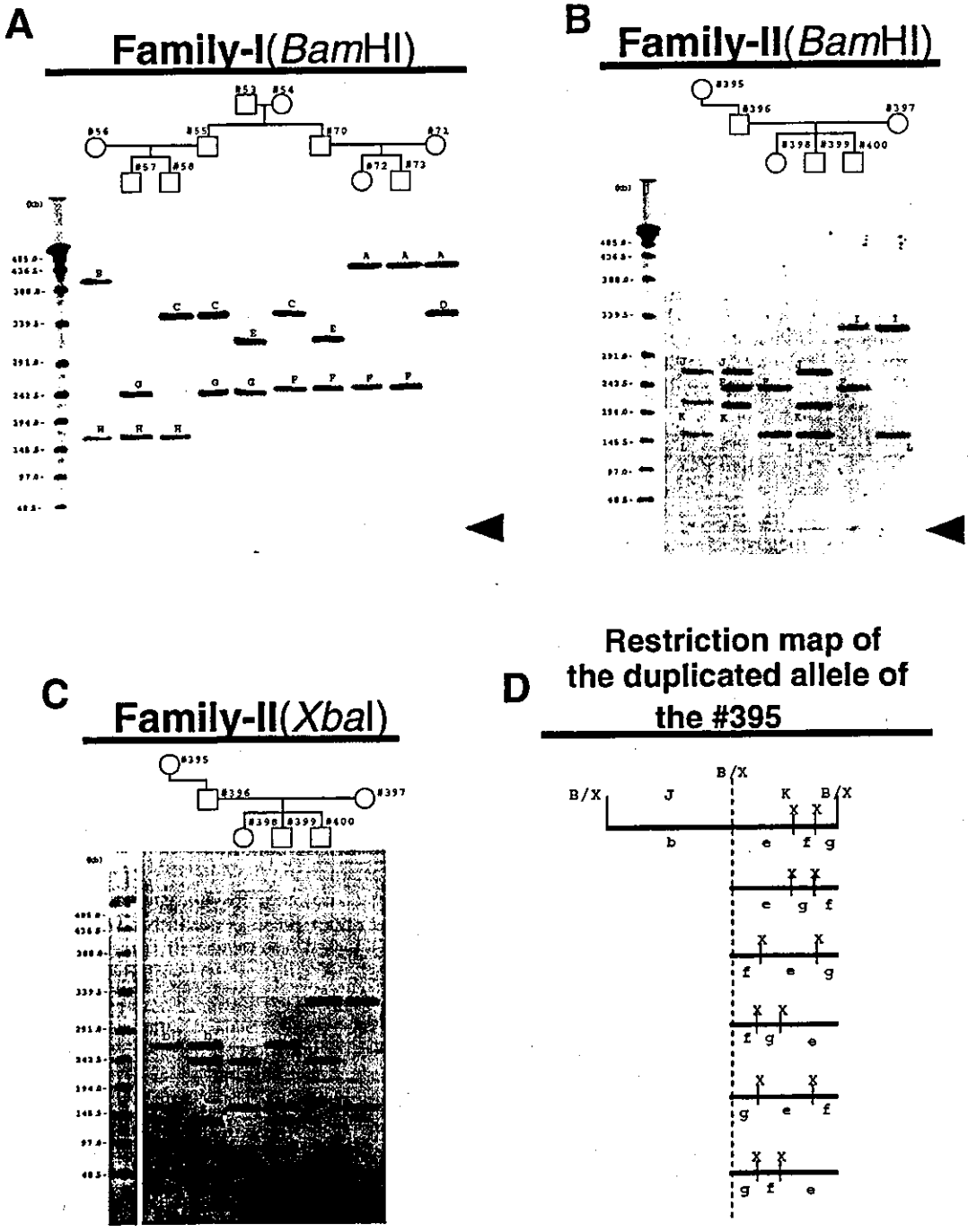


Fig. 1A–D Southern blots of PFGE gel showing Mendelian inheritance of the RS447 megasatellite DNA in Japanese pedigrees. Blots of *Bam*HI-digested DNA of Family-I (**A**) and Family-II (**B**) and *Xba*I-digested DNA of Family-II (**C**). DNA samples were separated by standard PEGE (voltage: 160 V; pulse time: 30 s; running time: 30 h) and blotted onto nylon membranes. Southern blot analysis was conducted by using the biotin-labeled 4.7-kb *Eco*RI fragment with the RS447 repeating unit as a probe. *Numbers* ID numbers for individuals, *uppercase letters* different sizes of the *Bam*HI fragments containing the RS447 DNA, *lowercase letters* different sizes of the *Xba*I fragments containing the RS447 DNA, *arrowheads right* (**A, B**) positions of bands derived from the minor RS447 DNA on chromosome 8p. The λ DNA concatemer *left* (**A–C**) was used as a size standard. **D** Predicted restriction map of the duplicated allele of the RS447 DNA derived from individual 395, according to results obtained from two Southern blot analyses (**B, C**). Sites of the restriction enzymes are shown as follows: *X XbaI* site, *B/X* closely located *Bam*HI and *Xba*I site with unknown order of their positions. *Upper case letters* Restriction fragments corresponding to bands shown in **B** or **C**. All six possible orders for three fragments (*e–g*) are shown

(Strata Linker; Stratagene). Hybridization was carried out in modified Church's buffer (Church and Gilbert 1984; Brilliant et al. 1991) containing 125 μg/ml Herring testis DNA (Sigma) at 65°C overnight with biotin-labeled RS447 and λ phage DNA probes. The final washing condition was: 0.1×SSC (1×SSC=150 mM NaCl, 15 mM sodium citrate, pH 7.0), 1% SDS at 65°C. Signals were detected by chemiluminescence in a NEBlot Phototope system (New England BioLabs) on X-ray film (X-OMAT; Kodak).

DNA sequence analysis and database searches

The DNA sequences were compared with the non-redundant nucleotide sequence database and the Human Genome DNA sequence database at NCBI (http://www.ncbi.nlm.nih.gov) by using the BLASTN program (Altschul et al. 1997). Dot-plot analyses were performed to display pair of homologous DNA sequences, with GeneWorks Version 2.5.1 (Oxford Molecular Group). Identification of repetitive sequences was carried out by using RepeatMasker (http://ftp.genome.washington.edu:80/cgi-bin/RepeatMasker).

Results

Transmission of RS447 megasatellite alleles in Japanese pedigrees

The determination of entire sizes for RS447 megasatellite DNA is achieved by separating *Bam*HI- or *Xba*I-digested genomic DNA samples by standard PFGE and Southern blot analysis (Gondo et al. 1998). By using this method, the major RS447 megasatellite alleles mapping to chromosome 4p can be distinguished from the minor RS447 alleles located on chromosome 8p. The latter locus, which is relatively non-polymorphic, contains an RS447 allele of an approximately 50 kb in size (Gondo et al. 1998).

To investigate the inheritance pattern of the major RS447 locus, we conducted PFGE analyses of 55 kindred from 10 independent Japanese families. Restriction fragments containing RS447 sequence were clearly separated and visualized by this method. A representative result for a three-generation family (Family-I) is shown (Fig. 1A). In this case, eight distinct *Bam*HI fragments, A–H, representing eight RS447 megasatellite DNA alleles were identified among these kindred. Interestingly, all individuals in Family-I carried two distinctive RS447 alleles (heterotypes), indicating a high level of heterozygosity. All transmissions from parents to offspring were clearly in accordance with Mendelian inheritance. Under these conditions, 51 of 55 cases (93%) clearly showed heterotypes for the RS447 tandem copy numbers. However, in the remaining four individuals (7%), a single band with stronger intensity (data not shown) was observed, indicating that they possessed two indistinguishable RS447 alleles (homotypes).

Southern blot analysis of DNA derived from a Chinese hamster ovary/human somatic cell hybrid harboring a single human chromosome 8 has shown that the faint ~20-kb bands represent the less variable minor RS447 DNA mapping to chromosome 8p (Gondo et al. 1998). Bands of a similar size were also observed in other families examined in this study (Fig. 1A, B and data not shown).

Unusual restriction pattern for the RS447 megasatellite alleles

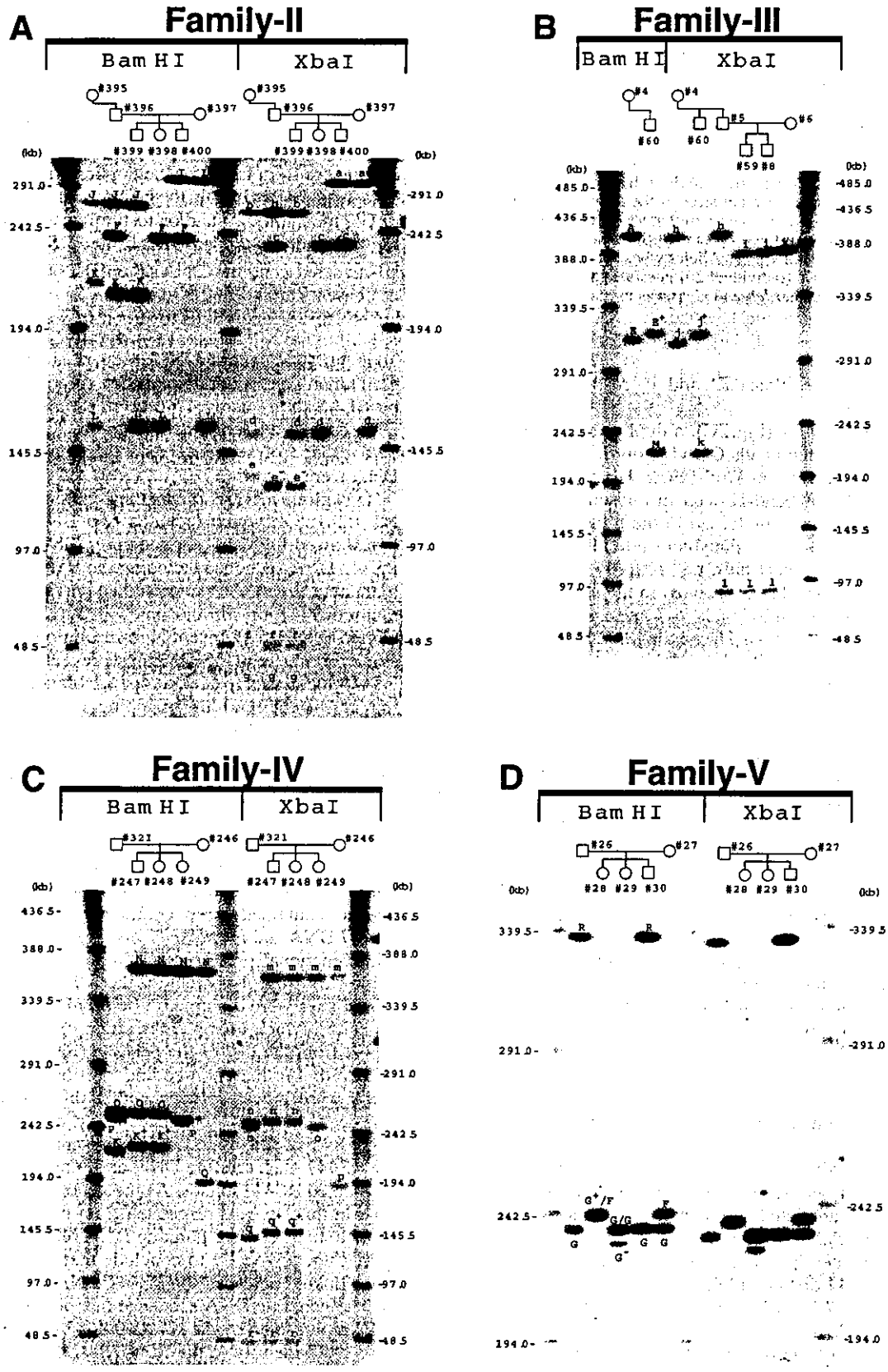
As neither *Bam*HI nor *Xba*I cut within the RS447 tandem repeat (Gondo et al. 1998), these would be expected to excise the intact RS447 repeat from the rest of the genome. However, we have identified six independent alleles in the 74 unrelated chromosomes (~8.1%); these alleles showed more than three bands for the RS447 megasatellite DNA on the PFGE-Southern blot analysis. In Family-II, individuals 395, 396, and 399 exhibit three bands on the *Bam*HI-digested genomic DNA (*Bam*HI-PFGE; Fig. 1B). In these cases, bands J and K were consistently transmitted through three-generations together. In addition, PFGE analysis of *Xba*I digested (*Xba*I-PFGE) revealed five bands in the same individuals (Fig. 1C), four of which (bands b, e, f, and g) co-segregated. As the sum of the *Bam*HI fragments J and K and the *Xba*I fragments b, e, f, and g were approximately equal, we assume that these multiple bands, which we refer to as the duplicated allele, originate from a single chromosome. A restriction map determining the six most likely orders of the *Xba*I positions within this allele is shown in Fig. 1D. The duplicated RS447 alleles observed in Family-II (Figs. 1B, 2C) and other two pedigrees (data not shown) clearly show Mendelian inheritance.

High frequency copy-number changes in the RS447 megasatellite DNA

To determine the RS447 megasatellite copy number for each allele precisely, we established three different high-resolution PEGE conditions depending on the targeted DNA sizes, allowing the detection of a single RS447 4.7-kb unit-copy difference. Re-analysis of the same Japanese pedigrees under these conditions revealed several previously undetected band-size shifts, reflecting copy-number differences, in four of 10 pedigrees (Fig. 2). The frequency of intergenerational copy-number alteration was 8.3% (five of 60 informative parent-to-offspring transmissions). Four cases showed a single repeat unit increment, whereas one case exhibited a single unit-length decrement (Fig. 2, Table 1).

Bands J and K on *Bam*HI-PFGE and bands b, e, f, and g on *Xba*I-PFGE representing duplicated RS447 alleles are observed in Family-II (Fig. 2A). High-resolution analysis revealed a 1-unit loss of band K in individual 395 as transmitted to her son (individual 396), reflecting an approximately 5-kb smaller band, i.e., K^- (217→212 kb, 46→45 copies). The *Xba*I band e in the same individual showed a similar reduction in size. In both cases, the smaller-sized bands K^- and e^- in individual 396 were stably transmitted to his son (individual 399), suggesting that loss of the repeat unit occurred in the germ line of individual 395.

In Family-III (Fig. 2B), both the *Bam*HI band E and the *Xba*I band j, which are seen in individual 4 (~315 kb, ~66 copies), are transmitted to her son (individual 60)


A
Family-II
Bam HI
XbaI
#395
#396
#397
#399 #398 #400
(kb)
291.0
242.5
194.0
145.5
97.0
48.5
B
Family-III
Bam HI
XbaI
#4
#5
#6
#60
#59 #8
485.0
436.5
388.0
339.5
291.0
242.5
194.0
145.5
97.0
48.5
C
Family-IV
Bam HI
XbaI
#321
#246
#247 #248 #249
436.5
388.0
339.5
291.0
242.5
194.0
145.5
97.0
48.5
D
Family-V
Bam HI
XbaI
#26
#27
#28 #29 #30
339.5
291.0
242.5
194.0

◀ **Fig. 2A–D** Unstable transmission of the RS447 megasatellite DNA in Japanese pedigrees. The *Bam*HI- or *Xba*I-digested genomic DNA samples were separated by a high-resolution PEGE method: **A** Family-II (voltage: 160 V; pulse time: 15 s; running time: 52 h), **B** Family-III (voltage: 160 V; pulse time: 26 s; running time: 42.5 h), **C** Family-IV (voltage: 160 V; pulse time: 26 s; running time: 42.5 h), **D** Family-V (voltage: 160 V; pulse time: 21 s; running time: 80 h). Southern blot analysis was conducted by using the biotin-labeled 4.7-kb *Eco*RI fragment, with the RS447 repeating unit as a probe. *Numbers* ID numbers for individuals, *uppercase letters* different sizes of the *Bam*HI fragments containing the RS447 DNA, *lowercase letters* different sizes of the *Xba*I fragments observed in each family. Six cases of single-copy alterations were identified. $^{+}$, $^{-}$ Alleles whose sizes are either increased or decreased, G^{-} (individual 29) possible somatic mosaicism. The λ DNA concatemer *left* and *right* was used as a size standard

as the larger alleles E^{+} and j^{+}, respectively (~320 kb, ~67 copies).

In Family-IV (Fig. 2C), a duplicated RS447 pattern is observed in the bands O and K on *Bam*HI-PFGE and the bands n, q, and r on *Xba*I-PFGE. High-resolution analysis revealed that band K seen in individual 321 (~220 kb, ~46 copies) was increased in the transmission to his son (individual 247) and daughter (individual 248) by approximately one repeat unit, i.e., K^{+} (~47 copies). The *Xba*I-band q (~141 kb, ~30 copies) in the same individual was transmitted to individuals 247 and 248 as a larger band, i.e., q^{+} (~31 copies). The finding that both children (individuals 247 and 248) demonstrate the same increase in allele size suggests that the changes occurred in the germ cells of individual 321.

Family-V showed a complicated inheritance pattern (Fig. 2D) of the *Bam*HI- and *Xba*I-PFGE bands, which were indistinguishable. The father (individual 26) and mother (individual 27) have R/G and F/G genotypes, respectively. Two children (individuals 28 and 29) are homotypes. Individual 28 with a single 241-kb band (~51 copies) appears to received the F allele from individual 27 and the other G^{+}, representing an increased-sized allele of G, from her father (individual 26). Individual 29 appears to have inherited two G alleles (~235 kb, ~50 copies) from each parent. Remarkably, this individual also has a minor G^{-} band, representing a one-copy decreased fragment (~49 copies) on both *Bam*HI- and *Xba*I-PFGEs, and suggesting a somatic mosaicism of the RS447 DNA in individual 29.

Allele size distribution of RS447 megasatellite DNA

In this study, we analyzed 110 chromosomes from 55 individuals belonging to 10 families and identified 54 genetically unrelated chromosomes originated from 30 individuals (data not shown). Further, 20 chromosomes from 10 unrelated individuals were also analyzed. We found a total of 44 distinct alleles ranging from 20 to 103 copies for the 4.7-kb RS447 unit in 74 unrelated chromosomes on the *Bam*HI-PFGE (Fig. 3; mean±SD=64±21 copies). The size of the duplicated RS447 alleles was calculated by summing two co-segregating bands on *Bam*HI-PFGE (Fig. 3, hatched bar). Interestingly, six of the alleles found in the duplicated RS447 DNA sequences (93, 100, 101, 102, 102, 103 copies estimated from *Bam*HI-PFGE) were among the largest observed. Although most of the copy numbers for RS447 estimated from *Xba*I-PFGE were consistent with those from *Bam*HI-PFGE, there were slight differences from those for the duplicated alleles (Table 1). This may be attributed to the limit of resolution of the PFGE methodology and/or to the consequence of structural diversity in the large duplicated RS447 alleles.

Genomic organization of the RS447 megasatellite DNA

To identify the genomic DNA sequences that flank RS447, the human genome draft sequence database was searched by BLASTN by using 4746 bp of the *Eco*RI-RS447 unit DNA sequence (GenBank account no. D38378) as a query. Two groups of sequence data that matched the RS447 DNA were identified.

One group represented by the genomic DNA sequence; NT_028165.3, which included two DNA sequence-contigs showed high sequence identity (98%–99%) with the RS447 DNA sequence including the *USP17* coding region (data not shown). Although we had previously mapped the major RS447 to chromosome 4p15 (Kogi et al. 1997; Gondo et al. 1998), recent high-resolution fluorescence in situ hybridization (FISH) analysis revealed that the major RS447 locus was located at 4p16.1 (data not shown), being consistent with the mapping of the NT_028165.3 DNA in the database. Unfortunately, no obvious flanking DNA sequence was identified within the genomic DNA sequence in the contigs. However, it was noteworthy that the first 4-kb region of a single contig consisting of 20,344 bp of the genomic DNA sequence demonstrated less homology to RS447 (<90% identity; data not shown). Dot-plot comparison of the 4.7-kb *Eco*RI-RS447 unit with this contig demonstrated that this genomic DNA sequence contained three complete *Eco*RI units for the RS447 DNA and four *USP17* open reading frames (ORFs) with functional promoters (Saitoh et al. 2000; Fig. 4A). Within the first 4-kb region, the 4.7-kb intervals of the *Eco*RI sites were disrupted. The *Bam*HI and *Xba*I restriction sites, which liberated the entire or duplicated RS447 repeat from the rest of the genome (Figs. 1, 2), were also observed. Further, non-coding portions for the RS447 DNA were repeated twice within this region (Fig. 4A). The decay in RS447 sequence observed in the 4-kb region suggests that this region may represent the outer copy of the RS447 megasatellite DNA or the linker region for the duplicated RS447 alleles. As the genomic DNA sequence used in this analysis originated from draft sequence data, further studies on the flanking DNA sequences will be required unequivocally to conclude that this represents the outer/linker copies for the RS447 DNA.

The genomic DNA sequence mapping to chromosome 8p23 represented by the NT_019483.7 sequence exhibited 94%–96% identity to the RS447 DNA sequence. The lo-

Table 1 Summarized results of 44 distinct RS447 alleles observed in 74 unrelated chromosomes

Allele ID	Chromosome no.	*BamHI*			*XbaI*			Allele type[c]
		Band size (kb)	Copy no[a]	Band ID in Figs. 1, 2	Band size (kb)[b]	Copy no.[a, b]	Band ID in Figs. 1, 2	
1	2	94~97	20		ITB	ITB	l	C
2	1	102	21		ITB	ITB		C
3	2	155~156	33	L	ITB	ITB	d	C
4	1	163	34	H	ITB	ITB		C
5	1	176	37		ITB	ITB		C
6	1	187	39		ITB	ITB		C
7	1	192	40	Q	ITB	ITB	p	C
8	1	198	42		ITB	ITB		C
9	1	208	44		ITB	ITB		C
10	1	227	48	M	ITB	ITB	k	C
11	2	231	49	G^-	ITB	ITB		C
12	4	235~238	50	G	ITB	ITB		C
13	8	240~244	51	F, G^+	ITB	ITB	c	C
14	3	246~249	52	P	ITB	ITB	o	C
15	1	255~258	54		ITB	ITB		C
16	2	265~267	56		ITB	ITB		C
17	2	278~279	59		ITB	ITB		C
18	2	283~285	60		ITB	ITB		C
19	1	298	63		ITB	ITB		C
20	1	304~305	64		ITB	ITB		C
21	4	312~315	66	E	ITB	ITB	j	C
22	1	320	67	E^+	ITB	ITB	j^+	C
23	2	321~325	68	I	ITB	ITB	a	C
24	1	333	70		ITB	ITB		C
25	2	337~338	71	R	ITB	ITB		C
26	1	341	72		ITB	ITB		C
27	1	348	73	D	ITB	ITB		C
28	2	350~351	74	C	ITB	ITB		C
29	1	361	76		ITB	ITB		C
30	2	368~371	78	N	ITB	ITB	m	C
31	2	386	81		ITB	ITB	i	C
32	1	396	83		ITB	ITB		C
33	2	406	86	B	ITB	ITB		C
34	1	415	87		ITB	ITB		C
35	1	425	90		ITB	ITB		C
36	4	437~439	92	A	ITB	ITB	h	C
37	1	136 304~305	29 64 [93]		72 132~134 235~238	15 28 50 [93]		D
38	1	454	96		ITB	ITB		C
39	1	462	97		ITB	ITB		C
40	1	217~220 255~258	46 54 [100]	K O	45 141~144 252	9 30 53 [92]	r q n	D

Table 1 (continued)

Allele ID	Chromosome no.	*BamHI*			*XbaI*			Allele type[c]
		Band size (kb)	Copy no[a]	Band ID in Figs. 1, 2	Band size (kb)[b]	Copy no.[a, b]	Band ID in Figs. 1, 2	
41	1	221~225	47	K^+	46~47	10	r	D
		255~258	54	O	147	31	q^+	
			[101]		252	53	n	
						[94]		
42	1	221~225	47		34	7		D
		259	55		53	11		
			[102]		141~144	30		
					257~258	54		
						[102]		
43	1	212	45	K^-	26	5	g	D
		269~271	57	J	46~47	10	f	
			[102]		127	27	e^-	
					265~268	56	b	
						[98]		
44	1	217~220	46	K	26	5	g	D
		269~271	57	J	46~47	10	f	
			[103]		132~134	28	e	
					265~268	56	b	
						[99]		

[a]Numbers in square brackets represent entire copy numbers for the RS447 units in duplicated alleles.
[b]In columns *Band size* and *Copy no.*, *ITB* indicates that both the estimated size and copy number of the RS447 alleles estimated from the *Xba*I-PFGE are identical to those observed in the *Bam*HI-PFGE.
[c]Allele types *C* and *D* represent the confined and duplicated RS447 alleles, respectively.

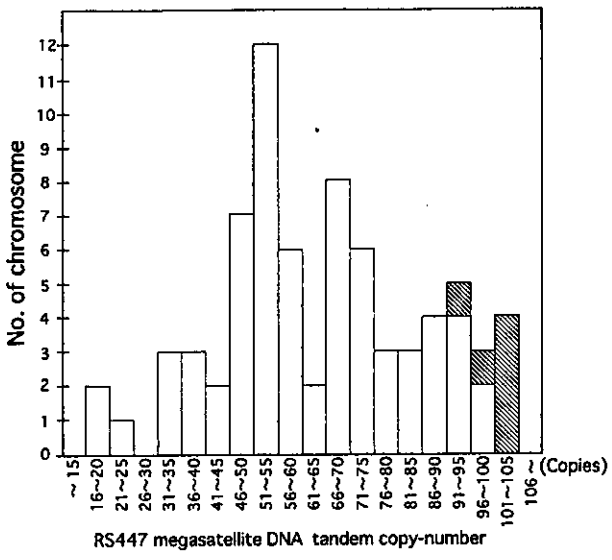


Fig. 3 Distribution of the RS447 repetitive unit sizes in 74 unrelated chromosomes estimating from the sizes of the *Bam*HI bands on PFGE-Southern blot analysis. *Hatched bars* Duplicated alleles whose copy numbers were calculated by summing co-segregated *Bam*HI fragments. Alleles that were predicted to carry two indistinguishable sizes of the RS447 DNA (homotypes) were counted twice (e.g., individual 6 in Family-III; Fig. 2B)

calization of these RS447 DNA sequences on this chromosomal region (minor RS447) is consistent with our previous FISH mapping data (Gondo et al. 1998). Dot-plot analysis of the 4.7-kb *Eco*RI-RS447 unit compared with the genomic DNA sequence (NT_019483.7; 1394~1418 kb region) demonstrated that this genomic DNA sequence contained three units of the RS447 DNA sequences within the 24-kb region (Fig. 4B). Although the ORF structure for *USP17* in the middle copy of the minor RS447 was conserved, two other copies coded for the truncated forms. The positions and/or intervals of restriction sites such as *Eco*RI, *Bam*HI, and *Xba*I in the minor RS447 repetitive DNA were also less conserved. Thus, it appears that the minor RS447 on chromosome 8p23 consists of both decayed and non-decayed copies of the RS447 repeat unit encoding both functional and non-functional USP17.

Finally, database searches of repetitive DNA sequences revealed the presence of direct repeat L1/LINE sequences outside the minor RS447 tandem repetitive array on chromosome 8q23 (Fig. 4B). These L1/LINE sequences are truncated at their 5' ends and are comprised of approximately 2.5 kb of the 3' portion of the L1/LINE consensus DNA sequence, including the poly-A tail. As integrated copies of L1/LINE sequence are usually truncated at the 5' end (Kazazian and Moran 1998), these flanking L1/LINE sequences may represent previous transposition and/or integration events in the genome.

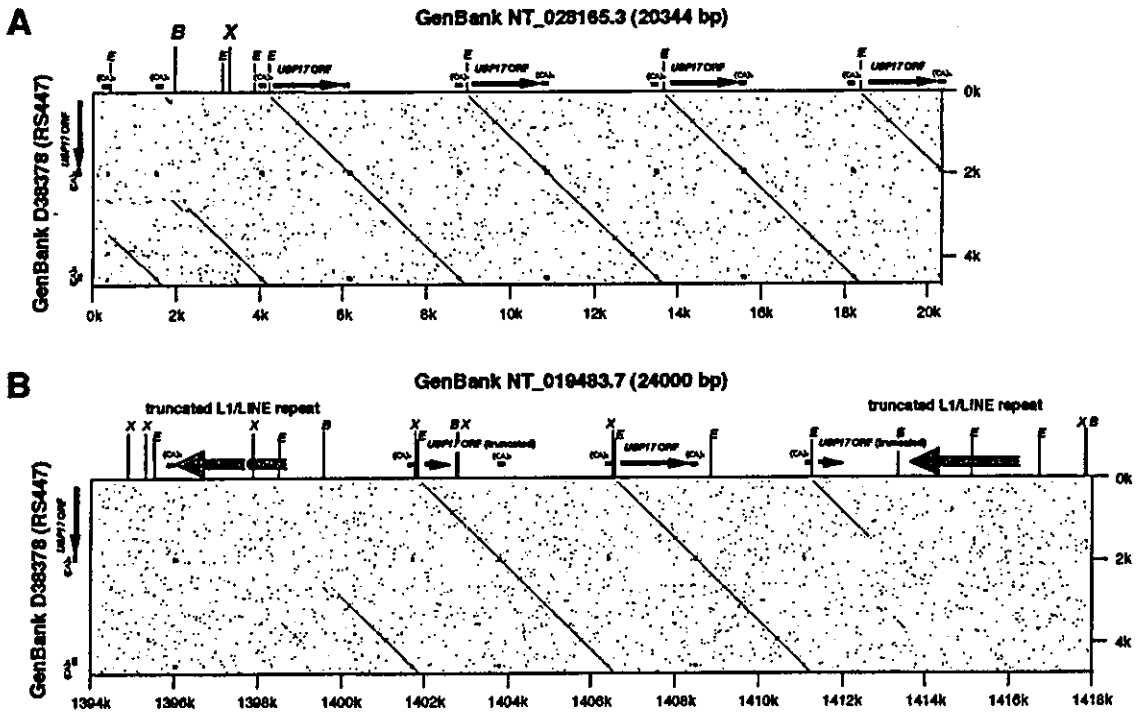


Fig. 4A, B Dot-plot analyses of the RS447 megasatellite DNA. A Comparison of the 4.7-kb *Eco*RI RS447 unit (*y axis*) with the genomic DNA sequence (NT_028165.3; 20,344 bp) mapping to chromosome 4p16.1 obtained from the Human Genome Draft sequence database (*x axis*). Positions of the ORF for the *USP17* gene, the CA-dinucleotide repeats, and three restriction enzymes, *Bam*HI (*B*), *Xba*I (*X*), and *Eco*RI (*E*), are shown along the *x axis*. B Analysis of the 4.7-kb *Eco*RI RS447 unit (*y axis*) compared with a portion of the genomic DNA sequence (NT_019483.7; 24000 bp) mapping to chromosome 8p23 (*x axis*). *Large arrows* Truncated L1/LINE repetitive sequences, *small arrows* either the complete ORF or the disrupted ORF for the *USP17* gene. Positions of three restriction enzymes, *Bam*HI (*B*), *Xba*I (*X*), and *Eco*RI (*E*), are shown along the *x* axis

Discussion

The RS447 megasatellite DNA is a highly polymorphic head-to-tail tandem repetitive sequence, which is widely conserved among many mammalian species (Gondo et al. 1998). We have previously demonstrated that each 4.7-kb unit comprising human RS447 contains the *USP17* gene encoding a 60-kDa functional deubiquitinating enzyme, together with functional promoter elements (Saitoh et al. 2000). Further, antisense RNAs for the *USP17* mRNA are also transcribed from RS447, consistent with an antisense-mediated control of *USP17* gene expression (Saitoh et al. 2000). However, the molecular mechanism underlying the hypervariability of the RS447 locus and the relationship between RS447 hypervariability and the expression of the intrinsic *USP17* and those genes adjacent to the RS447 DNA is unknown.

As a first step toward an understanding of these issues, we have investigated the nature and inheritance of the RS447 megasatellite DNA and of the tandem repeat copy number through meiotic transmission. Forty-four distinct alleles containing from 20 to 103 copies of the 4.7-kb RS447 unit were observed in 74 unrelated chromosomes in Japanese individuals. Of 60 parent-to-offspring transmissions, five alleles clearly showed changes in copy number, indicating a high frequency (~8.3%) of meiotic instability. One case of somatic mosaicism was also observed. Taken together, these data suggest that RS447 megasatellite DNA is hypervariable and comparatively unstable in both germinal and somatic cells, although the frequency of somatic mutation may be lower than that observed in meiotic transmissions. In view of the high frequency of copy-number alterations in meiotic transmissions, it is likely that the wide range of copy-number differences seen in RS447 alleles among Japanese individuals has resulted from the progressive accumulation of copy-number alterations in the RS447 alleles.

High frequencies (~10%) of minisatellite copy-number alterations in meiotic transmissions have previously been reported (Vergnaud et al. 1991; Andreassen and Olaisen 1998). The hypervariablity and high frequency of copy-number alteration observed in the RS447 megasatellite DNA in meiotic transmissions are comparable with those seen in the minisatellite DNA sequences, despite the significant difference in repeat unit size between the classes of DNA repeats (Charlesworth et al. 1994; RS447 vs minisatellite: 4.7 kb vs ~15 bp). Common mechanisms may exist underlying the unstable nature of these repetitive DNA sequences in germinal cells. One such model could be gene conversion through unequal crossing-over between sister chromatids (Smith 1976). Previous studies have demonstrated that copy-number changes are con-

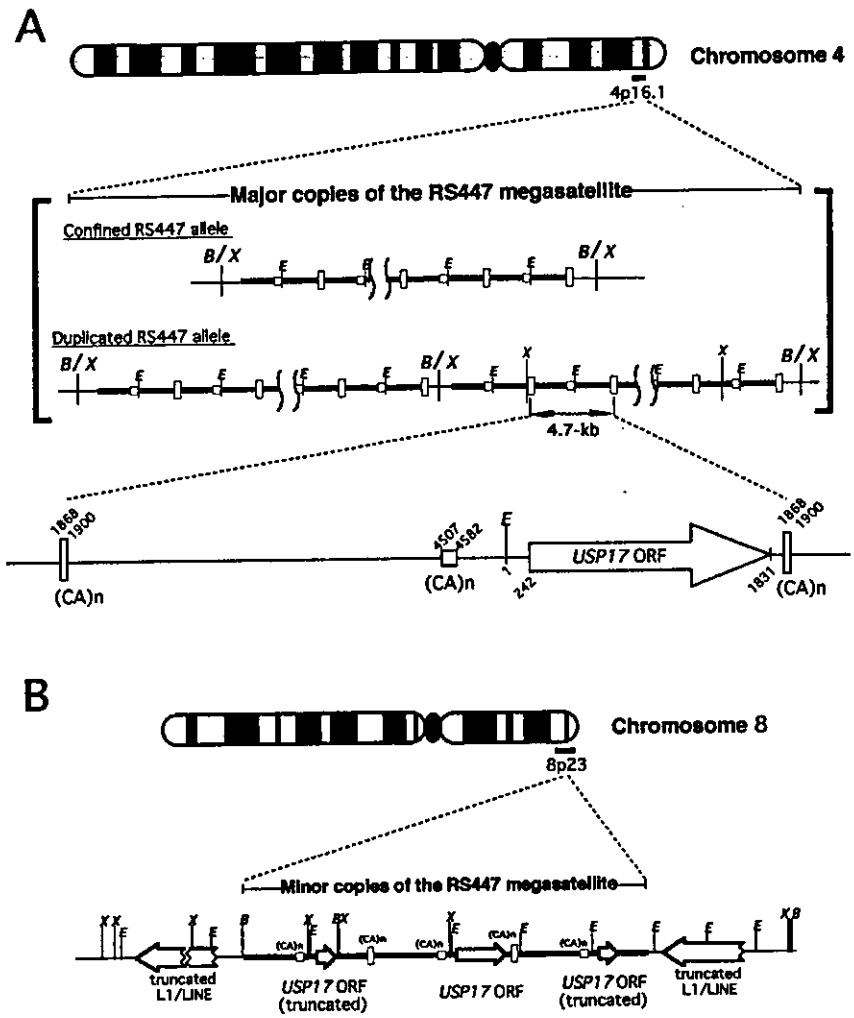


Fig. 5A, B Schematic representation of the RS447 megasatellite DNA. A RS447 megasatellite DNA mapping to chromosome 4p16.1. Idiogram of the human G-banded chromosome 4, organization of the RS447 megasatellite DNA comprised of tandem array of 4.7-kb repeat units, and multiple copies of the *USP17* gene encoded by these repeat units. Both the confined and the duplicated RS447 alleles are shown. The restriction map is depicted according to the results previously described (Kogi et al. 1997; Gondo et al. 1998); *Bam*HI (*B*), *Xba*I (*X*), *Eco*RI (*E*). **B** Minor copies of the RS447 megasatellite DNA mapping to chromosome 8p23. Idiogram of the human G-banded chromosome 8, organization of the RS447 megasatellite DNA with the flanking L1/LINE DNA sequences, and the *USP17* gene or decayed copies of the *USP17* genes. Orientations both of the major and of the minor RS447 DNAs are currently undetermined

fined to one end of the tandem arrays via complex gene-conversion-like events (Armour et al. 1993; Jeffreys et al. 1994). Allelic mutation-rate diversity of minisatellite DNA within the genome and the involvement of the flanking arrays in repeat-length mutation have also been demonstrated (Monckton et al. 1994; May et al. 1996). Finally, the high mutation rate observed in this study may reflect conditions at this particular genomic region, such as its being a hot spot of unequal crossing-over. The definition of the precise molecular mechanism for the small copy-number changes in RS447 megasatellite DNA requires further study.

One comparatively clear inference from this study is that the RS447 megasatellite DNA on chromosome 4 has been involved in a previous large-scale duplication with a subsequent increase in sequence diversity. The duplicated RS447 alleles are comprised of two co-segregating *Bam*HI fragments ranging between 136 kb and 304 kb. The size of the entire RS447 arrays for these duplicated RS447s range from 440 kb to 488 kb (93–103 copies). We assume that the duplicated RS447 alleles were generated by a large-scale duplication of the RS447 allele with the accompanying flanking *Bam*HI restriction sites. In view of the finding that flanking DNA sequences for the tandem arrays of repetitive DNA are known to contribute to large-scale duplications, the identification of these sequences for the major RS447 locus is a primary objective. In this regard, the identification of two truncated L1/LINE repetitive DNA sequences that flank the minor RS447 megasatellite DNA on chromosome 8p23 may be noteworthy.

The DNA sequence of RS447 was originally determined by sequencing plasmid DNA containing 4.7-kb *Eco*RI units (Kogi et al. 1997). However, the authentic RS447 repeat unit remained to be determined. In this study, DNA sequence analyses has shown that the minor RS447 megasatellite repeat is comprised of three 4.7-kb DNA units starting from and ending at the middle portions of the *Eco*RI units close to the CA-dinucleotide repeats (Figs. 4, 5). Disruption of the repetitive structure for the major RS447 DNA sequences on chromosome 4p16.1 was also observed in the vicinity to the CA-repeat in the

first 4-kb of the putative outer/linker copy of the RS447 DNA alleles (Fig. 4A). These data suggest that the RS447 megasatellite repeat unit consists of the 4.7-kb sequence flanked by the CA-dinucleotide repetitive sequences that reside in the middle of the *Eco*RI-RS447 DNA sequence. These flanking CA-dinucleotide repetitive sequences may contribute to the sequence diversity and hypervariability or instability of the RS447 DNA sequences.

Recently, a high frequency of mitotic chromosomal rearrangements has been identified between *D4Z4* repeats on chromosome 4 and highly homologous sequences on chromosome 10 (van der Maarel et al. 2000; van Overveld et al. 2000). Interestingly, a large-scale chromosome duplication has been found in chromosome 4p of patients with 4p duplication syndromes (Rodriguez et al. 1999; Xu et al. 1999). An inverted duplication of the 8p region, in which the minor RS447 DNA is closely located, has also been reported (Nelson et al. 1998). Collectively, the RS447 megasatellite DNA array could be associated with these chromosomal rearrangements.

The macro- or megasatellite DNA copy number may affect gene expression and phenotype and may sometimes cause genetic disease (Lupski 1998). In FSHD for example, a lower copy number of 3.3-kb tandem repeats *D4Z4* is strongly associated both with lower expression levels of adjacent genes and disease severity (van Deutekom et al. 1993). Further, hypermethylation of chromosomal DNA may lead to an increased level of he heterochromatin formation (Laird 1990; Sabl and Laird 1992), thereby altering the expression of adjacent genes. We have established that the RS447 allele is partially methylated (Y. Saitoh et al., unpublished). The copy number and the methylation status in RS447 DNA may affect both the chromatin structure and the expression of genes in the immediate region. Moreover, several studies have demonstrated that certain neurodegenerative diseases, such as Parkinson's disease, are associated with failure of the ubiquitin-proteasome system (McNaught et al. 2001). Thus, the expression level of the *USP17* gene, which may be associated with RS447 copy number and methylation status, may impact the activity of USP17 deubiqutinating enzyme and lead to dysfunction of the ubiquitin-proteasome system, which in turn may modify or even cause the disease(s).

In conclusion, we have elucidated the structure and hypervariable nature of RS447 megasatellite DNA. We have recently cloned tandem copies of RS447-like megasatellite DNA containing *USP17* homologs in other mammals and revealed the conservation of the head-to-tail tandem structure of the *USP17* homologs (Gondo et al. 1998; Y. Gondo et al., unpublished). These data clearly suggest a functional significance for the tandem array DNA structure for the *USP17* gene. We are currently undertaking both the isolation of the DNA sequences that flank the human major RS447 locus and the functional analysis of the deubiquitinating USP17 enzyme. These studies should provide further insights into the molecular mechanisms underlying RS447 hypervariability and the biological and possibly pathological roles of RS447 megasatellite DNA.

Acknowledgements We thank volunteers who donated blood for DNA extraction. We are also grateful to Dr. Alex E. MacKenzie at Ottawa University for help in the preparation of the manuscript. A part of this work was supported by a Grant-in-Aid for Scientific Research on Priority Areas, for Scientific Research (A), and for Scientific Research (C) from the Ministry of Education, Science, Sports, and Culture, Japan, and by a grant from the Ministry of Health and Welfare, Japan.

References

Altschul 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. Nucleic Acids Res 25:3389–3402

Andreassen R, Olaisen B (1998) De novo mutations and allelic diversity at minisatellite locus D7S22 investigated by allele-specific four-state MVR-PCR analysis. Hum Mol Genet 7:2113–2120

Armour JAL, Patel I, Thein SL, Fey MF, Jeffreys AJ (1989) Analysis of somatic mutations at human minisatellite loci in tumors and cell lines. Genomics 4:328–334

Armour JAL, Harris PC, Jeffreys AJ (1993) Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. Hum Mol Genet 2:1137–1145

Brilliant MH, Gondo Y, Eicher EM (1991) Direct molecular identification of the mouse pink-eyed unstable mutation by gemome scanning. Science 252:566–569

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371: 215–220

Church GM, Gilbert W (1984) Genomic sequenceing. Proc Natl Acad Sci USA 81:1991–1995

Deutekom JCT van, Wijmenga C, Tienhoven EAE van, Gruter A-M, Hewitt JE, Padberg GW, Ommen G-JB van, Hofker MH, Frants RR (1993) FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeat unit. Hum Mol Genet 2:2037–2042

Dubrova YE, Jeffreys AJ, Malashenko AM (1993) Mouse minisatellite mutations induced by ionizing radiation. Nat Genet 5: 92–94

Epstein ND, Karlsson S, O'Brien S, Modi W, Moulton A, Nienhuis AW (1987) A new moderately repetitive DNA sequence family of novel organization. Nucleic Acids Res 15:2327–2341

Eshleman JR, Lang EZ, Bowerfind GK, Parsons R, Vogelstein B, Willson JKV, Veigl ML, Sedwick WD, Markowitz SD (1995) Increased mutation rate at the *hprt* locus accompanies microsatellite instability in colon cancer. Oncogene 10:33–37

Fan YJ, Wang Z, Sadamoto S, Ninomiya Y, Kotomura N, Kamiya K, Dohi K, Kominami R, Niwa O (1995) Dose-response of a radiation induction of a germline mutation at a hypervariable mouse minisatellite locus. Int J Radiat Biol 68:177–183

Gacy AM, Goellner GM, Spiro C, Chen X, Gupta G, Bradbury EM, Dyer RB, Mikesell MJ, Yao JZ, Johnson AJ, Richter A, Melancon SB, McMurray CT (1998) GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. Mol Cell 1:583–593

Giacalone J, Friedes J, Francke U (1992) A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. Nat Genet 1:137–143

Gondo Y, Okada T, Matsuyama N, Saitoh Y, Yanagisawa Y, Ikeda J-E (1998) Human megasatellite DNA RS447: copy-number polymorphisms and interspecies conservation. Genomics 54: 39–49

Hatzistamou J, Kiaris H, Ergazaki M, Spandidos DA (1996) Loss of heterozygosity and microsatellite instability in human atherosclerotic plaques. Biochem Biophys Res Commun 225:186–190

Heyer E, Puymirat J, Dieltjes P, Bakker E, Knijff P de (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6: 799–803

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–73

Jeffreys AJ, Royle NJ, Wilson V, Wong Z (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 332:278–281

Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JAL (1994) Complex gene conversion events in germline mutation at human minisatellites. Nat Genet 6:136–145

Jeffreys AJ, Bois P, Buard J, Collick A, Dubrova Y, Hollies CR, May CA, Murray J, Neil DL, Neumann R, Stead JD, Tamaki K, Yardley J (1997) Spontaneous and induced minisatellite instability. Electrophoresis 18:1501–1511

Kazazian HH Jr, Moran JV (1998) The impact of L1 retrotransposons of the human genome. Nat Genet 19:19–24

Kodama H, Saitoh H, Tone M, Kuhara S, Sakaki Y, Mizuno S (1987) Nucleotide sequences and unusual electrophoretic behavior of the W chromosome-specific repeating DNA units of the domestic fowl, *Gallus gallus domesticus*. Chromosoma 96:18–25

Kogi M, Fukushige S, Lefevre C, Hadano S, Ikeda J-E (1997) A novel tandem repeat sequence located on human chromosome 4p: isolation and characterization. Genomics 42:278–283

Laird CD (1990) Proposed genetic basis of Huntington's disease. Trends Genet 6:242–247

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol 39:174–190

Lupski JR (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 14:417–422

Maarel SM van der, Deidda G, Lemmers RJLF, Overveld PGM van, Wielen M van der, Hewitt JE, Sandkuijl L, Bakker B, Ommen G-JB van, Padberg GW, Frants RR (2000) De novo facioscapulohumeral muscular dystrophy: frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. Am J Hum Genet 66: 26–35

Mahtani MM, Willard HF (1993) A polymorphic X-linked tetranucleotide repeat locus displaying a high rate of new mutation: implications for mechanisms of mutation at short tandem repeat loci. Hum Mol Genet 2:431–437

May CA, Jeffreys AJ, Armour JAL (1996) Mutation rate heterogeneity and the generaton of allele diversity at the human minisatellite MS205 (D16S309). Hum Mol Genet 5:1823–1833

McNaught KS, Olanow CW, Halliwell B, Isacson O, Jenner P (2001) Failure of the ubiquitin-proteasome system in Parkinson's disease. Nat Rev Neurosci 2:589–594

Monckton DG, Neumann R, Guram T, Fretwell N, Tamaki K, MacLeod A, Jeffreys AJ (1994) Minisatellite mutation rate variation assosiated with a flanking DNA sequence polymorphism. Nat Genet 8:162–170

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616–1622

Nelson TN, Dill FJ, Wood S (1998) A megasatellite repeat may be involved in the formation of inversion duplication 8p. Am J Hum Genet 63 (Suppl):A40

Overveld PGM van, Lemmers RJLF, Deidda G, Sandkuijl L, Padberg GW, Frants RR, Maarel SM van der (2000) Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. Hum Mol Genet 9:2879–2884

Richards RI, Sutherland GR (1992) Dynamic mutations: a new class of mutations causing human disease. Cell 70:709–712

Rodriguez C, Cimaroli T, Sciorra L, Guzman E, Smulian J, Day-Salvatore D (1999) Prenatal diagnosis of duplication 4p mosaicism. Am J Hum Genet 65 (Suppl):A369

Sabl JF, Laird CD (1992) Epigene conversion: a proposal with implications for gene mapping in humans. Am J Hum Genet 50: 1171–1177

Saitoh Y, Miyamoto N, Okada T, Gondo Y, Showguchi-Miyata J, Hadano S, Ikeda J-E (2000) The RS447 human megasatellite tandem repetitive sequence encodes a novel deubiquitinating enzyme with a functional promoter. Genomics 67:291–300

Smith GP (1976) Evolutions of repeated DNA sequences by unequal crossover. Science 191:528–535

Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M (1966) Frameshift mutations and the genetic code. Cold Spring Harbor Symp Quant Biol 31:77–84

Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations. Affecting DNA mismatch repair. Nature 365:274–276

Takahashi Y, Mitani K, Kuwabara K, Hayashi T, Niwa M, Miyashita N, Moriwaki K, Kominami R (1994) Methylation imprinting was observed of mouse mo-2 macrosatellite on the pseudoautosomal region but not on chromosome 9. Chromosoma 103:450–458

Tartof (1988) Unequal crossing over then and now. Genetics 120: 1-6

Vergnaud G, Mariat D, Apiou F, Aurias A, Lathrop M, Lauthier V (1991) The use of synthetic tandem repeats to isolate new VNTR loci: cloning of a hman hyermutable sequence. Genomics 11: 135–144

Weber JL (1990) Informativeness of human $(dC\text{-}dA)_n(dG\text{-}dT)_n$ polymorphisms. Genomics 7:524–530

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second-generation linkage map of the human genome. Nature 359:794–801

Wooster R, Cleton-Jansen A-M, Collins N, Mangion J, Cornelis RS, Cooper CS, Gusterson BA, Ponder BAJ, Deimling A von, Wiestler OD, Cornelisse CJ, Devilee P, Stratton MR (1994) Instability of short tandem repeats (microsatellites) in human cancers. Nat Genet 6:152–156

Xu J, Freeman V, Nowaczyk MJM (1999) Direct duplication of 4p14p15.31. Am J Hum Genet 65 (Suppl):A363

Zhu X, Skinner J, Burgoyne LA (1990) The macrosatellites of the Toulouse goose: the major tandemly repetitive DNA in the Toulouse goose genome. Genome 33:641-645

Human Molecular Genetics, 2002, Vol. 11, No. 23 2815–2828

HIP14, a novel ankyrin domain-containing protein, links huntingtin to intracellular trafficking and endocytosis

Roshni R. Singaraja[1], Shinji Hadano[1,2], Martina Metzler[1], Scott Givan[1], Cheryl L. Wellington[1], Simon Warby[1], Anat Yanai[1], Claire-Anne Gutekunst[3], Blair R. Leavitt[1], Hong Yi[3], Keith Fichter[1], Lu Gan[1], Krista McCutcheon[1], Vikramjit Chopra[1], Jennifer Michel[1], Steven M. Hersch[3], Joh-E Ikeda[2,4] and Michael R. Hayden[1,*]

[1]Centre for Molecular Medicine and Therapeutics, Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada V5Z 4H4, [2]NeuroGenes, International Cooperative Research Project, Japan Science and Technology Corporation, Tokai University School of Medicine, Isehara, Kanagawa 259-1193, Japan, [3]Emory University School of Medicine, Atlanta, GA 30322, USA and [4]Department of Molecular Neuroscience, Molecular Medicine Research Center, The Institute of Medical Sciences, Tokai University, Isehara, Kanagawa 259-1193, Japan

Received June 4, 2002; Revised and Accepted August 26, 2002 DDBJ/EMBL/Genbank accession nos. AB024494, AB024495

Huntington disease (HD) is caused by polyglutamine [poly(Q)] expansion in the protein huntingtin (htt). Although the exact mechanism of disease progression remains to be elucidated, altered interactions of mutant htt with its protein partners could contribute to the disease. Using the yeast two-hybrid system, we have isolated a novel htt interacting protein, HIP14. HIP14's interaction with htt is inversely correlated to the poly(Q) length in htt. mRNAs of 9 and 6 bp are transcribed from the *HIP14* gene, with the 6 kb transcript being predominantly expressed in the brain. HIP14 protein is enriched in the brain, shows partial co-localization with htt in the striatum, and is found in medium spiny projection neurons, the subset of neurons affected in HD. HIP14 localizes to the Golgi, and to vesicles in the cytoplasm. The HIP14 protein has sequence similarity to Akr1p, a protein essential for endocytosis in *Saccharomyces cerevisiae*. Expression of human HIP14 results in rescue of the temperature-sensitive lethality in *akr1Δ* yeast cells and, furthermore, restores their defect in endocytosis, demonstrating a role for HIP14 in intracellular trafficking. Our findings suggest that decreased interaction between htt and HIP14 could contribute to the neuronal dysfunction in HD by perturbing normal intracellular transport pathways in neurons.

INTRODUCTION

Although polyglutamine [poly(Q)] expansion, the causative mutation in Huntington disease (HD) was discovered several years ago, there is as yet no proven mechanism for the selective vulnerability of neurons in the striatum and cortex observed in the disease. The protein huntingtin (htt) is ubiquitously expressed in all tissues. In neurons, htt is associated with synaptic vesicles and microtubules, and is enriched in dendrites and nerve terminals (1,2).

In order to understand the normal function of htt and mechanisms of pathogenesis in HD, and to expand the pathways of interaction with htt, we performed yeast two-hybrid screens to identify htt-binding proteins and determined whether poly(Q) length modulated these interactions. In the event that the interaction is modulated by CAG expansion, this would suggest that the interaction is obviously disturbed in the presence of the mutation and that this could therefore contribute to the pathogenesis of HD. In this regard, it is notable that certain classes of proteins cluster in their pattern of interaction with htt. For example, functions of normal htt can be gleaned from analysis of its protein partners. If these partners do not show altered interactions in the presence of the mutation, then these interactions are unlikely to underlie HD. Mutant htt does not show altered interaction with proteins such as HIP2 and p53, suggesting that the role of htt in cell cycle control and cell survival is part of its normal function, and does not contribute directly to the pathogenesis.

*To whom correspondence should be addressed at: Centre for Molecular Medicine and Therapeutics, University of British Columbia, 980 West 28th Avenue, Vancouver, Canada BC V5Z 4H4. Tel: +1 6048753535; Fax:+1 6048753819; Email: mrh@cmmt.ubc.ca

Seven proteins involved in transcriptional regulation interact with htt. These are the nuclear receptor co-repressor (3), the transcriptional co-repressor C-terminal binding protein (4), HYP-A (FBP-11), HYP-B(p231HBP) (5), cAMP response element binding protein (CREB)-binding protein (CBP) (6), mSin3a (7) and p300/CBP-associated factor(P/CAF) (7). Interestingly, almost all of htt's interacting protein partners implicated in transcriptional regulation show altered interaction with htt in the presence of the poly(Q) expansion, suggesting that disturbance of this interaction could contribute to the observed pathogenesis.

Furthermore, mutant htt shows altered interaction with every one of its protein partners implicated as having functions in neuronal intracellular transport, providing strong evidence that altered transport could contribute to the pathogenesis observed in HD. Evidence in support of this is that several interacting proteins, including htt-interacting protein 1 (HIP1) (8), htt-associated protein 1 (HAP1) (9–11), calmodulin (12), SH3GL3 (13) and cystathione β-synthase (14), have been identified with functions in transport and cytoskeletal organization. We describe here the identification and functional determination of the htt-interacting protein, HIP14, which also shows altered interaction with mutant htt and has a function in endocytosis. HIP14 is predominantly expressed in neurons in the brain but not in glia, and localizes to the Golgi complex and to cytoplasmic vesicles. HIP14 shares a high degree of amino acid similarity to proteins in a variety of organisms, including Akr1p, a protein essential for endocytosis of pheromone receptors in the yeast *Saccharomyces cerevisiae* (15). Yeast cells lacking functional AKR1 (*akr1*Δ) show decreased viability at an elevated temperature and display defects in endocytosis. We determined that HIP14 can rescue the temperature-sensitive viability of *akr1*Δ yeast cells and restore their block in endocytosis, suggesting that HIP14 may play a role in intracellular transport processes, and providing further proof for the role of htt in neuronal transport. This data leads to the hypothesis that a disturbed interaction between htt and HIP14 could lead to alterations in neuronal transport contributing to the pathogenesis in HD.

RESULTS

Isolation of *HIP14* cDNA clones

We performed yeast two-hybrid library screens to identify htt's protein partners. The yeast two-hybrid screen identified 14 positive clones (16). Of these, a single *HIP14* clone (pGAD*HIP14*-500) was found to specifically interact with the pGBT9–1955-44 fusion protein, as yeast co-transformants gave a His^+, β-galactosidase$^+$ phenotype, and the transformation of pGAD*HIP14*-500 with pGBT9 vector did not induce β-galactosidase activity.

We identified five cDNA clones encoding *HIP14* by screening human brain cDNA libraries with pGAD*HIP14*-500. Database searches identified several expressed sequence tag (EST) clones encoding regions of *HIP14*. DNA sequences of these cDNA clones were determined and aligned. The composite *HIP14* cDNA sequence encompasses 5272 nucleotides containing a single open reading frame (ORF) that is 1902 nucleotides long, which is predicted to encode a HIP14 protein consisting of 633 amino acids. A htt-interacting protein *HYPH*, which is identical to the 5′ portion of *HIP14*, has been reported (5), providing independent confirmation of the interaction of HIP14 and htt. The nucleotide sequence for *HIP14* has been deposited in the DDBJ/EMBL/GenBank database under accession no. AB024494.

Identification of *HIP14-related protein* (*HIP14L*)

Database searches have identified human EST clones encoding a gene homologous to *HIP14*. Of these, one clone contained a single ORF that is 1869 nucleotides long, encoding a predicted protein product of 622 amino acids, with 48% identity and 57% similarity to *HIP14* (Fig. 1 and Table 1). We designated this EST *HIP14-related protein* (*HIP14L*). The nucleotide sequence for *HIP14L* has been deposited in the DDBJ/EMBL/GenBank database under accession no. AB024495. PCR-based somatic cell hybrid mapping revealed that *HIP14L* mapped to human chromosome 11 (data not shown).

HIP14 maps to human chromosome 12q14–q15

In order to determine the chromosomal localization of HIP14, FISH analysis was performed, revealing that *HIP14* mapped to a single genomic locus at 12q14–q15. The GDB and OMIM database search revealed that no genetic disease had been mapped to this region (data not shown).

HIP14 and HIP14L show homology to the ankyrin repeat containing the *S. cerevisiae* protein Akr1p

HIP14 and HIP14L amino acid sequences were subjected to a homology search, which identified significant homology between HIP14, HIP14L, two *S. cerevisiae* ankyrin repeat-containing proteins (Akr1p, a protein essential for endocytosis of pheromone receptors and Akr2p), and other predicted transcripts of unknown function in *Caenorhabditis elegans* (accession no. AAD12801), *Drosophila melanogaster* (accession no. AAF49554), *Mortierella alpina* (accession no. CAB56510) and *Schizosaccharomyces pombe* (accession no. CAA90497). Multiple sequence alignments between HIP14, HIP14L, Akr1p, Akr2p, *C. elegans H32C10.a*, *D. melanogaster*, *M. alpina* and *S. pombe* have confirmed significant sequence similarity to each other (Table 1). Structural and motif analyses of these proteins revealed that each has a series of 33-residue ankyrin repeats and four to six putative transmembrane helices (Fig. 1). The relative positions of the ankyrin repeats and the putative transmembrane helices in these proteins were found to be well conserved, suggesting that these proteins may have functional or structural similarities.

Decreased interaction of HIP14 with mutant htt

To confirm the interaction of htt with HIP14, and to assess the influence of poly(Q) length on this interaction, we performed β-galactosidase liquid assays (Fig. 2A). Five isolated colonies from each of three separate transformations of htt constructs pGBT9–1955-15 or pGBT9–1955-128, or pGBT9 as control, were individually assayed for their strength of interaction with

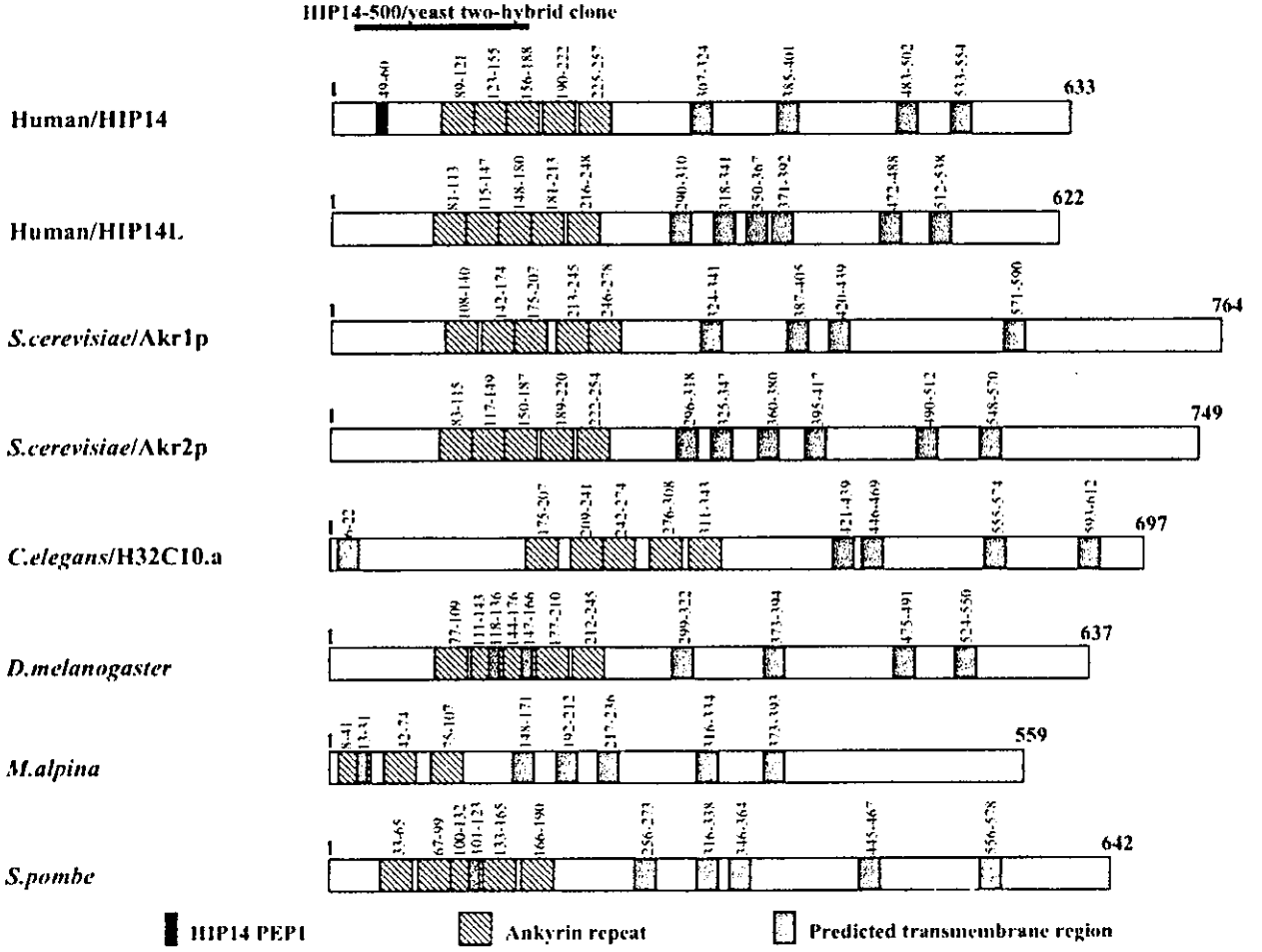


Figure 1. Schematic representation of the protein structure of human HIP14 (hHIP14), human HIP14L (hHIP14L), *Saccharomyces cerevisiae* AKR1, *S. cerevisiae* AKR2, and the related proteins in *Caenorhabditis elegans, Drosophila melanogaster, Mortierella alpina* and *Schizosaccharomyces pombe.* The positions of the HIP14PEP1 peptide sequence, ankyrin repeats, and predicted transmembrane helices are shown.

pGAD–HIP14 (Fig. 2A; pGBT9; 1.91 ± 0.05 SEM; $n = 15$, pGBT9–1955-15; 29.00 ± 0.48 SEM; $n = 15$, pGBT9–1955-128; 17.93 ± 0.82 SEM; $n = 15$). Poly(Q) expansion in htt resulted in a significant decrease in its interaction with HIP14 ($P < 0.001$). Western blot analysis of yeast cells transformed with pGAD–HIP14-500, and either pGBT9–1955-15 or pGBT9–1955-128 showed equivalent protein expression levels (data not shown), indicating that the decreased interaction between HIP14 and expanded htt in yeast was not attributable to differences in translation or to altered stability of htt with different poly(Q) lengths.

Table 1. Pairwise comparisons of the amino acid sequences of HIP14, HIP14RP, AKR1, AKR2 and similar proteins predicted to be encoded by genes in *Caenorhabditis elegans, Drosophila melanogaster, Mortierella alpina* and *Schizosaccharomyces pombe* using GAP (GCG)

	hHIP14	hHIP14L	Akr1	Akr2	*C. elegans*	*D. melanogaster*	*M. alpina*	*S. pombe*
								Identities
hHIP14	—	48	30	28	27	48	34	31
hHIP14L	57	—	30	27	26	38	32	28
Akr1	41	38	—	41	27	29	33	30
Akr2	38	37	51	—	25	29	30	28
C. elegans	38	35	38	37	—	27	26	29
D. melanogaster	57	49	39	38	36	—	32	28
M. alpina	45	41	41	40	37	42	—	33
S. pombe	42	39	43	39	39	40	43	—
	Similarities							

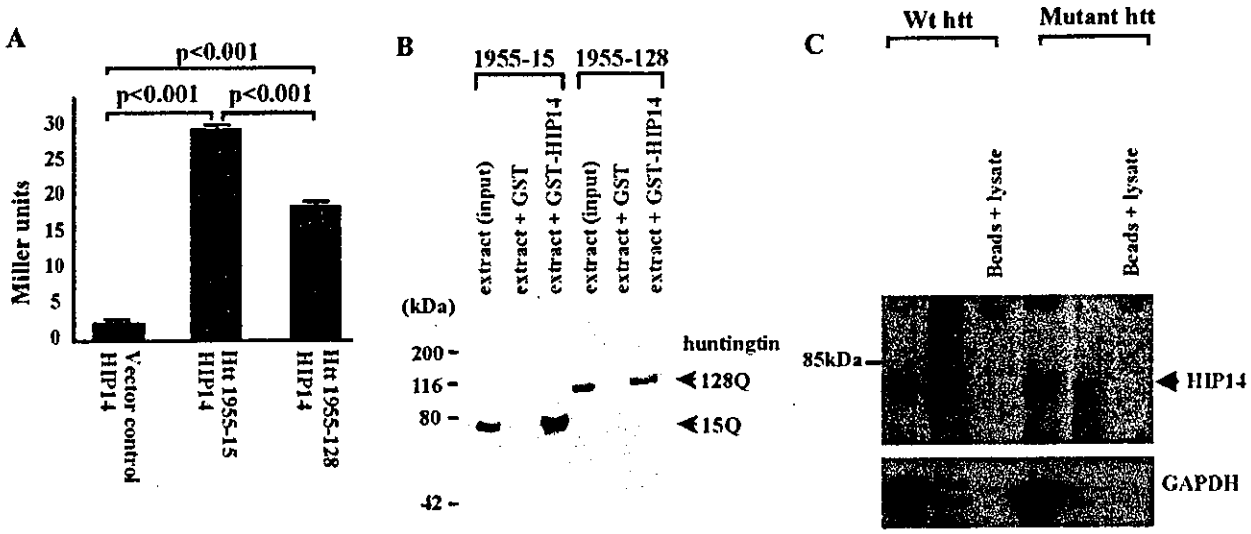


Figure 2. (A) Liquid β-galactosidase assays of the interaction between htt and HIP14. GAL4–BD htt fusion proteins with 15 and 128 CAG repeats (1955-15 and 1955-128) and pGBT9 vector control were co-transformed with pGAD–*HIP14*-500. Each bar represents the pooled results from 15 independently analyzed colonies from three independent transformations. (B) *In vitro* binding of GST–*HIP14* to the N terminus of htt in extracts prepared from human embryonic kidney cells (HEK 293T) transfected with htt 1955-15 and 1955-128. The htt fragments were immunodetected with BKP1 antibody. The htt fragments in the extracts were bound specifically to GST–HIP14 but not to GST alone. (C) Co-immunoprecipitation of HIP14 with htt *in vivo*. Cells transfected with wild-type (pc3-1955-15) htt immunoprecipitated more HIP14 than those cells transfected with mutant htt (pc3-1995-128) ($n=2$). htt was immunoprecipitated with the anti-htt 2166 antibody, and the immunoblotting was performed with anti-HIP14 antibody.

To confirm the dependence of the HIP14–htt interaction on poly(Q) length in htt, *in vitro* binding assays were performed using immobilized HIP14–GST fusion protein and extracts from HEK293T cells overexpressing the N-terminal region of htt containing either 15 or 128 poly(Q) residues. GST–HIP14 bound to glutathione sepharose beads consistently precipitated more wild-type than mutant htt ($n=2$). Neither wild-type nor mutant huntingtin was precipitated with GST alone (Fig. 2B). This finding confirms that the physical interaction between HIP14 and htt decreases as poly(Q) length increases.

In order to assess whether htt and HIP14 interact *in vivo*, we performed co-immunoprecipitation experiments from cells overexpressing both htt and HIP14. We found that lysates from cells expressing HIP14 and wild-type htt (pCI1955-15) immunoprecipitated more HIP14 ($n=2$) than lysates from cells containing HIP14 and mutant htt (pCI1955-128), proving that HIP14 and htt interact *in vivo*, and that this interaction is modulated by the poly(Q) tract length in htt.

HIP14 is not toxic in 293T cells

The interaction between HIP14 and mutant htt has been shown to be reduced, which could lead to excess unbound HIP14 in patient neurons generating toxicity. A similar finding has been shown for HIP1, with toxicity being mediated through its death effector domain (DED) (17,18). We studied the effect of overexpression of HIP14 in transiently transfected 293T cells. Measurement of toxicity in these cells using the modified MTT assay showed that, when overexpressed, HIP14 was not significantly more toxic ($P>0.05$) than a control LacZ-expressing vector (data not shown).

HIP14 generates two transcripts in various tissues

To determine expression of the *HIP14* transcripts, northern blot analysis was performed (Fig. 3), revealing the presence of two transcripts of ~9 and 6 kb in various tissues. Both transcripts show a similar pattern of expression, although the 6 kb transcript showed much higher expression when compared with the 9 kb transcript. Both transcripts were expressed in all tissues examined, with the highest expression of the 6 kb transcript occurring in the brain.

HIP14 is expressed at the highest level in the brain

Western blot analysis on human tissue detected a single immunoreactive band at 70 kDa in all brain regions. Preincubation of the anti-HIP14 antibody with the antigenic peptide completely eliminated the band (data not shown). HIP14 expression is highest in the cortex, cerebellum, occipital lobe and caudate, and lowest in the spinal cord (Fig. 4A). In human peripheral tissues, HIP14 protein was detected in testis, pancreas, heart and kidney, with no expression detected in the liver and lung (Fig. 4B). In mouse tissues, HIP14 was detected in heart, pancreas, kidney, testis, liver and brain, with the highest expression in the brain, and none in the lung (Fig. 4C). Therefore the pattern of HIP14 expression is similar in mouse and human tissues.

Since HD is characterized by cell death in the striatum and cortex, we further characterized the localization of HIP14 in

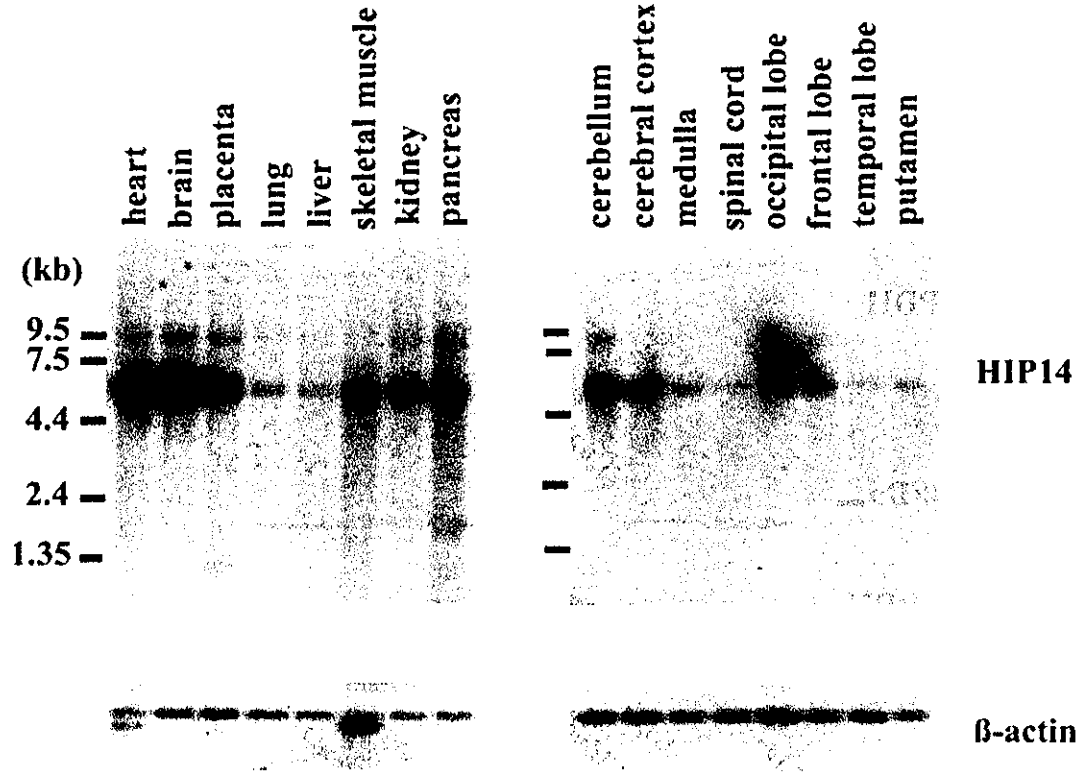


Figure 3. Northern blot analysis of human *HIP14* mRNA. Northern blots containing 2 µg of poly(A)$^+$ mRNA from various adult human tissues were hybridized with cDNA clone *HIP14*-500. The lower panel represents the same blot hybridized with human β-actin cDNA to confirm RNA quality and relative loading. RNA size markers are shown on the left. *HIP14* encodes a major transcript of 6 kb and a minor one of 9 kb. The 6 kb transcript levels are highest in the brain, with the cerebellum showing the highest levels.

mouse brain slices. Immunohistochemical analysis of HIP14 combined with neuronal and glial markers revealed that HIP14 localized to the perinuclear region of neurons from the cortex (Fig. 5J), striatum (Fig. 5K) and hippocampus (Fig. 5L), but not glial cells (data not shown). No HIP14 immunoreaction was detected in the nuclei. HIP14 was also found in a variety of other brain regions including cerebellum, brainstem and thalamus. No staining was observed in control sections when primary antibody was omitted or in the presence of blocking antigen (data not shown).

htt and HIP14 co-localize in medium spiny neurons of the striatum

To determine the localization of HIP14 and its co-localization with htt in cells of the striatum, we performed co-staining of htt and HIP14 in mouse brain sections. As shown in Figure 6, htt (Fig. 6A, green) and HIP14 (Fig. 6B, red) showed partial overlapping staining (Fig. 6C, yellow) in cytoplasmic and perinuclear regions of the striatal cells. These abundant and medium sized cells in the striatum had the morphology of medium sized spiny neurons. A low-powered image of HIP14 and htt staining in the striatum is shown in Figure 6D, also displaying partial overlap in the staining of HIP14 and htt (yellow).

HIP14 is found in projection neurons affected in HD

The specific subset of neurons that are selectively vulnerable to neurodegeneration in HD are the medium spiny neurons in the striatum that project to the medial and lateral globus pallidus (19). To determine whether HIP14 is present in the striato-pallidal neurons that took up the marker in the globus pallidus and transported it back to the striatum, we injected fluorogold tracer into the globus pallidus of mice. As seen in Figure 6E, HIP14 (red) is present in the specific subset of striatal neurons that project to the globus pallidus and uptake the fluorogold tracer (blue), conclusively indicating that HIP14 is found in the specific subset of neurons most affected in HD. Fluorogold injections were confined to the globus pallidus, as confirmed by anatomical assessment of the injection site in the immuno-histochemistry slices. The fluorogold injection sites did not overlap with the striatal marker DARPP-32 (Chemicon) (data not shown).

HIP14 localizes to the Golgi and to cytoplasmic vesicles in neuronal precursor NT2 cells and ES cell-derived neurons

In an effort to elucidate the subcellular localization of HIP14, we assessed its immunolocalization in neuronal NT2 cells and

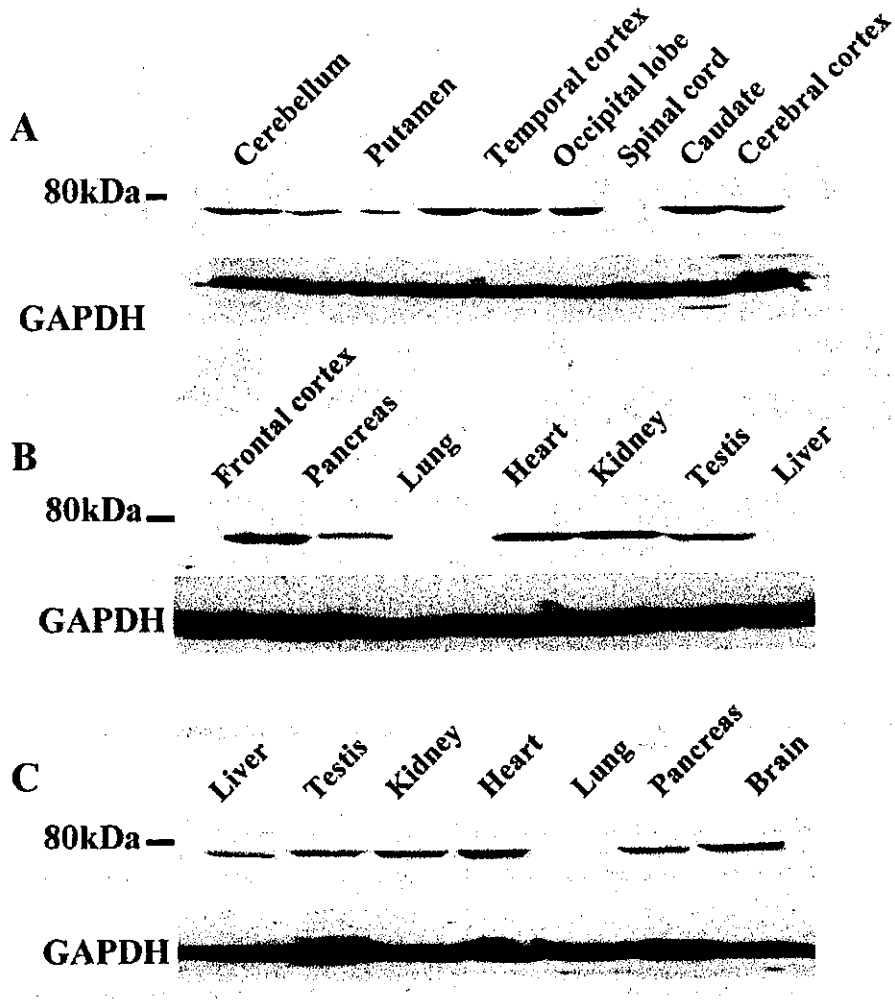


Figure 4. HIP14 protein expression in (A), normal human brain regions, (B), normal human peripheral tissues and (C), mouse tissues. All western blot analyses were carried out with HIP14PEP1 polyclonal antibody, and equivalent protein loading was assessed by GAPDH immunostaining. HIP14 levels are highest in the cortex, caudate, cerebellum and occipital lobe in human tissue. In human peripheral tissues, HIP14 levels are highest in the brain. In mice, the distribution of HIP14 in peripheral tissue is essentially similar, with the highest levels being observed in the brain.

ES cell-derived neurons. HIP14 immunoreactivity was localized to a perinuclear Golgi-like structure and to cytoplasmic vesicles, as analyzed by conventional and laser confocal microscopy. No overlap was found between HIP14 and the ER markers calreticulin and Grp78. By confocal microscopy, HIP14 showed partial co-localization with the *cis*-Golgi marker GM130 (Fig. 5A–C), which is involved in structural maintenance of the Golgi stack and interacts with the vesicle docking protein p115 (20). In addition, HIP14 was found to partially co-localize with the vesicle markers γ-adaptin, a component of the AP-1 complex (Fig. 5D–F), and clathrin (Fig. 5G–I) in the Golgi region. HIP14 also showed overlapping staining with the Golgi marker β-COP (data not shown). Biochemical fractionation of the human cortex, cerebellum and caudate using differential centrifugation revealed that HIP14 was highly enriched in the membrane (P3) fraction (data not shown) similar to the distribution of htt and HIP1 (21). These immunolocalization studies together with structural analysis suggested that HIP14 was associated with Golgi and cytoplasmic vesicle membranes.

To conclusively determine the localization of HIP14 within the Golgi apparatus, we performed pre-embedding immunogold labeling on mouse brain tissue sections using the anti-HIP14 antibody. At the electron-microscopic level, most of the immunogold particles were associated with the Golgi apparatus (Fig. 7A). Gold particles were also found in coated vesicles in the cytoplasm. Occasional staining was associated with the endoplasmic reticulum. Within the Golgi apparatus, particles were associated with the medial cisternae and the *cis*-Golgi cisternae (Fig. 7B). Particles were also associated with other Golgi compartments including *trans* cisternae, TGN and the pre-Golgi network.

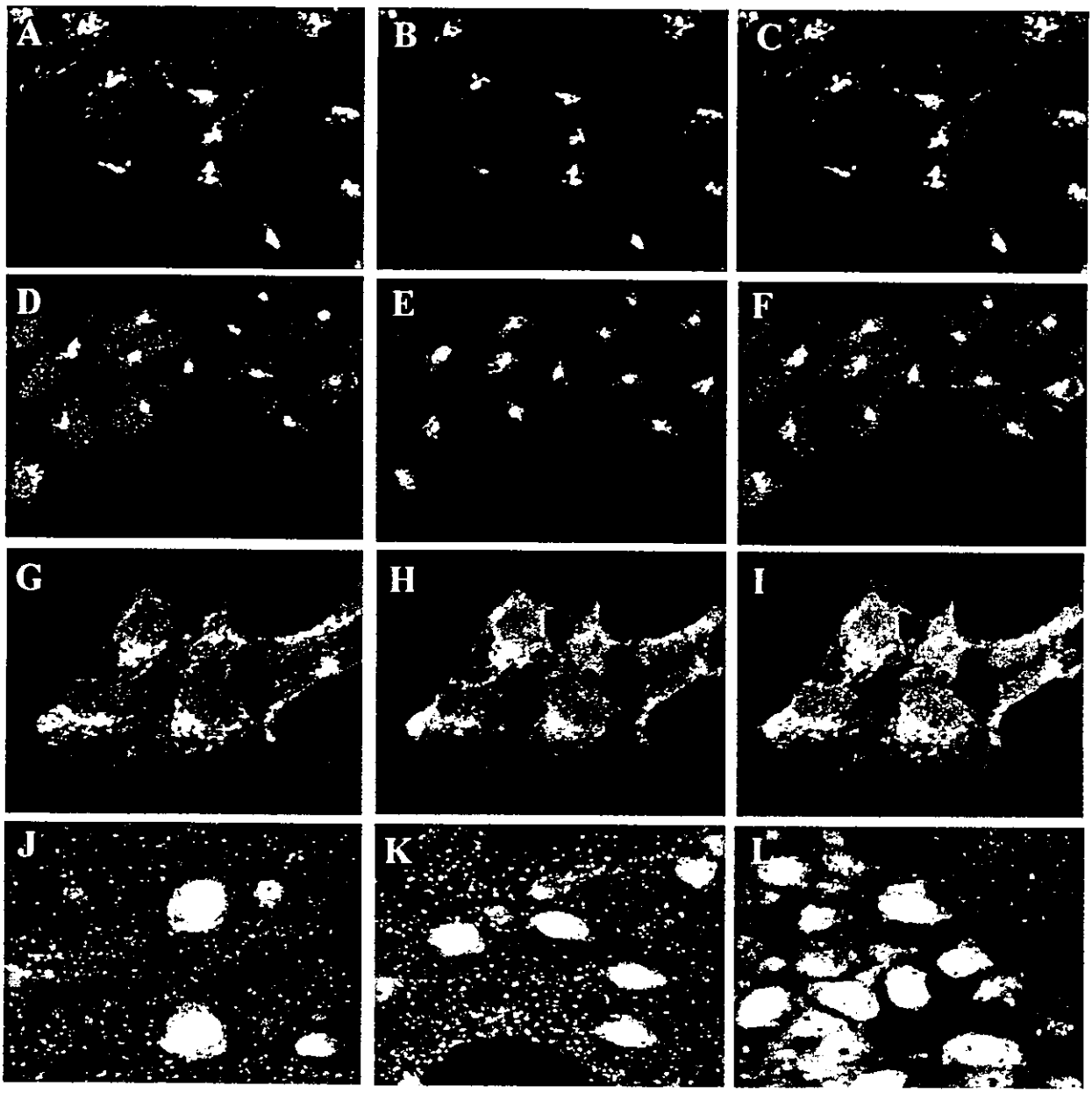


Figure 5. Immunolocalization of HIP14. (A–I) A representative neuronal precursor NT2 cell is shown by confocal microscopy. HIP14 was visualized using HIP14PEP1 antibody followed by incubation with an Alexa 488-labeled secondary antibody A, D and G, green. Expression of the *cis*-Golgi marker GM130 (B, red), and the vesicle markers γ-adaptin (E, red) and clathrin (H, red) were detected with an Alexa 568-conjugated secondary antibody. Images were captured with a CCD camera, overlaid and merged electronically (C, F and I). (J–L) Immunolocalization of HIP14 (red) with nuclear staining with neu-N mAb (green) in the mouse cortex (J), striatum (K), and hippocampus (L).

HIP14 rescues the temperature-sensitive phenotype and the defects in endocytosis of *akr1*Δ cells

In *S. cerevisiae*, it has been demonstrated that Akr1p is necessary for cellular viability at elevated temperatures. Considering that there is a 41% amino acid similarity between HIP14 and Akr1p, we explored whether HIP14 could rescue the temperature-dependant lethality of *akr1*Δcells. *AKR1* and *akr1*Δ yeast cells were transformed with the pAD5 vector alone, pAD5–HIP14, or the AKR1-containing positive control plasmid pPB575 (2μ, AKR1, LEU2) (22), streaked onto SD–leu plates and incubated at 25°C (permissive temperature for *akr1*Δ cells), and 37°C (restrictive temperature for *akr1*Δ cells). As can be seen in Figure 8A, *akr1*Δ cells carrying pPB575 grew well at both temperatures, whereas cells carrying pAD5 were viable only at 25°C. *akr1*Δ cells carrying pAD5–HIP14

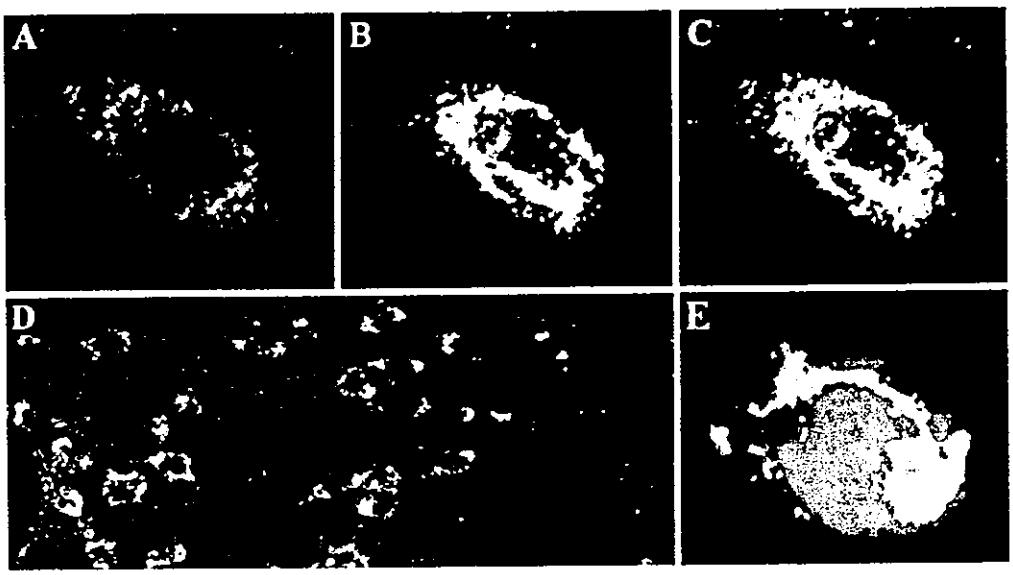


Figure 6. Subcellular localization of htt (A, green) and HIP14 (B, red) in medium spiny neurons of the striatum. Partial overlap of htt and HIP14 (C, yellow) is observed. A low-powered image of the co-stained striatum (D) is shown. HIP14 is localized in the medium spiny neurons that project into the globus pallidus (E) and are specifically affected in HD. Fluorogold was injected into the globus pallidus, taken up and transported, and it is present in the striatum (blue) along with HIP14 (red) in the subset of striatopallidal neurons.

grew well at 25°C, and were able to grow more vigorously at 37°C than cells carrying pAD5.

We determined whether HIP14 could specifically rescue the block in constitutive endocytosis of the yeast pheromone receptor Ste3p seen in *akr1*Δ cells. Akr1p is required for the constitutive endocytosis of Ste3p (15). Ste3p is normally targeted to the plasma membrane, then endocytosed and transported through vesicular intermediates to the vacuole, where it is degraded. In wild-type cells, Ste3p has a half-life of ~15 minutes, whereas in *akr1*Δ cells this half-life is increased >5-fold (15). The longer half-life of Ste3p in *akr1*Δ cells was shown to result from a block in constitutive endocytosis (15). We transformed *akr1*Δ cells with pAD5, pAD5–HIP14 or pPB575, subjected these strains to a cycloheximide-induced protein synthesis block, and removed samples at various time-points. Using a Ste3p antibody, we determined that Ste3p degradation is restored upon the expression of HIP14 in yeast, as can be seen in Figure 8B, demonstrating that HIP14 can functionally rescue the endocytosis defect in *akr1*Δ cells.

DISCUSSION

In this study, we have identified and characterized a novel htt-interacting protein, HIP14, that shows decreased interaction with mutant htt. A 542 bp fragment of HIP14 had been previously described (HYPH) (5) as an htt interacting protein, providing independent confirmation of this interaction. htt is widely expressed throughout the central nervous system (23,24). Unlike htt, which has ubiquitous expression, HIP14 is localized predominantly in the brain of both humans and mice. In the brain, it is localized, among other regions, in the medium spiny neurons of the striatum, the site of early and prominent pathology in HD.

HIP14 and htt show partial co-localization in the medium spiny neurons of the striatum, and, in addition, they are found in medium spiny neurons that project into the globus pallidus, which is the subset of neurons selectively vulnerable to neurodegeneration in HD. Furthermore, HIP14 is localized in the Golgi apparatus and in discrete cytoplasmic vesicles. Immunofluorescence microscopy has revealed that htt is also enriched on membranes of the Golgi, and associates with clathrin-coated and non-coated vesicles in the cytoplasm and along plasma membranes (25). The specific cellular and subcellular colocalization of HIP14 and htt provides further support for the functional interaction between these two proteins in living cells.

Nothing is known about the function of HIP14. However, clues to its normal function can be derived from the insights gained from the study of the gene in other organisms. The HIP14 protein contains several transmembrane and ankyrin repeat domains, and shares significant homology with the yeast protein Akr1, which, when deleted, causes defects in endocytosis in yeast (15). Interestingly, expression of HIP14 protein in *akr1*Δ yeast cells restores their endocytic function. This functional complementation by HIP14, along with its strong sequence similarity to Akr1p, provides evidence that Akr1p is the yeast homologue for HIP14 and gives clues to the role of HIP14 in intracellular protein trafficking. Akr1p has been recently shown to play a role in the targeting of type I casein kinases to the plasma membrane, where they are involved in the ubiquitin-mediated endocytosis of yeast pheromone receptors (26). Ubiquitin plays a central role in early endocytic events in yeast by providing a sorting determinant that initiates uptake. In addition, it has been