Downloaded from genome.cshlp.org on June 22, 2026 . Published by Cold Spring Harbor Laboratory Press

Isolation and mapping of novel mouse brain cDNA clones containing trinucleotide repeats, and demonstration of novel alleles in recombinant inbred strains.

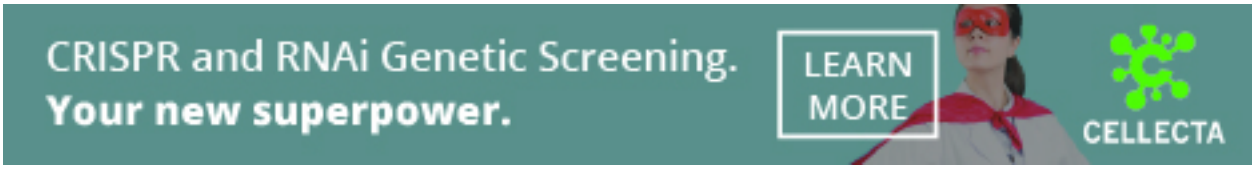
D M Chambers and C M Abbott

Genome Res. 1996 6: 715-723
Access the most recent version at doi:10.1101/gr.6.8.715

References This article cites 19 articles, 1 of which can be accessed free at:
http://genome.cshlp.org/content/6/8/715.full.html#ref-list-1

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**



To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Copyright © Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on June 22, 2026 . Published by Cold Spring Harbor Laboratory Press

RESEARCH

Isolation and Mapping of Novel Mouse Brain cDNA Clones Containing Trinucleotide Repeats, and Demonstration of Novel Alleles in Recombinant Inbred Strains

Doreen M. Chambers and Catherine M. Abbott[1]

Medical Research Council (MRC) Human Genetics Unit, Western General Hospital, Edinburgh, UK

Abnormal expansion of trinucleotide repeats (TRs) has now been implicated in the pathogenesis of at least nine human genetic disorders, particularly those in which anticipation and/or fragile sites have been demonstrated. Anticipation, the phenomenon of increasing severity of phenotype in successive generations, has never been seen in species other than man. Nevertheless, animal models for the dynamic mutation of TRs would be extremely valuable. We have screened a mouse brain cDNA library in an attempt to identify clones representing each of the 10 possible classes of trinucleotide repeat. Thirty-seven clones were analyzed in detail. Of the 37 sequences, 18 displayed significant levels of homology with sequences in GenBank, 10 of them with human expressed sequence tags (ESTs). We then analyzed 25 of the clones by PCR of the sequence containing the repeat in a number of different mouse strains and species to assess levels of variability of repeat length. Of the 25 clones analyzed in this way, 64% showed length variation between *Mus musculus* spp. and *Mus spretus*, and 32% showed variation between *Mus musculus musculus*-derived standard laboratory inbred strains. Where variation was detected (17 repeat-containing clones in all), the gene was mapped by linkage analysis. None of the repeats isolated showed any signs of extreme expansion. However, two of the repeats were shown to have undergone size changes during the establishment of a number of recombinant inbred strains, suggesting that these repeats are at least moderately unstable.

Abnormal expansion of trinucleotide repeats (TRs) is a relatively recently discovered phenomenon which has now been implicated in the pathogenesis of at least nine human genetic disorders, notably those in which anticipation and/or fragile sites have been identified. These include spinal bulbar muscular atrophy (SBMA), fragile X syndrome (FRAXA), myotonic dystrophy (DM), spinocerebellar ataxia (SCA1), Huntington's disease (HD), FRAXE syndrome, dentatorubral pallidoluysian atrophy (DRPLA), Machado-Joseph disease (MJD), and FRA11B (for review, see Ashley and Warren 1995). A recent addition to this list is Friedreich's ataxia (FRDA), which has been shown to result, in the majority of cases, from the expansion of a $(GAA)_n$ triplet repeat in the first intron of the X25 gene (Campuzano et al. 1996). This disorder is recessive and anticipation has never been seen. Surveys of human cDNA libraries and sequence data bases have been carried out to identify additional genes containing TRs (Riggins et al. 1992; Li et al. 1993). Many of these repeat sequences have been shown to be highly polymorphic in terms of repeat length and therefore might repesent candidates for other diseases in which anticipation has been demonstrated. The DRPLA gene was identified initially in such a survey (Li et al. 1993). The phenomenon of anticipation has never been seen in species other than man. Nevertheless, animal models for these diseases, and particularly for the process by which long TRs undergo abnormal expansion, would be extremely valuable. No spontaneous mouse models for any of the diseases listed above exist, but in each case where the homologous mouse gene has been sequenced, the TR exists only in a short, interrupted form, and would be unlikely to undergo expansion. Mouse models have been created for FRAXA by knocking out the mouse *Fraxa* gene (Bakker et al. 1994), but attempts to model SBMA and SCA1 by making transgenic mice carrying a human transgene containing an expanded CAG repeat have failed

[1]Corresponding author.
E-MAIL cma@srv0.med.ed.ac.uk; FAX 0131-537-1059.

Downloaded from genome.cshlp.org on June 22, 2026 . Published by Cold Spring Harbor Laboratory Press

to recreate instability of the repeat tract (Bingham et al. 1995; Burright et al. 1995).

We set out to survey mouse genes containing TRS for two purposes: (1) To examine the frequencies of different classes of repeat in transcribed regions of the mouse genome; and (2) to see if any of the genes isolated might be candidates for existing mouse mutants, potentially through expansion of the TR. We present the results obtained from an initial survey of a mouse brain cDNA library with oligonucleotides corresponding to all possible classes of TR. All of the repeats showing length variation between different inbred strains of *Mus musculus* subspecies, or between *M. musculus* subspecies and *Mus spretus,* were mapped and the position of the gene containing the repeat was aligned with consensus maps of mouse chromosomes to identify possible candidate mutations. Some of the repeats were analyzed in recombinant inbred (RI) strains, to establish whether the repeat length seen in the parental strains had been stably inherited during the establishment of the RI lines.

RESULTS

Distribution of Repeat Classes

A mouse brain cDNA library in λgt11 was screened with 15-mer oligonucleotides corresponding to all possible classes of TR. In two cases, alternative oligonucleotides were used that had different sequences designed to detect the same class of repeat, that is, $(CAT)_5$ and $(GAT)_5$ for ATC repeats, and $(CAG)_5$ and $(CTG)_5$ for AGC repeats. In each case, the same positive plaques were identified with the alternative oligonucleotide. The number of positively-hybridizing plaques for each class of repeat per 2000 plaques screened is as follows: $(AAT)_n$ 40, $(GTT)_n$ 60, $(CCA)_n$ 10, $(ACT)_n$ 0, $(CGA)_n$ 0, $(AGA)_n$ 15, $(CAG)_n$ 6, $(TCC)_n$ 40, $(CAT)_n$ 10, $(CCG)_n$ 0.

Similar numbers of positively hybridizing plaques were obtained when a mouse testis cDNA library was screened (data not shown). The distribution of different classes of repeats was clearly nonrandom, as would be expected from the known frequencies of particular nucleotide combinations, and from data base surveys of TRs (see Discussion). Most of the positive plaques were isolated for further analysis, and 37 were subcloned into plasmids and sequenced. Hybridization of selected clones back to others picked with the same repeat oligo revealed that in almost all cases the clones were unique within the subset, suggesting that many more repeat-containing clones remained undetected in the library. Two of the 37 clones, however, proved to contain overlapping sequence. The sequence of individual repeats is shown in Table 1. Although many were perfect, in some cases, notably for TCC and CAT repeats, the repeat sequence was complex and interrupted. The longest perfect repeat was a tract of $(TTC)_{34}$; a second clone contained $(AAG)_{27}$ adjacent to $(AGG)_{30}$. A number of clones contained more than one class of TR. This clustering has been observed previously (Stallings 1994; Braga et al. 1995)

Homologies with Other Sequences

Once an adequate amount of sequence information had been obtained, the data were analyzed using the BLAST program of the GCG package to check for repetitive sequences such as B1 repeats, and to compare sequence with known genes. Of the 37 clones, 18 displayed significant homology with gene sequences present in the GenBank data base. These homologies are detailed in Table 2. In 10 cases, homology was detected with human anonymous cDNAs or expressed sequence tags (ESTs). The other eight clones showed homology to genes of known function, namely mitochondrial malate dehydrogenase, nucleolar phosphoprotein B23, AT-BP1 (a zinc finger gene), glutamate decarboxylase GAD65, the extreme 3′ end of the mouse *Hnf1* gene, Ca^{2+}-ATPase isoform 2, the mouse homolog of the yeast *RAD23* gene and the mitochondrial 14-3-3 ζ gene. At least seven clones contained repetitive elements, such as B1 repeats.

Variability of Repeat Length

Wherever possible, PCR primers were designed flanking the TR to enable the size of the repeat to be estimated in different strains and species of mouse. In all, 25 TR sequences were successfully analyzed by PCR. The primer sequences and conditions used in each case are given in Table 3. A number of TRs could not be analyzed in this way, for several reasons, summarized in Table 1. This was particularly unfortunate in the case of clones TCC2, which contained the repeat sequence $(AAG)_{27}$ adjacent to $(AGG)_{30}$ and clone AGA6, which was found to contain 34 TTC repeats.

Downloaded from genome.cshlp.org on June 22, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 1. Sequences of Mouse Trinucleotide Repeats and Variability of Repeat Length in Different Mouse Strains

Clone name	Repeat sequence	PCR analysis[a]		Degree of variability[b]	No. of alleles[c]
AAT2	$(ATT)_{10}$	no	a		
AAT3	$(AAT)_{13}$	no	a		
AAT5	$(AAT)_{15}$	yes		+++++	5
AAT18	$(TTA)_6$	no	a		
AGA6	$(TTC)_{34}$	no	b		
AGA7	$(GAA)_{11}$	no	b		
AGA9	$(CTT)_{17}(CTC)_1(CTT)_3$	no	b		
CAG2	$(CAG)_{10}$	yes		+++++	7
CAG3	$(CAG)_6$	yes		+++++	4
CAG14	$(CCA)_5(CACTCA)_1(GCT)_{10}(TCT)_7$	no	b		
CAG17	$(CAG)_5$	yes		-	
CAG23	$(AGC)_{10}$	yes		+++++	6
CAT2	$(GAT)_6$	yes		-	
CAT6	$(GAT)_4(GAA)_1(GATGAC)_2$	yes		-	
CAT9	$(TCA)_4(TCG)_1(TCA)_2(TCG)_1(TCA)_3(TCG)_1$	yes		-	
CAT10	$(TCA)_6$	yes		+++	5
CAT11	$(CAT)_4(CGT)_1(CAT)_2(CGT)_1(CAT)_3(CGT)_1$	yes	d	- (3 loci)	
CAT19	$(CAT)_1(TAT)_1(CAT)_3$	no	b		
CCA3	$(CCT)_9(TCT)_1(CCT)_4(CCA)_4$	no	a		
CCA5	$(GGT)_8$	yes		+++++	6
AAT2	$(ATT)_{10}$	no	a		
AAT3	$(AAT)_{13}$	no	a		
AAT5	$(AAT)_{15}$	yes		+++++	5
AAT18	$(TTA)_6$	no	a		
AGA6	$(TTC)_{34}$	no	b		
AGA7	$(GAA)_{11}$	no	b		
AGA9	$(CTT)_{17}(CTC)_1(CTT)_3$	no	b		
CAG2	$(CAG)_{10}$	yes		+++++	7
CAG3	$(CAG)_6$	yes		+++++	4
CAG14	$(CCA)_5(CACTCA)_1(GCT)_{10}(TCT)_7$	no	b		
CAG17	$(CAG)_5$	yes		-	
CAG23	$(AGC)_{10}$	yes		+++++	6
CAT2	$(GAT)_6$	yes		-	
CAT6	$(GAT)_4(GAA)_1(GATGAC)_2$	yes		-	
CAT9	$(TCA)_4(TCG)_1(TCA)_2(TCG)_1(TCA)_3(TCG)_1$	yes		-	
CAT10	$(TCA)_6$	yes		+++	5
CAT11	$(CAT)_4(CGT)_1(CAT)_2(CGT)_1(CAT)_3(CGT)_1$	yes	d	- (3 loci)	
CAT19	$(CAT)_1(TAT)_1(CAT)_3$	no	b		
CCA3	$(CCT)_9(TCT)_1(CCT)_4(CCA)_4$	no	a		
CCA5	$(GGT)_8$	yes		+++++	6

Shown is the sequence of each individual trinucleotide repeat, together with an indication of the degree of variability of repeat length in each case.

Clone GTT7 is included with loci containing AGA repeats, as—although it was originally isolated with a (GTT)5 oligonucleotide—it also was found to contain an AGA repeat, and this was the repeat analyzed.

[a]The possibility of PCR analysis across the repeat region. Reasons why not possible: (a) the PCR reaction failed under all conditions tried; (b) the repeat was adjacent to the polylinker in the plasmid; (c) the repeat was embedded in a repetitive element such as a B1 repeat; (d) The PCR was effective; but more than one locus was amplified (for CAT 11, please see text; GTT 25 and CAG 3 primers amplified two bands, but only the band of the predicted size showed length variation).

[b]Variability is ranked according to the following criteria: (+++++) repeat length varies among *M.m. musculus*-derived strains of mouse; (++++) repeat length varies between *M.m. musculus*-derived strains and *M.m. castaneus* or *M.m. molossinus;* (+++) repeat length varies between *M.m. musculus*-derived strains and *M. spretus;* (++) repeat length varies between *M.m. musculus*-derived strains and *M. spretus* but variation is only detectable on sequencing gels; (+) repeat length varies between *M.m. musculus*-derived strains and *M. caroli;* no variation was detected.

[c]The total number of alleles seen is given.

Downloaded from genome.cshlp.org on June 22, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 2. Homologies and Map Positions of Mouse Trinucleotide Repeats

Clone name	Homology	Map location (locus symbol)[a]	Flanking markers	Mapping panel
AAT2	-			
AAT3	B1 repeat			
AAT5	-	Chr 14 (*D14Abb1e*)	*D14Mit2, Chat*	JAX BSS
AAT18	8 human ESTs (R34771)			
AGA6	MT repeat element			
AGA7	novel low copy repeat			
AGA9	-			
CAG2	human EST (T84949)	Chr 13 (*D13Abb1e*)	*D13Mit1, Plf*	JAX BSS
CAG3	mitochondrial malate dehydrogenase	Chr 5 (*Mor1*)	*D5Mit63, D5Mit62*	JAX BSB
CAG14	-			
CAG17	7 human ESTs (R59012), I mouse EST (R74808)			
CAG23	2 human ESTs (H19097)	Chr5 (*D5Abb2e*)	*Tcf1, Fla*	BXD, BXH
CAT2	nucleolar phosphoprotein B23			
CAT6	zinc finger protein AT-BP1			
CAT9	glutamate decarboxylase GAD65			
CAT10	-	Chr 6 (*D6Abb1e*)	*D6Bir12, D6Mit8*	JAX BSS
CAT11	10 human ESTs (H90077); TFIID components	see text		
CAT19	-			
CCA3	extreme 3' end of Hnf1; 18 human ESTs (H98215)			
CCA5	4 human ESTs (T17004)	Chr 18 (*D18Abb1e*)	*D18Bir2, Xltr2*	JAX BSS
GTT1	-	Chr 18 (*D18Abb2e*)	*D18Mit9, D18Mit7*	JAX BSS
GTT3	calcium ATPase isoform 2	Chr 6 (*Atp2b2*)	*Il5r, Rad52*	JAX BSS
GTT8	B1 repeat			
GTT9	MHR23A (mouse homologue of RAD23)	Chr8 (*Rad23*)	(*Ucp*)	AKXD, BXD
GTT11	B1 repeat			
GTT12	-	Chr 14 (*D14Abb2e*)	*Glud, Ms15-7*	BXD
GTT25	5 human ESTs (N34309)	Chr 2 (*D2Abb1e*)	*D2Mit93, D2Mit97*	JAX BSS
GTT28	2 human ESTs (R98066)			
GTT37	-			
TCC2	B1 repeat			
TCC3	12 human ESTs (T65567), 2 rat ESTs (H34280)			
TCC10	4 human ESTs (N24744), 1 mouse EST (R75548)	Chr 2 (*D2Abb2e*)	*D2Mit25, D2Mit74*	JAX BSS
TCC15	-	Chr 2 (*D2Abb3e*)	*Psp, Src*	BXH, NX129
TCC18	14-3-3 zeta isoform	Chr 15 (*D15Abb1e*)	*D15Mit13, Trhr*	AXB, BXA
TCC21	-	Chr 17 (*D17Abb1e*)	*D17Bir7, D17Mit9*	JAX BSS
TCC22	-	Chr 4 (*D4Abb1e*)	*D4Bir9, D4Mit4*	JAX BSS

Shown are the homologies detected for each clone in GenBank. In cases of multiple hits to human ESTs, the accession number is given only for the most significant homolog.

[a]The flanking markers shown have been selected as those that are closest to the locus being mapped, but also according to appearance on readily available consensus maps of mouse chromosomes. When a marker appears in brackets it means that this locus did not recombine with the TR-containing locus. The mapping data for clone TCC 15 should be treated with some caution, as the data obtained are consistent with the localization given, but only a relatively small number of RI strains were able to be analyzed, and it was not possible to amplify DNA from *M. spretus* with these primers. Data from the Jackson Laboratory Interspecific Backcross panels is available on the World Wide Web at http://www.jax.org/resources/documents/cmdata.

These two clones contained the longest repeats identified in this survey. In seven clones, the repeat was immediately adjacent to the polylinker. In other clones, notably those containing the repeat GTT, the repeat was associated with a B1 repetitive element, preventing the design of unique sequence flanking primers. In four further cases, the PCR failed under all conditions tried, even when several different pairs of primers had been used. These clones presumably contained sequences refractory to amplification, or, more probably, spanned introns that would only have been detectable by screening for, and analysis of, genomic clones. A third possibility is that sequence errors were introduced to the clones during PCR from the original phage clones, and this compromised the primer design.

DNA from a range of different inbred strains, subspecies, and species of mouse was subjected to PCR using the conditions shown in Table 3. DNA

Downloaded from genome.cshlp.org on June 22, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 3. PCR Primers and Conditions for amplification of Mouse Trinucleotide Repeats

Clone name	Primer sequences	Annealing temperature (°C)	Mg^{2+} (mM)	Other reagents	Product size
AAT5	5'GCCCTGGCTCGAAAAACCAAA 5'GAATGTTTACTGCAAAAACTGGAG	50	1.5	-	120 bp
CAG2	5'GTTGCACAGACTTCATGCAT 5'GAAGAAGGACCAGGCTTAGG	55	1.5	10% DMSO	260 bp
CAG3	5'AAAAAGGGCCTGGAGAAGAAC 5'CGCAAACTAGATTCTCAAGTGG	50	1.5	10% glycerol	210 bp
CAG17	5'AACCGCCGATGCCAGTGAAT 5'TTAATGCTGCCTGGACAGAG	55	1.5	10% glycerol	110 bp
CAG23	5'CAGCACACGTCCTTCAGAAT 5'GAGGACCTGGAGTTTGTGTA	50	1.5	10% glycerol	150 bp
CAT6	5'CACAGAAACAGAAGAAGCAG 5'CTTCTTGACCTTGTTTTGGG	55	1.5	-	187 bp
CAT9	5'CAGGTTACATGCATCGTGGC 5'TCACAGAATACAGCCGAGTC	55	1.5	-	130 bp
CAT10	5'GGTCAAGTCCTGATACACAT 5'CCCCTGTTGACCCTTAACAA	55	1.5	-	271 bp
CAT11	5'CACCACTTCTAAAGTACAGC 5'GGTATCACAGAATACAGCCG	55	1.5	-	210 bp
CCA5	5'TACCATTGTTTCACACCAGG 5'CACATACCGTCAACTTCACC	55	1.5	10% DMSO	135 bp
GTT1	5'CAGGTGAAGGAAGGAACCAA 5'TTCCTAGACAGTGAGACAGG	50	0.9		170 bp
GTT3	5'CCCTAGATGTACTCACTGGA 5'GAGGTTGAGGTCATACATGC	55	1.5		260 bp
GTT8	5'CAGCACTGGAAGGTTCTTGA 5'AACATCTGAAGCCAGGCAGT	55	1.5	10% DMSO	219 bp
GTT9	5'CCCTCATCATGGTGGAGGAA 5'CCTCCTGAGTCAGAACTTTG	50	1.5	10% DMSO	410 bp
GTT12	5'GTTCTAGAGTCCAAGAACTC 5'CCATTAACAGCTGTACTTGGC	55	1.5	10% DMSO	127 bp
GTT25	5'AGTGACCAGGAAGTGTGCCA 5'CATTGAATTACTAAACAAAGATGGTG	55	1.5		220/200 bp
GTT28	5'CCTCTTCCAGTCAAAGAAGG 5'CGGAGAGGCTATGGTTTGC	50	1.5	10% DMSO	133 bp
GTT37	5'AATTCCGGAAATGGTTGGGT 5'CTGCGAATTTCTTCCCAACC	50	1.5	10% DMSO	209 bp
TCC3	5'TCAAGCTACGCTGAGGAAGA 5'ACAAGTTTACCCTCCTCCCA	55	1.5	10% DMSO	194 bp
TCC10	5'GGTTCACTGGTAAGCAGTCT 5'GAAGTTCCAAAGTCCAGCAG	55	1.5	10% glycerol	140 bp
TCC15	5'TTGCAGCCCAGTGTCAAGA 5'GTTTGTTCTCCAGCCTCTGT	55	0.5	10% DMSO	210 bp
TCC18	5'CCCCTCATCCTCTCTACAG 5'CTTTGGGTGTGACTTAGCC	55	1.5	-	160 bp
TCC21	5'TCTTGTCTCCTTAAATGGGA 5'GAAAACGATGGTCACTGAGG	55	1.5	-	195 bp
TCC22	5'CCCAGTGAAGACAAGGAGAA 5'ACCTTGGTCAACAGGTTGAC	55	2.5		190 bp

Shown are the sequences of the PCR primers used to amplify across individual repeats, together with modifications to the basic PCR protocol used.

Downloaded from genome.cshlp.org on June 22, 2026 . Published by Cold Spring Harbor Laboratory Press

from 10 *Mus musculus musculus/Mus musculus domesticus* inbred strains was tested, along with DNA from mice of the subspecies *Mus musculus castaneus* (strain CAST/Ei) and *Mus musculus molossinus* (strain MOLF/Ei) and the species *Mus spretus* and *Mus caroli*. PCR products were initially analyzed by electrophoresis through 3%–6% NuSieve agarose, depending on the expected product size, and then, if no variation was detected, by running radiolabeled PCR products on sequencing gels. Of the 25 TRs analyzed in this way, only 7 appeared not to vary in size among mouse species, even though the repeats occur in transcribed regions. These seven include CAT2, which was not analyzed by PCR because the CAT repeat is conserved in human, rat, and mouse and is the same size in each. It can therefore be assumed not to vary among mouse strains. Eight repeats showed length variation between commonly used inbred *M. musculus* strains, one showed variation between standard laboratory inbred strains and CAST/Ei or MOLF/Ei, and a further seven showed variation between *M. musculus* and *M. spretus*. The remaining two showed length variation only between *M. musculus* spp. and *M. caroli*. These figures (64% of repeats varying between *M. musculus* and *M. spretus* and 24% between *M. musculus* strains) are lower than those quoted for dinucleotide repeats isolated from random regions of the mouse genome (90% and 50%, respectively; Dietrich et al. 1993). This may be a sampling error attributable to the relatively small number of clones analyzed in this study or be because the repeats isolated in this study all occur within transcribed regions and are therefore subject to greater evolutionary constraints. Alternatively, it may simply reflect the fact that a number of the repeats examined here are short and imperfect. An estimate of the degree of variability of each repeat is given in Table 1.

Mapping

Those repeats that displayed length variation were mapped in the mouse genome by linkage analysis using recombinant inbred strains or interspecific backcrosses between *M. spretus* and C57BL/6J (Rowe et al. 1994). The mapping data obtained are shown in Table 2. As expected, the genes containing TRs are distributed throughout the genome. There appears to be some clustering on MMU6, but this is presumably coincidental. The mapping of the clone designated GTT3 provides confirmation of the previous assignment of the gene for mitochondrial malate dehydrogenase to MMU5. This gene had been mapped previously by linkage analysis using electrophoretic variation of the protein. The clone designated GTT3, on the other hand, which is 90% identical at the DNA sequence level to rat Ca^{2+}–ATPase isoform 2 maps not to MMU12, the location of the mouse *Caa* gene, but to MMU6. *Caa* was mapped using a polymorphism in enzyme activity, and therefore probably represents a different isoform from that cloned as GTT3. This result is consistent with the mapping of the human gene for isoform 2 (*ATP2B2*) to chromosome 3p26–p25, a region that shows conservation of synteny with mouse chromosome 6 (Wang et al. 1994). The mapping of clone GTT9 on MMU8 provides confirmation of the recent cytogenetic assignment of the mouse homolog of the yeast *RAD23* gene to chromosome 8C3 (van der Spek et al. 1996). The clone designated TCC18 appears to represent the mouse 14-3-3 ζ gene, which we have now mapped to MMU15. Clone CAT11 is homologous to several components of the TFIID transcription initiation complex. The PCR primers flanking the CAT repeat amplified up to three distinct bands in different mouse strains. The band of the predicted size based on repeat length in the clone appeared not to vary in size, was present in all samples tested, and could not therefore be mapped. However, the bands above and below this that were seen in almost all inbred strains were not present in *M. spretus*. We therefore mapped these bands by presence/absence variation, and found them to map to MMU6 (between *D6Mit10* and *Rad52*) and MMU4 (between *D4Bir10* and *Xpa*). It is not known whether these bands represent pseudogenes (as seems likely) or related, expressed sequences.

Once the TRs had been mapped in mouse, the map location was compared with the consensus map of the relevant chromosome to assess whether any existing mouse mutant mapped to the same general region. Where mouse mutants whose DNA was available were found mapping to the region, this DNA was subjected to PCR using the same primers used for mapping, in order to establish whether any variation of the length of the repeat could be seen in the mutant compared with control mice from the same genetic background. This analysis is clearly incomplete because it is based on crude comparisons between maps, and rests on availability of DNA. In no case was such variation found (data not shown), although in the case of dominant mutations it is

Downloaded from genome.cshlp.org on June 22, 2026 . Published by Cold Spring Harbor Laboratory Press

impossible to demonstrate unequivocally that amplification has not occurred in the mutant allele to a size beyond that amplifiable by PCR.

Detection of New Alleles in Inbred Strains

Analysis of TRs in RI strains provides, in addition to mapping information, a way of evaluating stability of repeat length. RI strains have been inbred for at least 20 generations after the initial crossing of parental strains; each RI strain should carry alleles with the same repeat length as one or other of the parental strains. Of the five repeats isolated in this study that were analyzed in at least one set of RI strains, two (CCA5 and GTT9) had changed in size in at least one RI line compared with the parental strains. In each case, the strain concerned was homozygous for the new repeat length, suggesting the event had occurred some time previously. CCA5 exhibited a decrease in size in AXB1 relative to both parental strains. From the map position of CCA5, the most likely event is that the C57BL/6J allele had lost a single trinucleotide. GTT9 exhibited an increased size allele in AKXD12, which was most likely a gain of at least two trinucleotides from the original DBA/2 allele.

DISCUSSION

The distribution of trinucleotide classes among the mouse cDNA clones isolated here in most cases parallels that described previously from surveys of human, mouse, and rat sequences in data bases (Beckman and Weber 1992; Stallings 1994). No clones were found to contain $(ACT)_n$ or $(CGA)_n$ repeats. In contrast to data base surveys, no clones containing $(CCG)_n$ repeats, and far fewer clones containing $(CAG)_n$ repeats were found. This is presumably a reflection of the lack of 5′ sequences in the cDNA clones (which originated from an oligo(dT)-primed library), because CCG repeats are usually found in 5′ untranslated regions, or toward the 5′ end of coding regions. The under-representation of $(CAG)_n$ repeats is unexpected. The over-representation of clones containing $(AAT)_n$ and $(GTT)_n$ repeats is probably a result of the association of these sequences with Alu-type elements.

It is clear from this study that, as in humans, TRs isolated from cDNAs form useful genetic markers, as they are directly associated with identifiable genes and yet are highly variable and easily typed. They are potentially useful in comparative mapping, as TRs are often conserved between closely related species, and even if the repeat is not perfect, it may still display length variation within a species (Ricke et al. 1995).

Within the limited number of clones analyzed in this study, there is no obvious correlation between degree of variability of repeat length and factors such as repeat class, interruptions within the repeat, length of repeat in original clone, or where within the gene the repeat occurs (although this is not known for most of the clones in this study, other than that they are likely to be within 3′ untranslated or coding regions). It is therefore not possible to predict from sequence analysis alone which repeats are likely to be the most useful for mapping. This is in contrast to previous studies of $(CA)_n$ repeats (Weber 1990) and $(CAG)_n$ repeats in human genes (Riggins et al. 1992), but is consistent with our previous observations of mouse genes containing $(CAG)_n$ repeats identified from a data base survey (Abbott and Chambers 1994).

Abnormal TR expansion appears thus far to be a human-specific phenomenon. The reasons for this are unclear. No mouse mutation has been found to be caused by TR expansion, and anticipation has not been reported in the mouse. This may be because mouse mutagenesis programs have traditionally used reagents such as ethylnitrosourea, which generally induces point mutations, and X-irradiation which usually causes chromosomal rearrangements such as deletions (Rinchik and Russell 1990). It is hard to predict how a spontaneous TR expansion would behave in a mouse line. Because the generation time is so rapid compared with human, it is possible that such a mutation, assuming it was showing anticipation, would die out before it could be properly analyzed, particularly if the phenotype had been extremely mild in the early generations. However, the recent discovery of an expanded GAA repeat in the *FRDA* gene (Campuzano et al. 1996) suggests that there may be recessive disorders in the mouse, not necessarily showing anticipation, caused by TR expansion. Transgenic mice carrying long $(CAG)_n$ repeats have been made (Bingham et al. 1995; Burright et al. 1995), and in each case the repeat has been reported to be stably inherited through multiple meioses. This may reflect a fundamental difference between mice and humans during gametogenesis and/or early development. Alternatively, other factors may be involved. For example, predisposing haplotypes

Downloaded from genome.cshlp.org on June 22, 2026 . Published by Cold Spring Harbor Laboratory Press

have been described for a number of human disorders caused by TR expansion (e.g., Myers et al. 1993), but the transgenic mice reported thus far have been made using expanded repeats isolated from patients retrofitted into genomic clones of unknown haplotype. Alternatively, it may be necessary to investigate the behavior of TRs in transgenic mice compromised in some way with respect to DNA repair, for example, mice carrying mutations in mismatch repair genes.

METHODS

Library Screening

A mouse brain cDNA library in λgt11 was obtained from the ATCC (no. 37431). This library was constructed from an 18-day-old NIH Swiss mouse and has inserts averaging 0.8–1 kb. The library was plated out at 150,000 PFU/plate, and plaque lifts were made onto Hybond N (Amersham) in duplicate. Filters were prehybridized in 7% $NaDodSO_4$/0.5 M $NaPO_4$ for 1 hr at the relevant temperature (see below). 15-mer oligonucleotides were end-labeled with [γ-^{32}P] ATP, and filters were then hybridized with the labeled oligonucleotide in Church buffer at the following temperatures (T_m − 5°C) for 4 hr:

$(AAT)_5$: 32°C

$(AGA)_5$, $(CAT)_5$, $(GAT)_5$, $(GTT)_5$ and $(ACT)_5$: 45°C

$(CAG)_5$, $(CTG)_5$, $(TCC)_5$, $(CCA)_5$ and $(CGA)_5$: 59°C

$(CCG)_5$: 72°C

The filters were then washed at room temperature for 10 min in 6× SSC, 3 × 20 min in 6× SSC/0.1% SDS at the hybridization temperature, and 1 × 2 min at T_m. Autoradiographs were then set up overnight. Secondary screening of positive clones was carried out as described above. Positives were picked into SM buffer (50 mM Tris-HCL at pH 7.5/100 mM NaCl/8 mM $MgSO_4$/0.01% gelatin) and stored at 4°C.

Subcloning and Sequencing

Inserts were prepared by amplifying directly from plate lysates, by PCR using primers corresponding to the λgt11 arms. PCR was carried out under standard conditions using a 2-μl plate lysate in a 50-μl total reaction mix. The resulting PCR products were subcloned into pUC18. Sequencing of the insert DNA was carried out using Sequenase (United Biochemicals) under standard conditions; sequencing was continued until the TR had been sequenced throughout, with flanking regions. New sequencing primers were synthesized to walk through the insert where necessary. Sequences were compared with those in the GenBank and EMBL data bases using the BLAST program through the Human Genome Mapping Project–Resource Centre (HGMP–RC) computing facilities.

PCR

Primers were designed to amplify across each TR (see Table 3). PCR was carried out under standard conditions using *Taq* polymerase from Advanced Biotechnologies or Promega, with buffer supplied. Annealing temperatures, optimum magnesium concentrations, and additional reagents used for each pair of primers are given in Table 1. DNA samples from inbred strains were purchased from the Jackson Laboratory (Bar Harbor, ME).

Mapping

Each locus was mapped using RI strains (using DNA purchased from the Jackson Laboratory, Bar Harbor, ME) or the Jackson Laboratory Interspecific Backcross (Rowe et al. 1994). Either panel 1 (C57BL/6J × *M. spretus*)F_1 × C57BL/6J or panel 2 (C57BL/6JEi × SPRET/Ei)F_1 × SPRET/Ei DNAs were used. Each panel consists of 94 N_2 offspring. Results were analyzed by Lucy Rowe using the Map Manager program (Manly 1993).

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Ed Birkenmeier of the Jackson Laboratory, without whose foresight in setting up the Jackson Laboratory Interspecific Backcross this study could not have taken place. He will be sorely missed by friends and colleagues worldwide. We are very grateful to Dr. Susan Darling for her help in the early stages of this work with the library screening. Drs. Una Fairbrother, Lorna Pate, and Gordon McGurk have also given helpful advice. Lucy Rowe and Mary Barter have provided enormous help in the mapping of these loci, and we thank them. We are grateful to Lois Maltais for her help with nomenclature. We thank Adrian Bird, Nick Hastie, and Lucy Rowe for critical reading of the manuscript. This work was supported by the United Kingdom MRC HGMP. The sequence data described in this paper have been submitted to the EMBL data library under accession numbers Z74621–Z74661.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Abbott, C. and D. Chambers. 1994. Analysis of CAG trinucleotide repeats from mouse cDNA sequences. *Annu. Hum. Genet.* **58:** 87–94.

Ashley, C.T. and S.T. Warren. 1995. Trinucleotide repeat expansion and human disease. *Annu. Rev. Genet.* **29:** 703–728.

Bakker, C.E., C. Verheij, R. Willemsen, R. Vanderheim, F. Oerlemans, M. Vermey, A. Bygrave, A.T. Hoogeveen, B.A. Oostra, E. Reyniers, K. Deboulle, R. Dhooge, P. Cras, D. Vanvelzen, G. Nagels, J.J. Martin, P.P. Dedeyn, J.K.

Downloaded from genome.cshlp.org on June 22, 2026 . Published by Cold Spring Harbor Laboratory Press

Darby, and P.J. Willems. 1994. FMR1 knockout mice—a model to study fragile X mental retardation. *Cell* **78:** 23–33.

Beckman, J.S. and J.L. Weber. 1992. Survey of human and rat microsatellites. *Genomics* **12:** 627–631.

Bingham, P.M., M.O. Scott, S.P. Wang, M.J. Mcphaul, E.M. Wilson, J.Y. Garbern, D.E. Merry, and K.H. Fischbeck. 1995. Stability of an expanded trinucleotide repeat in the androgen receptor gene in transgenic mice. *Nature Genet.* **9:** 191–196.

Braga, E.A., B.I. Kapanadze, N.S. Kupriyanova, G.M. Ivanova, V.M. Brodyanskii, K.K. Nechvolodov, G.A. Skutov, A.P. Ryskov, V.V. Nosikov, and N.K. Yankovskii. 1995. Distribution analysis of 7 microsatellite motifs in cosmids of human chromosome 13 library. *Mol. Biol.* **29:** 584–590.

Burright, E.N., H.B. Clark, A. Servadio, T. Matilla, R.M. Feddersen, W.S. Yunis, L.A. Duvick, H.Y. Zoghbi, and H.T. Orr. 1995. SCA1 transgenic mice: A model for neurodegeneration caused by an expanded CAG trinucleotide repeat. *Cell* **82:** 937–948.

Campuzano, V., L. Montermini, M.D. Molto, L. Pianese, M. Cossee, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, A. Monticelli, F. Zara, J. Canizares, H. Koutnikova, S. Bidichandani, C. Gellera, A. Brice, P. Trouillas, G. De Michele, A. Filla, R. De Frutos, F. Palau, P.I. Patel, S. Di Donato, J.-L. Mandel, S. Cocozza, M. Koenig, and M. Pandolfo. 1996. Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271:** 1423–1427.

Dietrich, W., H. Katz, S.E. Lincoln, H.-S. Shin, J. Friedman, N.C Dracopoli, and E.S. Lander. 1993. A genetic map of the mouse suitable for typing intraspecific backcrosses. *Genetics* **131:** 423-447.

Li, S-H., M.G. McInnis, R.L. Margolis, S.E. Antonarakis, and C.A. Ross. 1993. Novel triplet repeat containing genes in human brain: Cloning, expression, and length polymorphisms. *Genomics* **16:** 572–579.

Manly, K.F. 1993. A Macintosh program for storage and analysis of experimental genetic mapping data. *Mamm. Genome* **4:** 303–313.

Myers, R.H., M.E. MacDonald, W.J. Koroshetz, M.P. Duyao, C.M. Ambrose, S.A.M. Taylor, G. Barnes, J. Srinidhi, C.S. Lin, W.L. Whaley, A.M. Lazzarini, M. Schwartz, G. Wolff, E.D. Bird, J.-P.G. Vonsattel, and J.F. Gusella. 1993. De novo expansion of a (CAG)n repeat in Huntington's disease. *Nature Genet.* **5:** 168–173.

Ricke, D.O., Q. Liu, B. Gostout, and S.S. Sommer. 1995. Nonrandom patterns of simple and cryptic triplet repeats in coding and noncoding sequences. *Genomics* **26:** 510–520.

Riggins, R.G., L.K. Lokey, J.L. Chastain, H.A. Leiner, S.L. Sherman, K.D. Wilkinson, and S.T. Warren. 1992. Human genes containing polymorphic trinucleotide repeats. *Nature Genet.* **2:** 186–191.

Rinchik, E.M. and L.B. Russell. 1990. Germ-line deletion mutations in the mouse: Tools for intensive functional and physical mapping of regions of the mammalian genome. In *Genetic and physical mapping* (ed. K.E. Davies and S.M. Tilghman), pp. 121–158. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Rowe, L.B., J.H. Nadeau, R. Turner, W.N. Frankel, V.A. Letts, J.T. Eppig, M.S.H. Ko, S.J. Thurston, and E.H. Birkenmeier. 1994. Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome* **5:** 253–274.

Stallings, R.L. 1994. Distribution of trinucleotide microsatellites in different categories of mammalian genomic sequence–implications for human genetic diseases. *Genomics* **21:** 116–121.

van der Spek, P.J., C.E. Visser, F.H. Hanaoka, B. Smit, A. Hagemeijer, D. Bootsma, and J.H.J. Hoeijmakers. 1996. Cloning, comparative mapping, and RNA expression of the mouse homologues of the *Saccharomyces cerevisiae* nucleotide excision repair gene RAD23. *Genomics* **31:** 20–27.

Wang, M.G., H. Yi, H. Hilfiker, E. Carafoli, E.E. Stehler, and O.W McBride. 1994. Localizations of two genes encoding plasma membrane Ca^{2+}-ATPases isoforms 2 (ATPB2) and 3 (ATPB3) to human chromosomes 3p26-p25 and Xq28, respectively. *Cyogenet. Cell Genet.* **67:** 41–45.

Weber, J.L. 1990. Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. *Genomics* **7:** 524–530.

Received April 5, 1996; accepted in revised form June 14, 1996.