 Human Molecular Genetics, 2002, Vol. 11, No. 2 191–198

Somatic expansion behaviour of the $(CTG)_n$ repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch–repair proteins

Walther J. A. A. van den Broek, Marcel R. Nelen, Derick G. Wansink, Marga M. Coerwinkel, Hein te Riele[1], Patricia J. T. A. Groenen and Bé Wieringa*

Department of Cell Biology, UMC Nijmegen, Nijmegen Center for Molecular Life Sciences, PO Box 9101, 6500 HB Nijmegen, The Netherlands and [1]Division of Molecular Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

Received October 3, 2001; Revised and Accepted November 16, 2001

The mechanism of expansion of the $(CTG)_n$ repeat in myotonic dystrophy (DM1) patients and the cause of its pathobiological effects are still largely unknown. Most likely, long repeats exert toxicity at the level of nuclear RNA transport or splicing. Here, we analyse *cis*- and *trans*-acting parameters that determine repeat behaviour in novel mouse models for DM1. Our mice carry 'humanized' myotonic dystrophy protein kinase (*Dmpk*) allele(s) with either a $(CTG)_{84}$ or a $(CTG)_{11}$ repeat, inserted at the correct position into the endogenous *DM* locus. Unlike in the human situation, the $(CTG)_{84}$ repeat in the syntenic mouse environment was relatively stable during intergenerational segregation. However, somatic tissues showed substantial repeat expansions which were progressive upon aging and prominent in kidney, and in stomach and small intestine, where it was cell-type restricted. Other tissues examined showed only marginal size changes. The $(CTG)_{11}$ allele was completely stable, as anticipated. Introducing the $(CTG)_{84}$ allele into an Msh3-deficient background completely blocked the somatic repeat instability. In contrast, Msh6 deficiency resulted in a significant increase in the frequency of somatic expansions. Competition of Msh3 and Msh6 for binding to Msh2 in functional complexes with different DNA mismatch-recognition specificity may explain why the somatic $(CTG)_n$ expansion rate is differentially affected by ablation of Msh3 and Msh6.

INTRODUCTION

Myotonic dystrophy [MIM 160900; dystrophia myotonica 1 (DM1)] is the most prevalent form of adult muscular dystrophy, with multisystemic features and a highly variable clinical phenotype (1). Careful clinical evaluation and genetic testing of large groups of patients with this dominantly inherited disorder has revealed that at least two forms of disease and underlying genetic causes can be distinguished (2–8). The genetic basis for the most common form of DM, DM1, is an expanded CTG repeat in the 3′-untranslated region (UTR) of the DM protein kinase (*DMPK*) gene (9–12) and in the promoter region of the immediately adjacent homeodomain *SIX5* gene (3) on chromosome 19q13. Myotonic dystrophy type 2 [MIM 602668; Dystrophia Myotonica 2 (DM2)] displays a remarkable identical phenotype (13) and is recently linked to a CCTG expansion in the first intron of the zinc finger protein 9 gene (*ZNF9*) on chromosome 3q21 (14).

Because of the striking parallels between these mutations, both of which eventually give rise to large microsatellite expansions in non-coding areas of muscle-expressed (pre)mRNAs, it has been speculated that the expansions create a distinct type of dominant gain-of-function mutation (6,8,15). In one proposed mechanism the mutations would have *cis*-effects on processing, routing or translation of the *DMPK* or *ZNF9* mRNAs proper. This mechanism assumes that DMPK and ZNF9 work in the same cellular pathway and the same set of cellular functions may be distorted. More likely is that the mutations cause the same type of *trans*-acting toxic effects on nuclear metabolism by sequestering specific proteins that are required for pre-mRNA splicing or export (16,17). Hypotheses that postulated how expanded repeats could alter expression of neighbouring genes by position effect variegation have lost attractiveness. Since there is no similarity between the DM1 and DM2 gene clusters it is difficult to see how deregulation of expression of completely different subsets of genes would produce such a remarkable similarity in disease phenotype.

In DM1 it is particularly well documented that the size of the repeat is correlated negatively with age of onset and positively with the severity of the disease phenotype. As repeats of >40 CTG will expand further in successive generations, and large repeats are very unstable in the germline and heavily biased

*To whom correspondence should be addressed. Tel: +31 24 3614329; Fax: +31 24 3615317; Email: b.wieringa@ncmls.kun.nl
Present address:
Patricia J. T. A. Groenen, Department of Pathology, UMC Nijmegen, Nijmegen, The Netherlands

towards further expansion, this provides the molecular explanation for the 'anticipation' phenomenon. In the literature there is a vast amount of data indicating that somatic mosaicism in DM1 is also very high. Small-pool PCR experiments have demonstrated that the CTG repeat in human tissues displays an age-dependent (18), expansion-biased (19), tissue-specific (20,21) length heterogeneity that appears to be mediated via multiple successive small length change events [mostly size increase (22)], starting early after gestation. In addition, transgenic animals that carry the DM CTG repeats show both intergenerational and somatic instability. Strikingly, a strong tissue-type dependence of repeat length instability has been observed, suggesting that tissue-specific factors may be involved (23,24).

Unfortunately, not much is known about the molecular basis underlying the repeat expansion, or the mechanisms involved in creating the tissue variability in CTG length. Recent studies on the expansion mechanism of the CAG repeat in Huntington disease (HD) in transgenic mice (25–27) pointed to the possible involvement of components of the DNA repair machinery. The data obtained support models in which expansion arises by single-stranded DNA gap repair and is not dependent on mitotic replication or recombination between homologous or sister chromatids.

Here we report on a study of a new knock-in mouse model with a long CTG_{84} repeat placed into the mouse DM1 locus, at a site exactly homologous to that where it is found in human patients. This enabled us to examine the (*cis*-acting) effects of proximal gene regions and (*trans*-acting) factors that relate to tissue type and ageing. Importantly, in order to clarify the role of individual key proteins in the DNA mismatch and repair process we crossed a CTG_{97}-DM1 locus into backgrounds with Msh3 and Msh6 deficiency. Our results demonstrate that, although Msh3 and Msh6 both form functional complexes with the Msh2 repair protein (28), they have a profoundly different role in the recognition and/or processing of anomalous DNA structures that may form in or around long CTG tracts at the DM1 locus.

RESULTS

Production of a DM1 CTG knock-in mouse model

To generate a mouse model with an expanded human $(CTG)_{84}$ repeat in the cognate position of the endogenous *Dmpk* gene, we used a multistep replacement mutagenesis procedure with targeted recombination-insertion and Cre-loxP based excision steps (29) in mouse 129/Ola-derived ES cells (Fig. 1). As a control for monitoring the effect(s) of the replacement of exons 13–15 by homologous human sequences, we created mutant mice with a non-expanded 'normal' $(CTG)_{11}$ repeat. Several independent ES cell clones were obtained from which highly chimeric $(CTG)_{84}$- or $(CTG)_{11}$-*Dmpk* allele transmitting males were derived. All heterozygous and homozygous mutant mice obtained during subsequent breeding were indistinguishable from their normal littermates, fertile, and displayed no overt outward anomalies. Northern and western blot analyses revealed that the mutant mice expressed the predicted chimeric mouse–human *Dmpk* transcripts, resulting from alternative splicing (12,30) of mouse exon 12 to either human exon 14 or 15, albeit at reduced levels. However, their tissue distribution profile was exactly similar to that observed for endogenous *Dmpk* mRNA and protein isoforms (12), indicating that the chimeric allele obeys the same gene-regulatory control as the wild-type allele (data not shown; D.G.Wansink, M.M.Coerwinkel, W.J.A.A.van den Broek, P.J.T.A.Groenen and B.Wieringa, manuscript in preparation).

Intergenerational segregation of the CTG repeat

We next tested how the $(CTG)_{84}$ allele behaves upon intergenerational transmission. In mice from cohorts with the mixed C57Bl/6 × 129/Ola background, the repeat was essentially stable and only $(CTG)_{84,85,86}$ alleles were found. However, during breeding efforts to place the $(CTG)_{84}$ allele in a pure C3H background we discovered one lineage of mice with slow but progressive increments in repeat size (Fig. 2A). The average repeat length in tail DNAs in this pedigree increased from 84 to 97 CTGs in three generational jumps and is now at 100 in the F7 generation. One major contraction was detected (–16 CTGs). In the cohorts analysed, similar numbers of contractions and expansions were seen: CTG repeat expansions (+1 to +8 triplets) in 42 out of 132 parent–sibling pairs, contractions (1–2 triplets, with the exception mentioned above) in 38 out of 132 parent–sibling pairs. This means that the frequency of mutation is different from the situation in human DM families, where even in non-manifesting carriers—with $(CTG)_{41–95}$ repeats—the incidence of significant repeat-length alterations is in the order of ~90–95% [mainly expansions (31)]. Sequence analysis of genomic tail DNA of $(CTG)_{84}$ mice revealed that the repeats were without imperfections, which theoretically could hamper repeat instability.

Cell-type-dependent factors are involved in CTG-repeat expansion

To study repeat behaviour at the somatic level and monitor the effects of genetic background we measured $(CTG)_{84}$ allele sizes in various tissues from several heterozygous transgenics (successive generations, 3–22 months of age) held on a mixed C57BL/6 × 129/Ola background or crossed once into a Balb/C, FVB or C3H environment. Repeat patterns obtained after PCR amplification consistently displayed size increments in all tissues when compared to the pattern generated from the plasmid-borne $(CTG)_{84}$ repeat (Figs 2B and 3A). When tissues of animals with different backgrounds were compared ($n = 2$, randomly selected animals from each group with mixed Balb/C, FVB, 129 or C3H background; analysis at both 6 and 12 months of age) those with a C3H background contribution ≥50% showed the highest fractional contribution of expanded alleles (Fig. 3A).

Strikingly, there was pronounced instability in DNA from stomach, kidney and small intestine, in all animals tested. The instability did not differ between male and female animals and appeared progressive with age (Figs 2B and 3B). This indicates that instability persists throughout adult life as seen in DM patients (19,32) and involves an increasing number of cells in a given tissue. Based on these observations we assume that genetic constitution is important for the frequency of expansion events, influencing both the length of individual repeats and the number of cells involved as a function of time.

When tracked as a function of cell type, repeat expansion was mainly seen in cells from the mucosa and sub-mucosa fraction and much less in cells from the muscularis externa

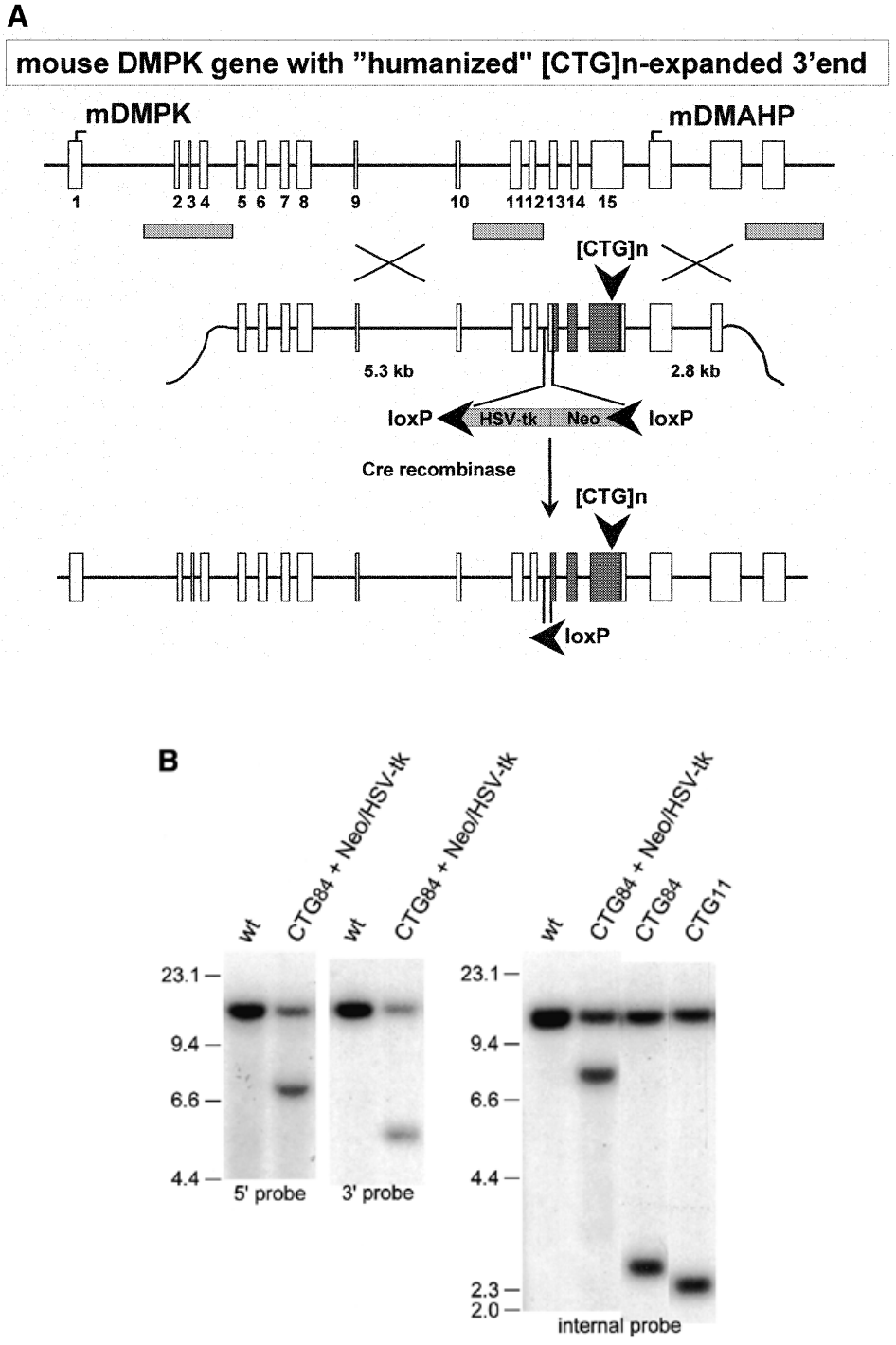

Figure 1. (**A**) Cre-Lox strategy for replacement of the 3′ end of the mouse DMPK gene by DNA segments from the normal $(CTG)_{11}$ or DM-affected $(CTG)_{84}$ human 19q area in mouse embryonic stem cells. Genomic structure of the mouse DMPK (exons 1–15) and DMAHP (exons A, B, C) genes, the structure of the targeting vector, and the predicted structure of the targeted DMPK allele are depicted. Exons are represented as open boxes and numbered; the small angled arrows on exons 1 and A indicate the translation initiation site of the DMPK and DMAHP/six5 gene, respectively. In the first targeting round, ES cell lines were generated that have one DMPK allele with a 'humanized' 3′ end encompassing exon 13–15 region, and bear a selectable loxP flanked Neo/HSV-tk cassette in the intron 12/exon 13 area. In the second round, the Neo/HSV-tk cassette was excised upon transfection with a Cre-recombinase expression plasmid. In the resulting chimaeric gene, the mouse intron 12/human exon 13 junction segment is replaced by a loxP-vector sequence. The locations of the 1.5 kb *Bgl*II–*Bam*HI 5′ probe, the 1.5 kb *Eco*RI–*Bgl*II 3′ probe and the 1 kb internal *Hin*dIII–*Sty*I probe are indicated. (**B**) Left, Southern blot analyses of genomic DNAs from a wild-type control (wt) and a mutant Neo/HSV-tk targeted ES cell clone with $(CTG)_{84}$ after *Bgl*II digestion are shown. The expected sizes of genomic *Bgl*II fragments that hybridize with the 5′ probe are 12.5 kb for the wild-type, and 7 kb for the allele with the Neo/HSV-tk cassette and $(CTG)_{84}$ segment [6.8 kb for the corresponding $(CTG)_{11}$ allele, not shown]. For the 3′ probe, bands of 12.5 and 5.8 kb are diagnostic for the wild-type and Neo/HSV-tk with $(CTG)_{84}$ alleles [5.6 kb for the corresponding $(CTG)_{11}$ allele, not shown], respectively. Right, Southern blot analysis of *Hin*dIII-digested DNAs from wild-type ES cells (11 kb signal) and targeted ES cell clones carrying the humanized $(CTG)_{84}$ segment with and without the Neo/HSV-tk cassette (bands of 7.2 and 2.7 kb), or the pure $(CTG)_{11}$ segment without Neo/HSV-tk cassette (2.5 kb signal).

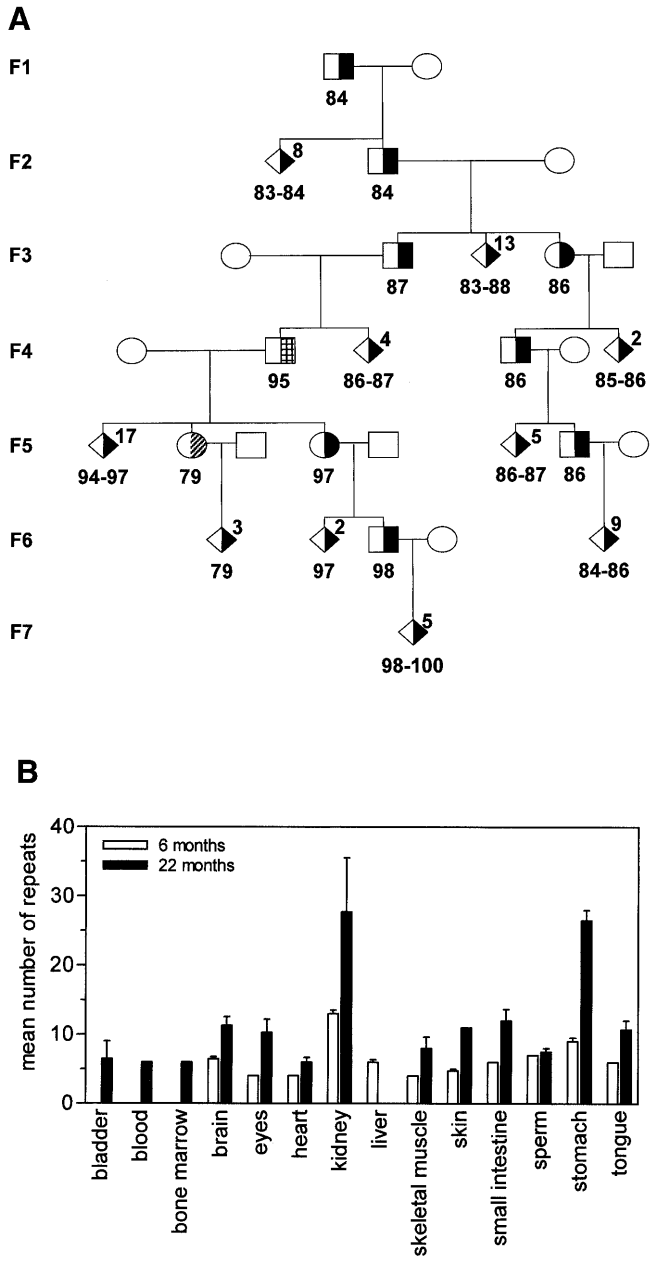

Figure 2. Intergenerational and somatic instability in $(CTG)_{84}$ mice. (**A**) The intergenerational instability in a pedigree of a transgenic mouse line crossed onto a C3H background is shown. The F1 heterozygous male carrying a $(CTG)_{84}$ allele has a 50% C57BL\6 × 129/Ola background. Open symbols indicate C3H wild-type mice used for breeding. For the sake of clarity, only heterozygous offspring are shown; diamonds mean multiple offspring as indicated by numbers. The $(CTG)_n$ repeat length(s) is indicated below each symbol. Note that the right branch is expanding only moderately (maximally +1–2 CTG triplets between generations), whereas in the left branch expansions of +3 (F2–F3) and +8 triplets (F3–F4) are observed. In the same branch one major contraction was observed (–16 CTGs). (**B**) Average increment in $(CTG)_n$ repeat size in different tissues of 6- and 22-month-old mice. In nearly all tissues analysed the $(CTG)_{84}$ was the highest peak (arrow indicated in Figure 3) and was taken as a reference point for determining the maximal detectable expansion size. The maximal number of additional (CTG) triplets, averaged over one to four mice for each data point, is given. Only increases in repeat size and no repeat contractions were observed when 6- and 22-month-old mice were compared. The repeat expansion is strongly tissue-dependent with an apparent basic mosaicism of about +4 (CTG) triplets at 6 months and +6–8 (CTG) triplets at 22 months.

fraction of the stomach [i.e. in which Dmpk is predominantly expressed (12)] (Fig. 3B). Also, in cells scraped from the luminal side of the small intestine repeat expansion was obvious. Interestingly, no mosaic length distribution with large expansions was seen in other mixed cell preparations with actively proliferating (stem) cells like sperm, bone marrow, blood or skin (Fig. 2B). This suggests that expansion is not, or not consistently, related to cell-cycle progression *per se* and may involve non-mitotic events. We are currently analysing different cell types of kidney, lung and bladder to explore the possibility that a high frequency of repeat expansion events might be a consistent feature of (epithelial?) cell types present in areas that connect with the outside of the body, i.e. in the alimentary, respiratory or genitourinary tracts.

Guarding role of Msh3 and Msh6 repair proteins

The somatic mutation frequency and clear ageing effects in stomach, kidney and intestine of our $(CTG)_{84}$ mice offered us a unique context for searching specific *trans*-acting factors that could stimulate or suppress instability, or promote transition to full $(CTG)_n$ repeat instability, as seen in humans. Previous studies have indicated that MMR protein Msh2 may have a crucial role in repeat expansions (25,26,33). To obtain insight in the significance of other repair complex proteins we placed a $(CTG)_{97}$-*Dmpk* allele, obtained from the pedigree shown in Figure 2, into an *Msh3*$^{-/-}$ or *Msh6*$^{-/-}$ background (34). Multiple comparisons of repeat size distribution between tissues from DM-transgenic *Msh6*$^{-/-}$ $(CTG)_{97}^{+/-}$ and *Msh6*$^{+/-}$ $(CTG)_{97}^{+/-}$ littermates ($n = 3$) at 6 and 12 months of age showed a consistently greater repeat instability in animals that lack Msh6. Although the distribution over tissues with high, intermediate and low expansion frequencies remained qualitatively similar, the stomach, kidney and also liver now showed a significantly stronger culmination of expansion events (Fig. 3C). It may be of interest here that we have found alleles with almost double-sized repeat lengths in an isolated liver tumour that had developed in one of our 12-month-old *Msh6*$^{-/-}$ $(CTG)_{97}^{+/-}$ mice (data not shown). Repeat length profiles in other tissues, like heart, muscle, brain, tongue, tail and sperm, were similar to those in wild-type $(CTG)_{97}^{+/-}$ animals.

In contrast, the $(CTG)_{97}$-*Dmpk* allele placed in an *Msh3*$^{-/-}$ background yielded profiles similar to that of the CTG_{84}-control plasmid, pointing to a complete absence of somatic CTG-repeat instability (Fig. 3D). Furthermore, Msh3 deficiency also completely protected against age-related (at least up to 12 months) instability.

DISCUSSION

In this study we have analysed the instability of a perfect $(CTG)_{84}$ repeat (cloned from the *DMPK* locus of a DM family) in a new 'second generation' mouse model. When placed in the cognate chromosomal environment of the mouse genome, i.e. the 3′-UTR of the *Dmpk* gene on the proximal arm of mouse chromosome 7, repeat length behaviour was grossly reminiscent of that in somatic tissues of human patients and strikingly, also of that of randomly inserted $(CTG)_{50–500}$ repeats in soma of different transgenic mouse lineages (17,23,24,35,36). Furthermore, the $(CTG)_{84}$ repeat was only marginally unstable during intergenerational transmission. This is unlike the human situation,

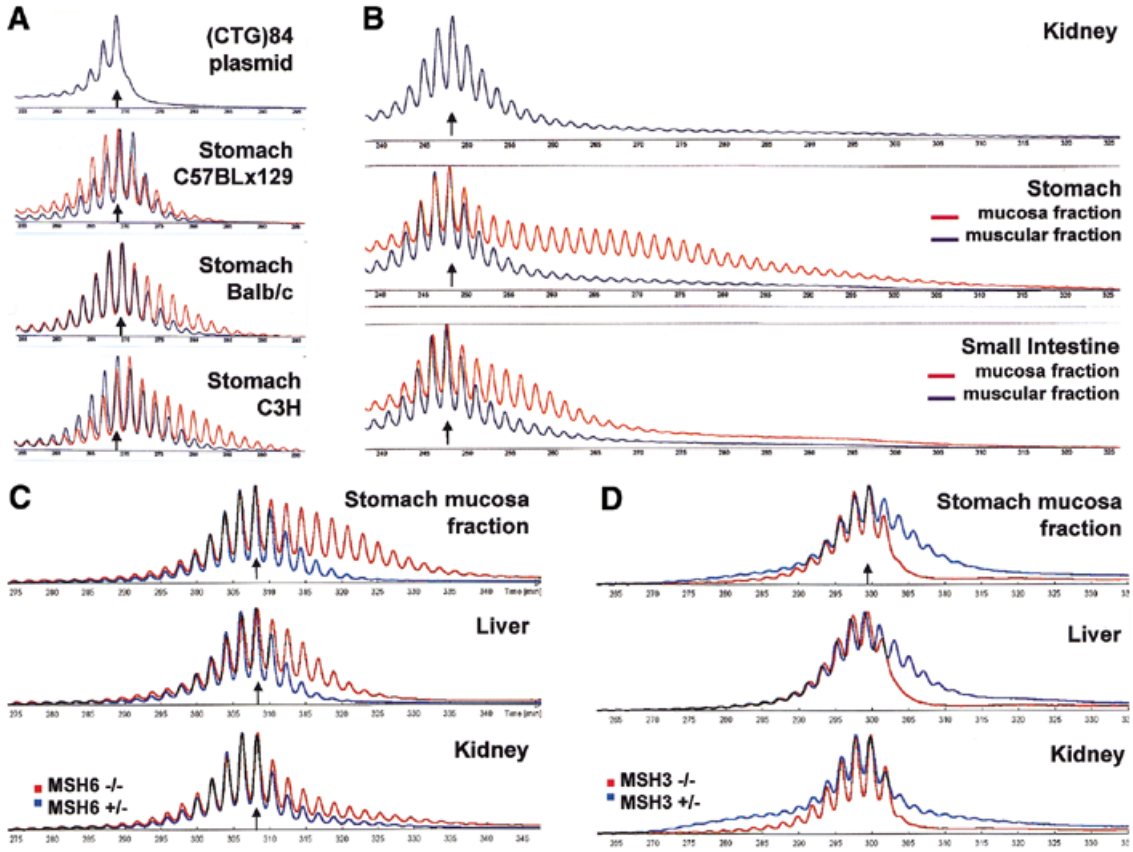

Figure 3. Somatic instability of the CTG repeat in tissues from mice with different genetic backgrounds and in different cell types. (**A**) Typical PCR profiles of the control CTG_{84} gene-targeting plasmid and of genomic DNA isolated from the stomach of a 3- (blue) and an 11-month-old (red) male mouse on a C57BL\6 × 129/Ola background, a 6- (blue) and an 11-month-old (red) male once crossed onto Balb\c (50% C57BL\6 × 129/Ola: 50% Balb\c), a 6- (blue) and an 11-month-old (red) male crossed once onto C3H (50% C57BL\6 × 129/Ola: 50% C3H). Arrows indicate $(CTG)_{84}$ size. (**B**) PCR profiles of genomic DNA isolated from kidney and mucosa and muscular fractions of stomach and small intestine of a 22-month-old male crossed once onto a C3H background (50% C57BL\6 × 129/Ola: 50% C3H). Arrow indicates $(CTG)_{84}$. (**C** and **D**) Effects of mismatch repair proteins Msh3 and Msh6 on CTG repeat instability. (C) PCR profiles of genomic DNA isolated from stomach mucosa fraction, liver and kidney of 6-month-old mice carrying one $(CTG)_{97}$ repeat allele crossed into a $Msh6^{-/-}$ (red) or a $Msh6^{+/-}$ (blue) background. CTG repeat stability appears to be promoted in an $Msh6^{-/-}$ background ($n = 3$). Other tissues examined show no increased CTG repeat instability. (D) PCR profiles of genomic DNA isolated from stomach mucosa fraction, liver and kidney of 6-month-old mice carrying one $(CTG)_{97}$ repeat allele crossed into a $Msh3^{-/-}$ (red) or a $Msh3^{+/-}$ (blue) background. CTG repeat mosaicism is completely absent in an $Msh3^{-/-}$ background ($n = 3$). [Note that the PCR profiles of different tissues were analysed on separate gels and therefore show different migration times; $(CTG)_{97}$ is arrow indicated.]

where in 90–95% of transmissions from a non-manifesting carrier to offspring a length change (predominantly expansion) would be expected for repeats in the size range between 41 and 95 triplets (31). Similar semi-stable intergenerational expansion behaviour has been reported for the transgenic DM mouse models with randomly positioned inserts and different segments of human flanking DNA. Also, for mouse strains carrying trinucleotide-repeat segments from the human HD, DRPLA or fra-X genes (37–39) repeat length changes were relatively modest and appeared less frequent than in the human situation. Taken together, these observations seem to exclude a species-specific and dominant role for flanking sequences. Other factors that may determine intergenerational repeat expandability may relate to species-specific parameters involved in setting the reproductive life-span (40), the number of cell divisions during gametogenesis (32,38) or the efficiency of DNA metabolism and repair in the germline compartment (27,33). Interestingly, Kovtun and McMurray (27) reported that expansion in a HD mouse model is limited to post-meiotic haploid cells and most likely involves gap repair by Msh2 containing complexes. Future follow-up studies in which well defined trinucleotide repeats can serve as genomic reporter elements should therefore be directed to inter-species and inter-individual comparison of mechanistic steps in this process.

Some of the steps involved in length mutation of trinucleotide repeats may be shared between cells in the germ-line and somatic tissues. Of particular interest is the observation that our $(CTG)_{84}$ knock-in model exhibited a significant and age-dependent somatic instability in all tissues examined, but most pronounced in stomach, kidney and small intestine. Closer inspection of specific cell types of both stomach and small intestine revealed that repeat expansion was mainly present in the (sub) mucosa fraction, with a high fractional content of actively dividing cells. However, such simple correlation between expandability and cell division or genome replication rate did not exist for kidney (a tissue with relatively few dividing cells and the highest rate of expandability), bone

marrow (high content of actively proliferating cells, but low expandability) or brain (very low number of dividing cells, moderate instability) (Fig. 2B). This confirms other studies [including those using cells with repeats in tissue culture (41)] and suggests that tissue-specific factor(s) rather than the frequency of replication *per se* controls the rate of somatic repeat expansion. Of particular interest seems the finding that the most conspicuous instability in our model and in other DM-transgenic lines was observed in the kidney. In cells present in the lining of the gastro-intestinal tract the repeat was also very unstable, but a comparison to tissues in other studies is not possible, because they were not analysed. Merely as a suggestion, we would like to point out that the kidney and inner intestinal layers are particularly rich in epithelial cells. These cells may perhaps need highly active DNA repair, because they are constantly in direct contact with the variable external milieu (sometimes containing genotoxic waste products) in the body. Further study and comparison of tissues like lung or bladder, and the DNA repair activity therein, is necessary to see if this speculative correlation holds.

Involvement of proteins from the DNA repair and recombination machinery is a salient feature of many studies that have pointed to the importance of post-mitotic events in trinucleotide instability in patients and cellular and animal model systems. One mechanism implicated is the activity of the mismatch–repair (MMR) machinery. For example, there is solid evidence that a core component in the MMR complex, Msh2, has affinity for DNA structures formed by CTG repeats (25). Most direct evidence comes from the group of Manley *et al.* (26) who have shown that absence of Msh2 has a stabilizing effect on the HD trinucleotide repeat. Here we report that Msh3 deficiency mimics this effect of Msh2 deficiency, in that somatic repeat instability of the $(CTG)_{84}$ repeat in our knock-in model is completely absent. In contrast, in an Msh6 deficient background higher instability was seen; particularly in stomach and kidney. Our finding that Msh3 and Msh6 have opposed effects on trinucleotide repeat instability is an unexpected result. For mismatch recognition, MSH2 forms heterodimers with either MSH6 or MSH3, depending on the type of mutation that needs to be repaired (42,43). It is generally thought that the MSH2/MSH6 complex functions in base:base mispairs whereas the MSH2/MSH3 complex is primarily responsible for the recognition of larger insertion/deletion loops. However, there is functional redundancy, as both Msh2–Msh3 and Msh2–Msh6 repair complexes recognize single base insertion/deletion mispairs (43). The $(CTG.CAG)_{97}$ DNA may be entangled in either slipped-strand structures or misaligned recombination products between sister chromatids (44,45). Our data suggests that recognition and stabilization of such large anomalous structures by Msh2/Msh3 heterodimers promotes misincorporation of additional (CTG) triplets. Over age, this could lead to involvement of an increasing number of cells, each of which could undergo multiple rounds of misincorporation, resulting in gains of relatively large tracts of CTGs. As the stability of individual Msh3, Msh6 or Msh2 proteins depends on their opportunity to heterodimerize (34) the level of the Msh2/Msh3 complex in Msh6-deficient cells could even exceed that in wild-type cells, leading to increased instability. In contrast, the presence of only the Msh2/Msh6 complex in *Msh3*$^{-/-}$ mice may help to protect against somatic expansion. Unfortunately, the precise mode of action of this complex in the process of recognition, untangling, or gap-repair of anomalous structures in trinucleotide repeats is currently not precisely known. The mechanistic aspects of repeat (de)stabilization in our repair-deficient model therefore remain enigmatic.

In summary, our data suggest that species context, cell-type and genetic background parameters work together with the MMR machinery in determining somatic $(CTG)_n$ repeat instability in the DM locus in mice. Whether this picture also holds for the somatic and meiotic instability of the repeat in DM patients remains to be seen.

MATERIALS AND METHODS

Construction of the targeting vector

The targeting vector was derived from cosmid mDMRc8 (mouse 129/Sv specific) (14) and contains the 5.3 kb *Eco*RV–*Sty*I fragment (spanning nucleotide positions 4873–10200 of the genomic mDMPK sequence) as the upstream segment, and the 2.7 kb *Nsi*I–*Eco*RI fragment (spanning nucleotide position 11865 in the *Dmpk* gene to nucleotide position 2347 in the *Dmahp* gene) (5) as the downstream segment of homology. Between these homologous arms, bracketed by a blunted *Sty*I site at position 10200 and the natural *Nsi*I site at position 11865 of the mouse sequence, we placed a loxP flanked Neo/HSV-tk cassette (courtesy of Dr P.Krimpenfort, The Netherlands Cancer Institute, Amsterdam, The Netherlands) and a human segment containing the DMPK exon 13–15 area. One type of human fragment used carried a normal $(CTG)_{11}$ repeat, the other contained an expanded $(CTG)_{84}$ repeat, and these were obtained by PCR on DNA from cosmid F100263 [=3C7 (46)] and whole genomic blood DNA from a non-manifesting DM carrier, respectively. Fragments were tailored in several cloning steps, during which the mouse *Sty*I (intron 12)–*Bam*HI (position 10379, exon 13, including the splice acceptor site) was removed. As a result, the endogenous mouse exon 13–15 gene segment [between *Bam*HI and *Nsi*I (position 11865)] was exchanged for a human segment between the *Bam*HI at position 12129 and an artificially introduced *Nsi*I site at position 13655 (corresponding to the naturally occurring *Nsi*I at position 11865 in mDMPK exon 15) carrying an expanded $(CTG)_{84}$ or non-expanded $(CTG)_{11}$ repeat.

Targeted 'humanization' of the 3′ part of the DMPK locus in ES cells and generation of mouse models

E14 ES cells, a kind gift of M.L.Hooper, were maintained as described by Bradley (47) and van Deursen and Wieringa (48) and 10^7 ES cells were mixed in medium with *Xba*I (*Dmpk* gene position 5235) linearized targeting vector DNA (12 µg) and electroporated at 250 V and 500 µF using a Gene Pulser (Bio-Rad) apparatus. Cells were seeded onto SNLH9 feeder cells and positive selection was started after 24 h with 300 µg/ml G418. Cell colonies were picked after 8–10 days of selection and screened by Southern blot analysis after *Bgl*II digestion of genomic DNA using 5′ (*Bgl*II–*Bam*HI) and 3′ (*Eco*RI–*Bgl*II) probes that flank the sequences present in the targeting vector (Fig. 1A and B). Independent homologous recombinants with normal karyotypes were used for Cre-mediated excision of the Neo/HSV-tk cassette from intron 12, leaving only the loxP

sequence (plus polylinker sequences from the vector) as a chromosomal mark. To this end, clones were transfected with a CMV promoter driven Cre-recombinase expression plasmid and selection with FIAU (0.2 μM, a kind gift of Bristol Myers) starting at day 6 after electroporation. Cell colonies were picked at day 10 of selection. Successful excision of cassette sequences was monitored by Southern blot analyses of *Hin*dIII-digested ES DNA and hybridization with a genomic *Hin*dIII–*Sty*I probe (Fig. 1A and B).

Clones that had undergone the correct genomic change(s) were identified at >95% frequency for the $(CTG)_{84}$ allele and at <10% frequency for the $(CTG)_{11}$ allele and were karyotyped again. Ultimately, ES cells from independently obtained $(CTG)_{84}$ and $(CTG)_{11}$ clones were microinjected into C57BL/6 blastocysts and used to generate chimeric mutant offspring (47). Transmitting males were used in further breeding programs. All procedures involving mice were approved by the Animal Care Committee of the University of Nijmegen and conformed to the Dutch Council for Animal Care and NIH guidelines.

Genotyping and analysis of instability of the $(CTG)_n$ repeat

Msh3 and *Msh6* heterozygous knock-out male mice on a 129/Ola background were crossed with heterozygous *Dmpk* $(CTG)_{97}$ repeat female mice [on a mixed C3H (~94%)-129/Ola/C57BL/6 (6%) background]. F1 (50% C3H/50% 129/Ola) generation mice were intercrossed to obtain mice homozygous or heterozygous for *Msh3* or *Msh6* knock-out alleles and heterozygous for the CTG_{97} allele. Littermates with the mixed inbred background were analysed for repeat instability as described below. *Msh3* or *Msh*6 genotypes were determined by PCR (34). DNA was prepared from tail biopsies or tissues samples according to standard procedures. For certain experiments stomach and small intestine were first fractionated into a mucosa-enriched fraction and a (smooth) muscle-enriched fraction. Briefly, stomach and small intestine were cut open and rinsed in PBS. Using a scalpel, the mucosa fraction (mainly the tunica mucosa and the tela submucosa) was scraped off, leaving the muscular fraction (mainly the tunica muscularis, tela subserosa and tunica serosa) behind and both preparations were analysed separately. For routine typing, PCR was performed with the forward primer 226 (5′-GAAGGGTCCTTGTAGCCGGGAA-3′, matching the DMPK gene sequence between nucleotide positions 13206 and 13227), and the reverse primer 225 (5′-GGAGGATGGAACACGGACGG-3′, matching nucleotides 13353–13372). The position of the $(CTG)_{11}$ repeat is assigned nucleotide positions 13230–13262. Genomic DNA (100 ng) was PCR amplified in a mixture containing 100 ng of each primer by a standard protocol (34 cycles at 94°C for 1 min, 58°C for 1 min and 72°C for 2 min) and amplified products were electrophoretically resolved on agarose gels. The PCR yields a 166 bp band for the humanized $(CTG)_{11}$ allele and a 385 bp band for the humanized $(CTG)_{84}$ allele as originally cloned. For more precise repeat length estimates PCR was performed with primer 226 and a Cy-5 fluorescently labelled primer 225. Amplified products were mixed with formamide loading buffer, heated for 10 min at 96°C and subjected to electrophoresis on 6% denaturing acrylamide–urea gels in the ALF DNA sequencing system (Amersham Pharmacia Biotech, Uppsala, Sweden). To determine the number of (CTG) repeats, the highest peak in the trace of the $(CTG)_{84}$ plasmid was used as a reference point. Somatic instability was analysed by comparing peak area distributions in repeat profiles. Intergenerational instability was studied by comparing the lengths of the main peak(s) in profiles of individual tail DNA products. For repeat sequence verification genomic DNA samples were PCR amplified using the forward primer 14 (5′-GCCAATGACGAGTTCGGA-3′, matching nucleotides 12674–12691) and the reverse primer 227 (5′-TTTGCACTTTGCGAACCAAC-3′, matching nucleotides 13390–13409). The sequencing reaction was performed with the fluorescently labelled forward primer DM-A [5′-CAGTTCACAACCGCTCCGAGC-3′ (32), matching nucleotide positions 13091–13111] and DNA sequence determined on the ALF system.

Accession numbers

Numbering of sites is according to GenBank database accession nos L08835 and Z38015 for the human and mouse *DMPK* genes, respectively.

ACKNOWLEDGEMENTS

We thank Dr P.Krimpenfort, Netherlands Cancer Institute, Amsterdam for help and advice with the Cre/loxP experiments. This study was supported by grants from the American Muscular Dystrophy Association (MDA), the Association Française contre Les Myopathies (AFM), the Prinses Beatrix Fonds (grant 98-0102) and the Dutch Cancer Society (grants NKI 2000-2233 and KUN 99-1915).

REFERENCES

1. Harper,P.S. (1989) *Myotonic Dystrophy*. W.B. Saunders Co., London.
2. Meola,G. (2000) Clinical and genetic heterogeneity in myotonic dystrophies. *Muscle Nerve*, **23**, 1789–1799.
3. Boucher,C.A., King,S.K., Carey, N.,Krahe,R., Winchester,C.L., Rahman,S., Creavin,T., Meghji,P., Bailey,M.E., Chartier,F.L. *et al.* (1995) A novel homeodomain-encoding gene is associated with a large CpG island interrupted by the myotonic dystrophy unstable $(CTG)_n$ repeat. *Hum. Mol. Genet.*, **4**, 1919–1925.
4. Jansen,G., Bachner,D., Coerwinkel,M., Wormskamp,N., Hameister,H. and Wieringa,B. (1995) Structural organization and developmental expression pattern of the mouse WD-repeat gene DMR-N9 immediately upstream of the myotonic dystrophy locus. *Hum. Mol. Genet.*, **4**, 843–852.
5. Alwazzan,M., Newman,E., Hamshere,M.G. and Brook,J.D. (1999) Myotonic dystrophy is associated with a reduced level of RNA from the DMWD allele adjacent to the expanded repeat. *Hum. Mol. Genet.*, **8**, 1491–1497.
6. Harris,S., Moncrieff,C. and Johnson,K. (1996) Myotonic dystrophy: will the real gene please step forward! *Hum. Mol. Genet.*, **5**, 1417–1423.
7. Reddy,P.S. and Housman,D.E. (1997) The complex pathology of trinucleotide repeats. *Curr. Opin. Cell Biol.*, **9**, 364–372.
8. Groenen,P. and Wieringa,B. (1998) Expanding complexity in myotonic dystrophy, *Bioessays*, **20**, 901–912.
9. Brook,J.D., McCurrach,M.E., Harley,H.G., Buckler,A.J., Church,D., Aburatani,H., Hunter,K., Stanton,V.P., Thirion,J.P., Hudson,T. *et al.* (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell*, **68**, 799–808.
10. Fu,Y.H., Pizzuti,A., Fenwick R.G.,Jr, King,J., Rajnarayan,S., Dunne,P.W., Dubel,J., Nasser,G.A., Ashizawa,T., de Jong,P. *et al.* (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science*, **255**, 1256–1258.

11. Mahadevan,M., Tsilfidis C., Sabourin L., Shutler,G., Amemiya,C., Jansen,G., Neville,C., Narang,M., Barcelo,J., O'Hoy,K. *et al.* (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. *Science*, **255**, 1253–1255.
12. Jansen,G., Mahadevan,M., Amemiya,C., Wormskamp N., Segers,B., Hendriks,W., O'Hoy,K., Baird,S., Sabourin,L., Lennon,G. *et al.* (1992) Characterization of the myotonic dystrophy region predicts multiple protein isoform-encoding mRNAs. *Nat. Genet.*, **1**, 261–266.
13. Meola,G. (2000) Myotonic dystrophies. *Curr. Opin. Neurol.*, **13**, 519–525.
14. Liquori,C.L., Ricker,K., Moseley,M.L., Jacobsen,J.F., Kress,W., Naylor,S.L., Day,J.W. and Ranum,L.P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*, **293**, 864–867.
15. Tapscott,S.J. (2000) Deconstructing myotonic dystrophy. *Science*, **289**, 1701–1702.
16. Miller,J.W., Urbinati,C.R., Teng-Umnuay,P., Stenberg,M.G., Byrne,B.J., Thornton, C.A. and Swanson, M.S. (2000) Recruitment of human muscle-blind proteins to $(CUG)_n$ expansions associated with myotonic dystrophy. *EMBO J.*, **19**, 4439–4448.
17. Mankodi,A., Logigian,E., Callahan L., McClain,C., White,R., Henderson,D., Krym,M. and Thornton Charles,A. (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science*, **289**, 1769–1772.
18. Wong,L.J., Ashizawa,T., Monckton,D.G., Caskey,C.T. and Richards,C.S. (1995) Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am. J. Hum. Genet.*, **56**, 114–122.
19. Monckton,D.G., Wong,L.J., Ashizawa,T. and Caskey,C.T. (1995) Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum. Mol. Genet.*, **4**, 1–8.
20. Thornton,C.A., Johnson,K. and Moxley,R.T.,III (1994) Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann. Neurol.*, **35**, 104–107.
21. Ishii,S., Nishio,T., Sunohara,N., Yoshihara,T., Takemura,K., Hikiji,K., Tsujino,S. and Sakuragawa,N. (1996) Small increase in triplet repeat length of cerebellum from patients with myotonic dystrophy. *Hum. Genet.*, **98**, 138–140.
22. Martorell,L., Johnson,K., Boucher,C.A. and Baiget,M. (1997) Somatic instability of the myotonic dystrophy $(CTG)_n$ repeat during human fetal development. *Hum. Mol. Genet.*, **6**, 877–880.
23. Lia,A.S., Seznec,H., Hofmann-Radvanyi,H., Radvanyi,F., Duros,C., Saquet,C., Blanche,M., Junien,C. and Gourdon,G. (1998) Somatic instability of the CTG repeat in mice transgenic for the myotonic dystrophy region is age dependent but not correlated to the relative intertissue transcription levels and proliferative capacities. *Hum. Mol. Genet.*, **7**, 1285–1291.
24. Fortune,M.T., Vassilopoulos,C., Coolbaugh,M.I., Siciliano,M.J. and Monckton,D.G. (2000) Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum. Mol. Genet.*, **9**, 439–445.
25. Pearson,C.E., Ewel,A., Acharya,S., Fishel,R. and Sinden,R.R. (1997) Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum. Mol. Genet.*, **6**, 1117–1123.
26. Manley,K., Shirley,T.L., Flaherty,L. and Messer,A. (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.*, **23**, 471–473.
27. Kovtun,I.V. and McMurray,C.T. (2001) Trinucleotide expansion in haploid germ cells by gap repair. *Nat. Genet.*, **27**, 407–411.
28. Acharya,S., Wilson,T., Gradia,S., Kane,M.F., Guerrette,S., Marsischky,G.T., Kolodner,R. and Fishel,R. (1996) hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc. Natl Acad. Sci. USA*, **93**, 13629–13634.
29. Rajewsky,K., Gu,H., Kuhn,R., Betz,U.A., Muller,W., Roes,J. and Schwenk,F. (1996) Conditional gene targeting. *J. Clin. Invest.*, **98**, 600–603.
30. Groenen,P.J., Wansink,D.G., Coerwinkel,M., van den Broek,W., Jansen,G. and Wieringa,B. (2000) Constitutive and regulated modes of splicing produce six major myotonic dystrophy protein kinase (DMPK) isoforms with distinct properties. *Hum. Mol. Genet.*, **9**, 605–616.
31. Brunner,H.G., Bruggenwirth,H.T., Nillesen,W., Jansen,G., Hamel,B.C., Hoppe,R.L., de Die,C.E., Howeler,C.J., van Oost,B.A., Wieringa,B. *et al.* (1993) Influence of sex of the transmitting parent as well as of parental allele size on the CTG expansion in myotonic dystrophy (DM). *Am. J. Hum. Genet.*, **53**, 1016–1023.
32. Jansen,G., Willems,P., Coerwinkel,M., Nillesen,W., Smeets,H., Vits,L., Howeler,C., Brunner,H. and Wieringa,B. (1994) Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in $(CTG)_n$ repeat variation and selection against extreme expansion in sperm. *Am. J. Hum. Genet.*, **54**, 575–585.
33. Lipkin,S.M., Wang,V., Jacoby,R., Banerjee Basu,S., Baxevanis,A.D., Lynch,H.T., Elliott,R.M. and Collins,F.S. (2000) MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat. Genet.*, **24**, 27–35.
34. de Wind,N., Dekker,M., Claij,N., Jansen,L., van Klink,Y., Radman,M., Riggins,G., van der Valk,M., van't Wout,K. and te Riele,H. (1999) HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch–repair protein functions. *Nat. Genet.*, **23**, 359–362.
35. Gourdon,G., Radvanyi,F., Lia,A.S., Duros,C., Blanche,M., Abitbol,M., Junien,C. and Hofmann-Radvanyi,H. (1997) Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice. *Nat. Genet.*, **15**, 190–192.
36. Seznec,H., Lia Baldini,A.S., Duros,C., Fouquet,C., Lacroix,C., Hofmann Radvanyi,H., Junien,C. and Gourdon,G. (2000) Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum. Mol. Genet.*, **9**, 1185–1194.
37. Mangiarini,L., Sathasivam,K., Mahal,A., Mott,R., Seller,M. and Bates,G.P. (1997) Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat. Genet.*, **15**, 197–200.
38. Sato,T., Oyake,M., Nakamura,K., Nakao,K., Fukusima,Y., Onodera,O., Igarashi,S., Takano,H., Kikugawa,K., Ishida,Y. *et al.* (1999) Transgenic mice harboring a full-length human mutant DRPLA gene exhibit age-dependent intergenerational and somatic instabilities of CAG repeats comparable with those in DRPLA patients. *Hum. Mol. Genet.*, **8**, 99–106.
39. Lavedan,C.N., Garrett,L. and Nussbaum,R.L. (1997) Trinucleotide repeats $(CGG)_{22}TGG(CGG)_{43}TGG(CGG)_{21}$ from the fragile X gene remain stable in transgenic mice. *Hum. Genet.*, **100**, 407–414.
40. Leeflang,E.P., Tavare,S., Marjoram,P., Neal,C.O.S., Srinidhi,J., MacDonald,M.E., de Young,M., Wexler,N.S., Gusella,J.F. and Arnheim,N. (1999) Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. *Hum. Mol. Genet.*, **8**, 173–183.
41. Gomes Pereira,M., Fortune,M.T. and Monckton,D.G. (2001) Mouse tissue culture models of unstable triplet repeats: *in vitro* selection for larger alleles, mutational expansion bias and tissue specificity, but no association with cell division rates. *Hum. Mol. Genet.*, **10**, 845–854.
42. de Wind,N., Dekker,M., Berns,A., Radman,M. and te Riele,H. (1995) Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination and predisposition to cancer. *Cell*, **82**, 321–330.
43. Kolodner,R.D. and Marsischky,G.T. (1999) Eukaryotic DNA mismatch repair. *Curr. Opin. Genet. Dev.*, **9**, 89–96.
44. Wells,R.D. (1996) Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem.*, **271**, 2875–2878.
45. Pearson,C.E. and Sinden,R.R. (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry*, **35**, 5041–5053.
46. Aslanidis,C., Jansen,G., Amemiya,C., Shutler,G., Mahadevan,M., Tsilfidis,C., Chen,C., Alleman,J., Wormskamp,N.G., Vooijs,M. *et al.* (1992) Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature*, **355**, 548–551.
47. Bradley,A. (1987) Teratocarcinomas and embryonic stem cells: a practical approach. In Robertson, E.J. (ed.), *Production and Analysis of Chimeric Mice*. IRL Press, Oxford, pp. 113–151.
48. van Deursen,J. and Wieringa,B. (1992) Targeting of the creatine kinase M gene in embryonic stem cells using isogenic and nonisogenic vectors. *Nucleic Acids Res.*, **20**, 3815–3820.