NATURE|Vol 447|21 June 2007|doi:10.1038/nature05977

Expandable DNA repeats and human disease

Sergei M. Mirkin[1]

Nearly 30 hereditary disorders in humans result from an increase in the number of copies of simple repeats in genomic DNA. These DNA repeats seem to be predisposed to such expansion because they have unusual structural features, which disrupt the cellular replication, repair and recombination machineries. The presence of expanded DNA repeats alters gene expression in human cells, leading to disease. Surprisingly, many of these debilitating diseases are caused by repeat expansions in the non-coding regions of their resident genes. It is becoming clear that the peculiar structures of repeat-containing transcripts are at the heart of the pathogenesis of these diseases.

One of the central principles of classical (mendelian) genetics is that mutations are stably transmitted between generations. As long ago as 1918, however, a different type of inheritance was described for a human neurological disorder, myotonic dystrophy[1]. This type of inheritance was characterized by increased expressivity: that is, a decreased age of onset and increased severity in individuals of subsequent generations. A similar hereditary pattern was later observed for other neurological diseases: for example, Huntington's disease, spinal and bulbar muscular atrophy, and several ataxias. The penetrance — that is, the probability that a given mutation results in disease — can also increase in successive generations, as was first demonstrated for fragile X syndrome[2]. This unusual type of inheritance — characterized by a progressive increase in the expressivity and, sometimes, the penetrance of a mutation as it passes through generations — was called genetic anticipation.

Understanding this genetic anomaly became possible when the mutations that result in fragile X syndrome[3,4] and spinal and bulbar muscular atrophy[5] were characterized, which was soon followed by cloning of the gene that causes myotonic dystrophy[6,7]. In all three cases, mutation seemed to arise from the continuous intergenerational expansion of simple trinucleotide repeats — $(CNG)_n$ (where N denotes any nucleotide) — in different human genes. The progressive character of repeat expansion across generations provided a clue about the mechanism of genetic anticipation: the longer a repeat is, the more probable it is that it expands, and the more severe the phenotype. Thus, such mutations are classified as dynamic to account for the perpetual nature of the expansion process.

At present, expansions of simple DNA repeats are implicated in nearly 30 human hereditary disorders, and the list continues to grow (see refs 8 and 9 for recent updates). Various disease-causing repeats are depicted in the context of a fictitious human gene in Fig. 1. Most of these disorders are caused by the expansion of the triplet repeats $(CGG)_n$•$(CCG)_n$, $(CAG)_n$•$(CTG)_n$, $(GAA)_n$•$(TTC)_n$ and $(GCN)_n$•$(NGC)_n$. But disease can also result from the expansion of the tetranucleotide repeat $(CCTG)_n$•$(CAGG)_n$, the pentanucleotide repeat $(ATTCT)_n$•$(AGAAT)_n$, and even the dodecanucleotide repeat $(C_4GC_4GCG)_n$•$(CGCG_4CG_4)_n$. For a given hereditary disorder, only one repeat expands in a particular gene, strongly indicating that the molecular events leading to repeat expansions occur *in cis*. Expandable repeats can be located in various regions of their resident genes: first, the coding regions, as occurs in numerous diseases mediated by polyglutamine or polyalanine runs in proteins; second, the 5′ untranslated regions (5′-UTRs), as in the case of fragile X syndrome, fragile X mental retardation associated with the *FRAXE* site, fragile X tremor and ataxia syndrome, and spinocerebellar ataxia 12; third, 3′-UTRs, as is observed for myotonic dystrophy 1, spinocerebellar ataxia 8 and Huntington's-disease-like 2; fourth, introns, as in the case of myotonic dystrophy 2, Friedreich's ataxia and spinocerebellar ataxia 10; and fifth, promoter regions, as occurs in progressive myoclonic epilepsy 1.

Normal alleles of the genes associated with expansion-mediated diseases mostly contain either very short repetitive runs ('short-normal' alleles) or longer runs with several stabilizing interruptions ('long-normal' alleles): for example, AGG inserts within $(CGG)_n$ runs in the gene associated with fragile X syndrome; CAT inserts within $(CAG)_n$ runs in the gene associated with spinocerebellar ataxia 1; or GAG inserts within $(GAA)_n$ runs in the gene associated with Friedreich's ataxia. Expansions begin when the length of an uninterrupted repetitive run exceeds a threshold of ~100–150 bases, often as a result of the loss of stabilizing interruptions at the end of the repetitive run in long-normal alleles[10,11]. After this threshold is overcome, further expansions become progressively more likely, leading to the accumulation of dozens of repeats (for those that encode polyglutamine) to thousands of repeats (for those in non-coding regions) in just a few generations. Polyalanine-coding repeats behave differently from other expandable repeats[12]. They are encoded by the imperfect triplet $(GCN)_n$. The threshold length for their expansion is extremely low (30–60 bases), and they rarely expand more than 1.5-fold. In addition, expanded $(GCN)_n$•$(NGC)_n$ repeats are stable during both intergenerational and somatic transmission. Thus, progressive repeat lengthening, which is responsible for genetic anticipation, is not observed in polyalanine-mediated disorders.

This review concentrates on two questions. First, what are the mechanisms of repeat expansion? Studies carried out during the past decade suggest that expandable repeats are predisposed to instability, as a result of 'confusion' between the DNA replication, repair and recombination machineries (see refs 9, 13 and 14 for recent reviews). Second, how do repeat expansions result in disease? This question is particularly intriguing when considering the diseases that are caused by repeat expansions in the non-coding regions of human genes. It is becoming increasingly clear that a toxic 'gain of function' at the RNA level (see ref. 15 for a review) could be responsible. Here, I argue that the unusual structural characteristics of repetitive DNA and RNA, respectively, are central to these two issues.

[1]Department of Biology, Tufts University, Medford, Massachusetts 02155, USA.

Molecular mechanisms of repeat expansions

Unusual structures of repetitive DNA

The first molecular model of how repeat expansions occur was based on DNA strand slippage during replication[16]. Looping out one or several repeats in the newly synthesized DNA strand should convert the loop into expansions after a second round of replication. However, this simple idea fails to explain why only a handful of all repeats expand and what determines the threshold length and the large-scale character of the expansions. The next breakthrough was the realization that all expandable repeats have unusual structural characteristics (see refs 13 and 17 for reviews). Single-stranded $(CNG)_n$ repeats form hairpin-like structures that consist of both Watson–Crick base pairs and mismatched base pairs[18] (Fig. 2a). In physiological conditions, the stability of such imperfect hairpins decreases according to the sequence of the triplet, in the order CGG>CTG>CAG=CCG, as a consequence of the energy contribution of the mismatched base pairs. Individual strands of $(CCTG)_n$•$(CAGG)_n$ repeats have also been shown to fold into hairpin-like structures[19]. In addition to hairpins, single-stranded $(CGG)_n$, $(CCG)_n$ and $(CGCG_4CG_4)_n$ repeats can fold into tetrahelical structures[20,21] stabilized by intertwining G quartets and i motifs (Fig. 2b).

The denaturation and renaturation of double-stranded DNA fragments that contain expandable repeats promote the formation of the 'slipped-stranded' DNA conformation. In this case, an out-of-register realignment of the complementary repetitive strands gives rise to 'slip-outs' that are folded into hairpin-like structures (Fig. 2c). These hairpins kinetically 'trap' repetitive DNA in the otherwise unfavourable slipped-stranded configuration[22]. Owing to the difference in hairpin-forming potential between expandable repeats in the complementary strands, slipped-stranded DNA is intrinsically asymmetrical. For example, when the $(CTG)_n$•$(CAG)_n$ repeat converts into the slipped-stranded form, CAG slip-outs are mainly in the random-coil state, whereas CTG slip-outs are in the hairpin state[23]. This asymmetry has important biological implications, because one of the complementary repetitive strands is usually more structure-prone than the other.

Slipped-stranded structures are not the only unusual structures formed by expandable repeats within double-stranded DNA. A $(GAA)_n$•$(TTC)_n$ (homopurine–homopyrimidine) repeat can convert into an intramolecular triplex called H-DNA under the influence of negative supercoiling[24] (Fig. 2d). A different structure associated with the longer forms of this repeat is called sticky DNA[25]. The main element of sticky DNA is a composite triplex (Fig. 2d), which is formed by the two distant, directly repeated $(GAA)_n$•$(TTC)_n$ runs within circular DNA[26]; the exact configuration of the fourth repetitive strand remains to be elucidated. Lastly, an (A+T)-rich repeat that is responsible for spinocerebellar ataxia 10, $(ATTCT)_n$•$(AGAAT)_n$, belongs to the class of DNA-unwinding elements (Fig. 2e); that is, it unwinds progressively with increasing negative superhelical stress[27].

DNA replication models

It is generally thought that the unusual structural features of expandable repeats predispose them to instability. Indeed, repeats that are not structure-prone are considerably more stable genetically[18,25,28]. Furthermore, the stabilizing effect of interruptions within the repetitive run in long-normal alleles is probably a result of their destabilizing effect on these unusual DNA structures[18,29]. This has led to the idea that a misalignment between the two repetitive strands during DNA replication, further stabilized by unusual conformations of repetitive slip-outs (Fig. 3a), is the basis of repeat instability. After the next round of replication, either expansions or contractions occur, depending on the origin of the slipped-out strand[30]. These slipped-stranded intermediates can be formed in the course of genetic processes that involve the separation of DNA strands: for example, DNA replication, repair and recombination. At present, each of these processes has been implicated in repeat expansions in one or another experimental system. It should be noted, however, that these unusual structures would be only transient intermediates during those processes, making their direct detection challenging.

Many models for repeat expansions assume that they occur during DNA replication for two main reasons. First, rapid accumulation of repetitive DNA cannot be explained without synthesis of massive amounts of DNA. Second, during the progression of the replication fork, a portion of the lagging-strand template that is known as the Okazaki initiation zone (OIZ) is always single stranded. The appearance of a repetitive run within this region could facilitate its folding into an unusual secondary structure.

Studies of DNA synthesis of expandable repeats *in vitro* support these ideas. Unusual DNA structures that are formed by expandable repeats during DNA synthesis *in vitro* stall various DNA polymerases[20,24]. Occasionally, this stalling results in the misalignment of repetitive DNA strands, causing repeat expansions or contractions[31].

In vivo, data from bacterial, yeast and mammalian cells show that the stability of expandable repeats depends substantially on their orientation relative to replication origins[30,32–34]. Although these studies differ in important experimental details and interpretations, they generally agree

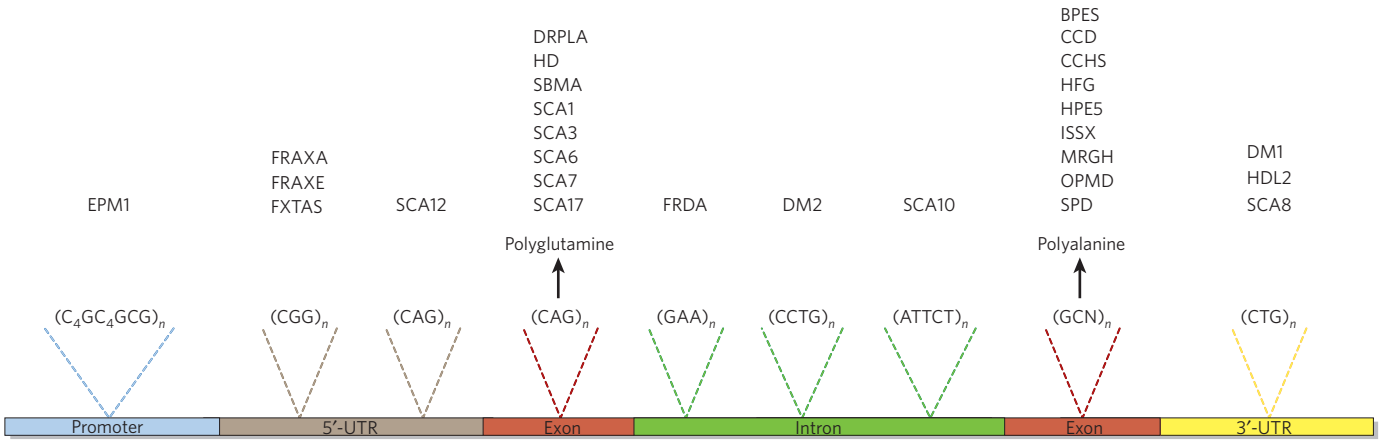

Figure 1 | Location of expandable repeats responsible for human diseases. The sequence and location within a generic gene of expandable repeats that cause human diseases are shown, and the associated diseases are listed. BPES, blepharophimosis, ptosis and epicanthus inversus; CCD, cleidocranial dysplasia; CCHS, congenital central hypoventilation syndrome; DM, myotonic dystrophy; DRPLA, dentatorubral–pallidoluysian atrophy; EPM1, progressive myoclonic epilepsy 1; FRAXA, fragile X syndrome; FRAXE, fragile X mental retardation associated with *FRAXE* site; FRDA, Friedreich's ataxia; FXTAS, fragile X tremor and ataxia syndrome; HD, Huntington's disease; HDL2, Huntington's-disease-like 2; HFG, hand–foot–genital syndrome; HPE5, holoprosencephaly 5; ISSX, X-linked infantile spasm syndrome; MRGH, mental retardation with isolated growth hormone deficiency; OPMD, oculopharyngeal muscular dystrophy; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; SPD, synpolydactyly.

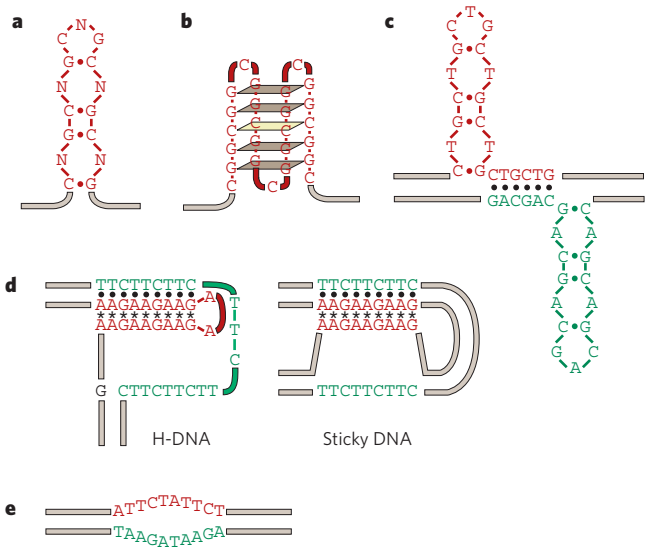

Figure 2 | Unusual DNA structures formed by expandable repeats. Repetitive DNA can form several unusual structures, examples of which are shown. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige. **a**, An imperfect hairpin formed by $(CNG)_n$ repeats. **b**, A quadruplex-like structure formed by the $(CGG)_n$ repeat. Brown rectangles indicate G quartets, and the yellow rectangle indicates an i motif. **c**, A slipped-stranded structure formed by the $(CTG)_n$•$(CAG)_n$ repeat. **d**, H-DNA and sticky DNA formed by the $(GAA)_n$•$(TTC)_n$ repeat. Only one possible isoform, in which the homopurine strand is donated to the triplex, is shown for both structures. Reverse Hoogsteen pairing is indicated by asterisks. **e**, A DNA-unwinding element formed by the $(ATTCT)_n$•$(AGAAT)_n$ repeat.

that instability is the most marked when the structure-prone strand of a repetitive run functions as the lagging-strand template. The stability of repeats also depends on their distance from the replication origin[34,35]. That is, the functioning of the structure-prone repetitive strand as the lagging-strand template and the precise location of the repetitive run in the OIZ are important factors in determining repeat instability.

Strong support for the role of DNA replication in repeat expansions came from studying yeast replication mutants. The frequencies of repeat expansions and contractions are affected markedly by mutations in several genes that encode proteins involved in replication: flap endonuclease (Fen1; also known as Rad27), DNA polymerase-δ, proliferating cell nuclear antigen, the large subunit of the clamp-loading complex, the helicase Srs2 (also known as Hpr5), and several other genes (see ref. 8 for a review). These proteins are involved in the synthesis of the lagging strand, the coordination between leading- and lagging-strand synthesis, and/or the restarting of the replication fork. Some of these mutations, most notably deletion of *FEN1*, destabilize various microsatellites and minisatellites. By contrast, others, such as deletion of *SRS2*, affect only expandable repeats[36]. Recent genetic evidence from yeast studies suggests that the helicase Srs2 inhibits repeat expansions at a stage of post-replicative repair[37]. It can therefore be concluded that some repetitive DNA intermediates, left behind after the replication fork passes on, can be converted into expansions if they remain unrepaired.

This view is in accord with direct observations of anomalous replication fork progression through expandable repeats in prokaryotic and eukaryotic cells[38–40]. Various expandable repeats were found to stall the replication fork in all systems that have been studied. In almost all cases, replication stalling was evident when the length of the repetitive run approached the expansion threshold, and it was more marked when the structure-prone strand of the repetitive run was part of the lagging-strand template. In the region of the stall site, the lagging strand seemed to be under-replicated[38], implicating problems with the lagging-strand synthesis. Finally, repeats were particularly unstable in the orientations that were associated with replication stalling.

Together, these data led to the replication model for repeat instability (Fig. 3b), which is based on stalling and restarting of the replication fork (see ref. 14 for a review). In brief, formation of a stable secondary structure by a repetitive run in the lagging-strand template stalls lagging-strand synthesis and disrupts coordination with leading-strand synthesis. This could lead to synthesis continuing on the leading strand alone, while synthesis of the lagging strand resumes after skipping one or more Okazaki fragments, leaving a gap in the nascent lagging strand. Repeat contractions occur if a DNA polymerase involved in repair of the gap skips the structured portion of the lagging-strand template (Fig. 3b, upper pathway). Alternatively, replication stalling within a repetitive run can trigger replication fork reversal[41]. This would create a peculiar 'chicken-foot' structure with a single-stranded repetitive extension in the nascent leading strand, which can easily fold into a hairpin-like conformation. When the reversed replication fork is flipped back to restart replication, extra repeats can be added to the leading strand (Fig. 3b, lower pathway).

This model can account for several genetic features of repeat expansions at a molecular level. First, formation of unusual secondary structures in the lagging-strand template is more likely as the length of the repetitive run becomes comparable with the size of an OIZ. Thus, the similarity in expansion thresholds for various repeats might simply reflect the average size of the eukaryotic OIZ (~200 bases). Second, genetic anticipation could be explained in terms of consecutive replication stalls and restarts within longer repetitive runs, which would progressively increase their instability. Last, the disparity in the propensities of repeats to expand or contract in various model organisms could be explained by the differential probability of fork reversal and fork bypass in those organisms.

DNA-repair models

The unwinding of DNA during repair and recombination processes can also lead to the formation of slipped-stranded structures, which are implicated in repeat expansions (as mentioned earlier). The role of mismatch repair (MMR) in repeat instability has attracted particular attention. This is largely a result of studies using transgenic mouse models of Huntington's disease and myotonic dystrophy. In these mice, mutational inactivation of the gene encoding MSH2 (MutS homologue 2) or MSH3 markedly decreased the frequency of repeat expansions during intergenerational transmission and in non-dividing somatic cells, shifting the pattern of repeat instability towards contractions[42–45]. A heterodimer of MSH2 and MSH3 normally repairs single-base insertions and small loop-outs formed during DNA replication. So how could this complex promote, rather than prevent, repeat expansions? One possibility is that MSH2–MSH3 has an affinity for repetitive hairpins because these structures contain numerous mismatches. This interaction could therefore sequester MSH2–MSH3 to stabilize slipped-stranded intermediates instead of repairing them[45]. In support of this idea, binding of MSH2–MSH3 to repetitive hairpins *in vitro* leads to discoordination of ATP hydrolysis and hairpin stabilization rather than to repair[46].

How could this hijacking of MMR result in repeat expansions? Repeat expansions that occur during intergenerational transmission in humans and transgenic mice seem to happen in dividing cells about to undergo meiosis[43,47,48]. As discussed earlier, replication fork stalling and restarting within a repetitive run might lead to the formation of a hairpin-like slip-out in the nascent DNA strand (Fig. 3b, bracketed intermediate). MSH2–MSH3 can be tricked into binding this hairpin through its similarity to mismatched DNA. Because the MMR machinery is thought to operate on nascent DNA strands, repetitive hairpins present in these strands would be stabilized preferentially over repetitive structures in template strands, shifting the equilibrium towards repeat expansions.

Repeat expansions are also observed in tissues in which cells do not divide, such as brain and skeletal muscle tissue in humans and mice[49–52]. These events also require a functional MMR system[42–45], a somewhat unexpected finding given the lack of DNA replication in these cells. Furthermore, repeat expansions depend on the oxidative damage of

repetitive DNA in ageing non-dividing cells. Although expansions can occur during the repair of DNA nicks or gaps generated directly by oxygen radicals[45], recent studies using a transgenic mouse model of Huntington's disease have implicated base-excision repair as central to this process[53]. Strikingly, in these mice, age-dependent repeat expansions in somatic cells depended on a single base-excision repair enzyme, 8-oxoguanine DNA glycosylase (OGG1) (see page 941). Removal of an oxidized guanine by OGG1 generates a nick in the repetitive run (Fig. 4a), and DNA-repair synthesis is then needed to heal this lesion. During this repair synthesis, the non-template DNA strand is displaced, forming a flap (Fig. 4b). Normally, a flap is removed by FEN1; this endonuclease is loaded onto the 5′ end of the flap, migrates to the junction with the duplex, and cleaves the flap. If a flap contains the structure-prone strand of a repetitive run, however, it can fold into a hairpin-like conformation, complicating FEN1 loading[54,55]. MSH2–MSH3 can further stabilize this hairpin (as discussed earlier), preventing flap removal (Fig. 4c). Completion of the repair process will yield a stable slipped-stranded DNA intermediate with a repeat extension in the nicked strand (Fig. 4d).

Can these unprocessed flaps be converted into expansions in non-dividing cells? This was indeed the case in a study of the fate of various slipped-stranded DNA intermediates in extracts of terminally differentiated neuron-like cells[56]. Intermediates with repetitive slip-outs in the nicked DNA strand (Fig. 4d) were repaired to become a set of products with differentially expanded repeats (Fig. 4e), as though the hairpin was incompletely excised during the repair synthesis. This 'error-prone' repair might therefore account for the final stages of repeat expansions in non-dividing cells.

Unprocessed flap structures can also contribute to repeat expansions in dividing cells, providing an elegant explanation of the bias towards repeat expansions that has been observed in human pedigrees. In support of this, inefficient flap removal leads to repeat expansions in yeast[57]. The situation is less clear in mice, however, because $(CTG)_n$ repeat stability was essentially unaltered in *Fen1*-knockout mice[58].

DNA recombination models

Various pathways of genetic recombination can also contribute to repeat instability. The simplest mechanism could be unequal crossing-over between the repetitive runs on homologous chromosomes during meiosis, resulting in reciprocal expansions and contractions (Fig. 5a). This process has been implicated in expansions of $(GCN)_n$ repeats, which encode polyalanine[59]. For polyalanine-mediated disorders, every allele in a given pedigree has sequence variations due to the redundancy in the third position of the repeat. Tracking these variations in alleles containing repeat expansions leads to the conclusion

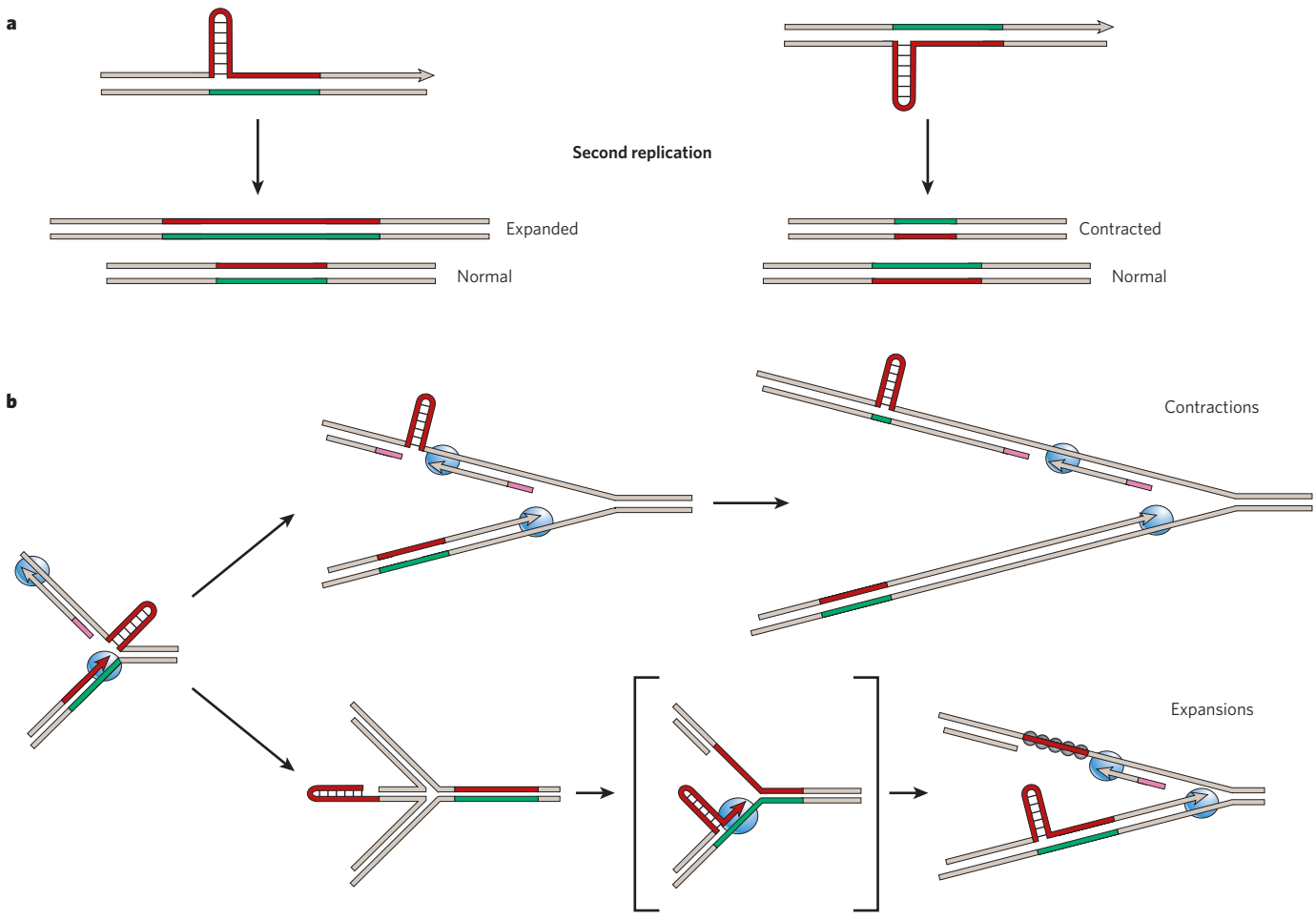

Figure 3 | Replication mechanisms for repeat expansion. a, After two rounds of replication, formation of a repetitive hairpin on the nascent strand results in repeat expansions (left panel), whereas the presence of the same structure on the template strand results in repeat contractions (right panel). **b**, A model for repeat instability based on replication fork stalling and restarting within the repetitive run is shown. Repeat contractions (upper pathway) occur when the machinery for the lagging-strand synthesis skips the repetitive hairpin on the lagging-strand template. Repeat expansions (lower pathway) can occur during replication fork reversal and restart, leading to the formation of a repetitive hairpin on the nascent leading strand. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige. DNA polymerases are shown in blue, primers for Okazaki fragments in pink, and single-stranded-DNA-binding proteins as grey circles. The bracketed intermediate contains a hairpin on the nascent strand, which can also be stabilized by MSH2–MSH3.

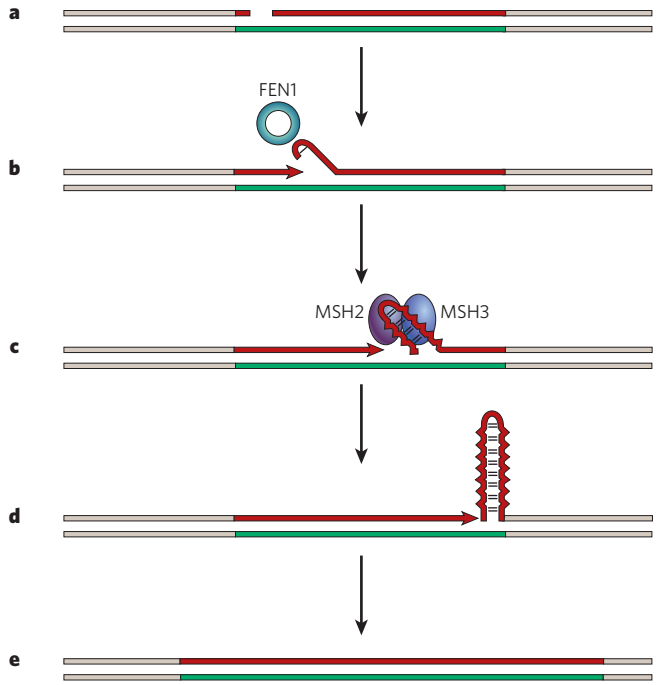

Figure 4 | Gap repair model for repeat expansions in non-dividing cells. **a**, Oxidizing radicals generate a small gap in the structure-prone strand of a repetitive run. **b**, The loading of FEN1 onto a repetitive flap generated during the DNA-repair synthesis is impaired by hairpin formation. **c**, The binding of MSH2–MSH3 stabilizes the repetitive hairpin, preventing flap removal. **d**, A stable slipped-stranded DNA intermediate is formed on the completion of the repair synthesis. **e**, The slipped-stranded intermediate is converted into an expansion by an error-prone repair pathway. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige.

that the repeat expansions resulted from unequal crossing-over between $(GCN)_n$ blocks in normal alleles.

Meiotic crossing-over has, however, been ruled out as the source of instability for all other repeats, because these expansions are not accompanied by an exchange of the flanking markers[60]. This leaves mitotic recombination as the most plausible source. Indeed, expandable repetitive runs stimulate recombination in mitotically dividing cells, undergoing length changes during this process. In bacteria, $(CAG)_n$ (ref. 61), $(GAA)_n$ (ref. 62) and $(CCTG)_n$ (ref. 63) repeats increase the rate of both inter- and intramolecular plasmid recombination. This increase is proportional to the length of the repetitive run, being particularly marked when a structure-prone repetitive strand functions as the lagging-strand template, pointing to the role of DNA replication (discussed later). Furthermore, expanded and contracted versions of these repeats were frequently detected among the recombination products. In yeast, long $(CAG)_n$•$(CTG)_n$ runs were shown to cause chromosomal breakage, triggering ectopic recombination[64]. They also stimulated spontaneous unequal sister-chromatid exchange[65] and underwent frequent expansions and contractions during gene conversion[66,67]. Finally, in mitotically dividing mammalian cells, $(CAG)_n$•$(CTG)_n$ repeats were shown to stimulate homologous recombination, undergoing contractions and other rearrangements during this process[68].

To account for the recombinogenic activity of expandable repeats, it is safe to assume that, in one way or another, the repeats trigger the formation of double-strand breaks in DNA. Notably, data from bacteria and yeast show that stimulation of recombination depends on the orientation of repeats within a replicon, pointing to the connection between their replication and recombination. One possible mechanism of such a connection is presented in Fig. 5b. As discussed earlier, the formation of a stable secondary structure in the lagging-strand template by an expandable repeat leads to stalling of the replication fork. Cleavage and processing of this structure by an unidentified eukaryotic endonuclease — a functional homologue of the bacterial nuclease SbcCD — would generate a DNA fragment with a single-stranded 3′ repetitive extension capable of invading the sister chromatid. An out-of-register invasion would then lead to repeat expansions or contractions (Fig. 5b). Another plausible recombinogenic intermediate could be the product of fork reversal (Fig. 3b), which contains a single-stranded repetitive run at the 3′ end of the leading strand.

It is worth noting that single-stranded repetitive tails (discussed earlier) can also invade homologous repeats at ectopic positions. Subsequent recombination events would lead to chromosomal rearrangements, such as translocations and deletions. Expandable repeats have indeed been shown to induce gross chromosomal rearrangements (see ref. 69 for a review).

Initial stages

All of the molecular models described here considered repeats beyond the expansion threshold. But what happens at the initial stages, when a normal allele converts into the pre-mutation allele capable of expanding in future generations? These early events could involve either the accidental lengthening of short-repetitive runs in normal alleles or the loss of stabilizing interruptions in long-normal alleles. Most of the data point to the second possibility.

Two prospective mechanisms leading to the loss of stabilizing interruptions are presented in Fig. 6. An interruption that has slipped out of the template strand during DNA replication would simply be lost if this slippage were unrepaired (Fig. 6a). This simple model cannot, however, account for two observations: first, more than one interruption is usually lost en route from the long-normal allele to its expandable counterpart; and, second, expansions in genetic carriers usually occur at one end of a repetitive run[10,11]. An elegant explanation that has emerged from yeast studies is presented in Fig. 6b. Expandable repeats in yeast are stabilized by interruptions, as is the case in humans. This stable maintenance seems to require the MMR system, because knockdown of expression of the genes involved in MMR led to the frequent loss of interruptions. This stabilization is explained by the 'co-excision' mechanism[70]. Formation of a slipped-stranded intermediate during replication of a repetitive run that contains several interruptions would place these interruptions out of register in both the hairpin and the duplex part of this intermediate. Co-excision of the hairpin and the mismatches in a duplex part of such an intermediate is needed to maintain the original sequence of the repeat. Failure to do so would generate a non-interrupted expansion at one end of the repeat after another round of replication.

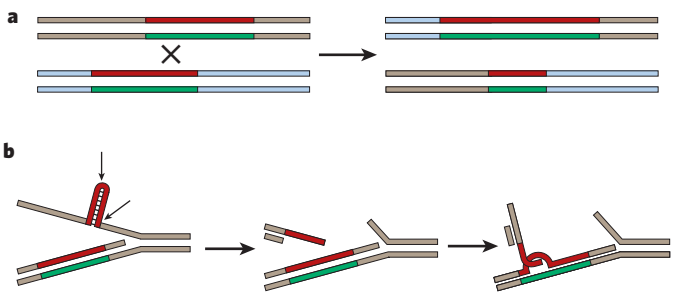

Figure 5 | Recombination models for repeat expansions. **a**, Unequal crossing-over results in the reciprocal appearance of expanded and contracted repeats. Homologous chromosomes are shown in beige and blue, and repetitive DNA strands are shown in red and green. **b**, Cleavage of a stable DNA structure on the lagging-strand template (left), formed during DNA replication, generates a single-stranded 3′ repetitive extension (centre). Out-of-register invasion of such an extension into a sister chromatid (right) might lead to repeat expansions or contractions. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige. Small arrows show potential cleavage sites.

As discussed earlier, formation of slipped-stranded replication intermediates depends on the mode of replication fork progression through the repeat: that is, its orientation relative to the replication origin, as well as its exact position in the lagging-strand template. These considerations led to three hypotheses, linking early expansion stages with the position and orientation of the repetitive run in the replication unit. First, a hypothesis known as 'ori-switch'[71] suggests that inactivation of the replication origin on one side of a repeat, combined with the activation of a cryptic origin on its other side, triggers expansions by placing the structure-prone strand of the repetitive run as the lagging-strand template (Fig. 7a). Second, a hypothesis known as 'ori-shift'[71] assumes that early stages of expansions depend on the position of a repeat within the OIZ; therefore, a change in the distance between the replication origin and the repeat (caused by insertion of a mobile element, for example) could induce expansions (Fig. 7b). Third, a hypothesis known as the fork-shift model[72] proposes that a change in the mode of replication fork progression, caused by an epigenetic event in the vicinity of the repeat, could alter the position of this repeat within the OIZ, leading to expansions (Fig. 7c). These models remain to be substantiated by fine analysis of replication fork progression through expandable repeats in cells from normal and affected individuals. However, a particularly intriguing possibility is that the predicted changes could depend on the developmental or tissue-specific mode of origin usage in mammals.

Molecular mechanisms of repeat-mediated RNA toxicity

Most diseases associated with repeat expansions show dominant inheritance. Classical genetics explains dominance through the effects of mutations on protein function. Those effects include loss of function, leading to haploinsufficiency, and gain of function, when a mutant protein has a dominant-negative effect on its normal counterpart or acquires a novel, deleterious function. This explanation seems to account for diseases caused by repeat expansions in coding sequences. In the case of polyglutamine- or polyalanine-mediated disorders, mutant proteins acquire a deleterious ability to aggregate, which might trigger cell death and tissue degeneration (see ref. 73 for a review).

Unexpectedly, however, the expression of expanded repeats in non-coding sequences also gives rise to dominant mutations (see ref. 15 for a review). This is true for myotonic dystrophy 1 and 2, spinocerebellar ataxia 8, 10 and 12, Huntington's-disease-like 2, and fragile X tremor and ataxia syndrome. These diseases are caused by various tri-, tetra- and pentanucleotide repeats, which are situated in the various non-coding regions (5′-UTRs, 3′-UTRs and introns) of their resident genes (Fig. 1). The scale of expansions also differs between these diseases, from as many as 11,000 repeats for myotonic dystrophy 2 to fewer than 100 repeats for Huntington's-disease-like 2 and spinocerebellar ataxia 8.

Recent progress in the field indicates that these dominantly inherited diseases could be caused by gain of function at the RNA level. Originally, this idea came from studies of myotonic dystrophy. First, transcripts of *DMPK* (which encodes myotonic dystrophy protein kinase 1) containing expanded CUG repeats were shown to be retained in the nuclei of fibroblasts and myoblasts, forming distinct foci[74]. Second, the ability of normal myoblasts to undergo myogenic differentiation in cell culture seemed to be suppressed by overexpression of RNA containing a $(CUG)_{200}$ repeat[75]. Most strikingly, a transgenic mouse expressing $(CUG)_{250}$ repeats within the 3′-UTR of a heterologous gene (skeletal actin) showed major symptoms of myotonic dystrophy[76].

Similar, somewhat less compelling, observations have begun to accumulate for other diseases. In patients with fragile X tremor and ataxia syndrome, transcripts containing intermediate-size $(CGG)_n$ repeats are overproduced, forming intranuclear inclusions in the neurons and astrocytes[77]. Furthermore, overexpression of $(CGG)_n$-containing transcripts leads to the appearance of intranuclear inclusions, together with neurodegeneration, in mice and *Drosophila melanogaster*[78]. A role for 'toxic' RNA is also favoured as a possibility in the development of spinocerebellar ataxia 8, because transcription of the human *SCA8* gene containing an expanded $(CTG)_n$ repeat in the retina of *D. melanogaster* causes its neurodegeneration[79]. In addition, overexpression of the repeat associated with spinocerebellar ataxia 10 leads to the formation of intranuclear foci containing $(AUUCU)_n$ transcripts in cell culture[80]. Finally, there are promising data that point to the existence of similar mechanisms for the development of spinocerebellar ataxia 12 and Huntington's-disease-like 2.

RNA gain-of-function effects could be grounded in the unusual structural features of repeat-containing RNAs (Fig. 8A). The earliest support for this idea came from the observation that $(CUG)_n$ repeats in the natural sequence context of the *DMPK* transcript formed imperfect, mismatched hairpins, the stability of which increased with the length of the repeat[81]. This study was subsequently extended to all 20 possible triplet repeats in RNA[82]. It seems that six repetitive motifs — $(CGU)_n$, $(CGA)_n$, $(CAG)_n$, $(CUG)_n$, $(CCG)_n$ and $(CGG)_n$ — can form stable RNA hairpins. The stability of RNA hairpins that consist of $(CNG)_n$, which are implicated in diseases associated with repeat expansions, was found to depend on the nature of mismatched base pairs, decreasing in the order of CGG>CUG>CCG>CAG. A tetranucleotide repeat involved in myotonic dystrophy 2, $(CCUG)_n$, was also shown to form an RNA hairpin with twice as many mismatches. Finally, a pentanucleotide repeat responsible for spinocerebellar ataxia 10, $(AUUCU)_n$, folded into an unusual RNA hairpin-like structure stabilized by non-Watson–Crick A•U and U•U base pairs[83]. Interestingly, stabilizing interruptions, such as AGG inserts within long-normal $(CGG)_n$-containing alleles, led to the formation of branched RNA hairpin structures[84] (Fig. 8A, c). It

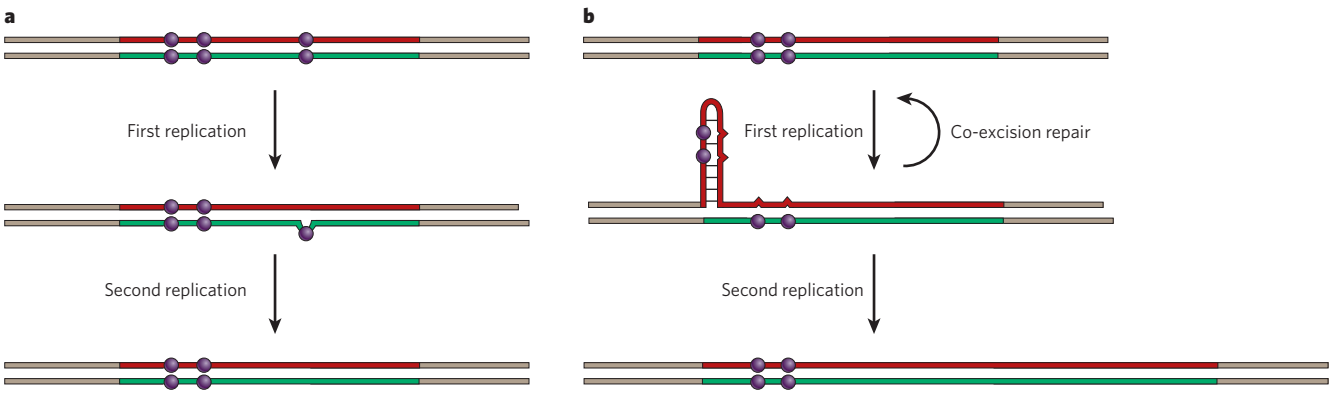

Figure 6 | Loss of stabilizing interruptions within expandable repeats. Two possible mechanisms for the loss of stabilizing interruptions. **a**, An interruption (purple circle) that has slipped out of the template strand (green) is lost during DNA replication. **b**, Misalignment of nascent and template DNA strands in long-normal alleles creates mismatches in both the hairpin and duplex part of the slipped-stranded structure during replication. These mismatches can be repaired by co-excision repair. Failure to repair them leads to expansions at the 3′ end of the repetitive run. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige.

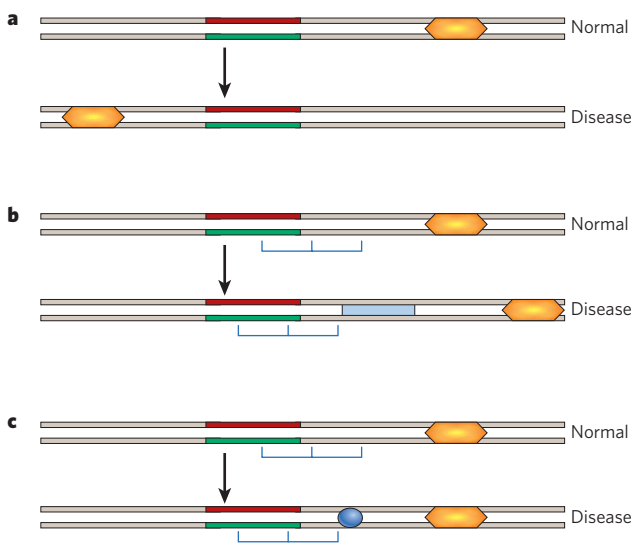

Figure 7 | Three models of early events in repeat expansions. Three hypotheses have been put forward to explain early events in repeat expansions: the ori-switch model, the ori-shift model and the fork-shift model. **a**, The ori-switch model proposes that the initial event leading to repeat expansion is a reversal in the direction of replication (replication origin shown in orange) through the repetitive run, so the structure-prone strand of repetitive run becomes the lagging-strand template. **b**, The ori-shift model assumes that a change in the distance between the replication origin and the repetitive run (for example, as a result of insertion of a mobile element; blue rectangle) changes the position of the repetitive run within the OIZ (consecutive OIZs shown as blue brackets), eliciting repeat expansions. **c**, The fork-shift model proposes that an expansion-prone change in the position of the repetitive run within the OIZ is triggered by an epigenetic event (blue oval) that does not affect the orientation of a repetitive run or its distance from the replication origin. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige.

was therefore suggested that this branching precludes repetitive RNA encoded by long-normal alleles from becoming toxic.

How these RNA structures can cause gain of function and disease remains to be determined. The most intensively discussed potential mechanisms can be referred to as protein sequestration, RNA degradation and chromatin silencing (Fig. 8B). Although there is considerable experimental evidence for the protein sequestration model, the other two mechanisms are only beginning to be tested.

The main support for the protein sequestration model comes from studies of myotonic dystrophy 1 and 2. Intranuclear foci that are characteristic of these diseases contain at least seven RNA-binding proteins associated with $(CUG)_n$- and $(CCUG)_n$-containing transcripts, including three Muscleblind-like (MBNL)-family proteins[85,86] and two different CUG RNA-binding proteins (CUG-BPs)[87]. It was suggested that sequestration of these proteins could lead to RNA gain of function if these proteins are required for the normal expression of muscle-, heart- and brain-specific genes. It is becoming increasingly clear that a key molecular event leading to myotonic dystrophy is the deregulation of alternative RNA splicing during development[88]. At least 13 splicing events are disturbed in muscle, heart and brain tissues from patients with myotonic dystrophy, and an embryonic 'blueprint' for splicing is almost always retained at the expense of the adult splicing pattern (see ref. 89 for a review). Both MBNL1 and CUG-BP1 are implicated in these splicing events, and the splicing pattern characteristic for myotonic dystrophy is consistent with the loss of MBNL1 and gain of CUG-BP1 activities[88,90].

Although the fine details of the interplay between MBNL1 and CUG-BP1 remain to be understood, most of the data suggest that MBNL1 sequestration by long mismatched hairpins (Fig. 8B, a) is a key event leading to splicing deregulation and, eventually, disease. MBNL1 binds strongly to $(CUG)_n$ and $(CCUG)_n$ hairpins, but not to perfect $(CUG)_n \bullet (CAG)_n$ RNA duplexes, implicating a role for non-Watson–Crick U•U mismatches[91]. Most importantly, *Mbnl1*-knockout mice show splicing deregulation and phenotypic manifestations characteristic of myotonic dystrophy, such as skeletal muscle myotonia and cataracts[92]. Finally, the myotonic-dystrophy-like phenotype of mice expressing $(CUG)_n$ from a heterologous gene[76] could be reversed by overproduction of MBNL1 (ref. 93).

Protein sequestration has also been implicated in other diseases caused by repeat expansions in non-coding RNA. For example, MBNL1 is present in RNA foci that have been observed in cells isolated from patients with fragile X tremor and ataxia syndrome[94]. Furthermore, neurodegeneration caused by expression of a mutant human *SCA8* gene in *D. melanogaster* retina is modified by mutations in the *D. melanogaster muscleblind* gene[79]. This model, however, seems more plausible for the diseases that are caused by massive repeat expansions, where exceptionally long RNA hairpins could efficiently sequester these RNA-binding proteins, rather than for diseases with moderate-sized repeat expansions.

What could be an alternative mechanism for RNA gain of function? One exciting possibility is that repetitive RNA hairpins are reminiscent of unprocessed microRNAs. Processing of such hairpins by the RNA interference (RNAi) pathway could lead to silencing of genes that contain short complementary repeats in their transcripts, resulting in disease[95]. This hypothesis, the RNA degradation model (Fig. 8B, b), was initially supported by observations that $(CGG)_n$ hairpins are 'digested', albeit inefficiently, by the human protein Dicer, an RNase that is central to the

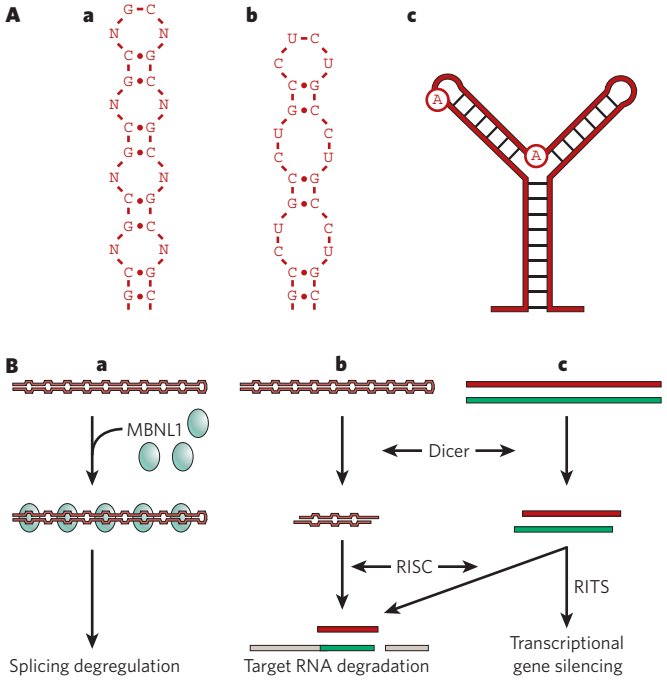

Figure 8 | Disease-associated RNA gain of function. A, Mismatched RNA hairpins can be formed by various expandable repeats: $(CNG)_n$ (**a**), $(CCUG)_n$ (**b**), and $(CGG)_n$ with two stabilizing AGG interruptions, with the A encircled (**c**). **B**, Molecular models of RNA gain of function. **a**, A repetitive RNA hairpin sequesters the protein MBNL1, which is required for normal splicing. **b**, A repetitive RNA hairpin is cleaved by Dicer and processed by the RNA-induced silencing complex (RISC), leading to the degradation of transcripts carrying short complementary repeats. **c**, Bidirectional transcription through repetitive runs generates double-stranded RNA. This leads to either target RNA degradation, if the processing is carried out by RISC, or chromatin silencing, if the processing is carried out by the RNA-induced transcriptional silencing complex (RITS). The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige.

RNAi response[96]. Recently, a variety of $(CNG)_n$-containing transcripts were shown to be targets of Dicer both *in vitro* and *in vivo*. Remarkably, the resultant short $(CNG)_n$ repeats functioned as siRNAs and triggered downstream silencing effects[97].

Another intriguing model involves chromatin silencing (Fig. 8B, c). This model is derived from recent data showing that in some diseases (such as myotonic dystrophy 1 and spinocerebellar ataxia 8), transcription across expandable repeats can proceed in both directions[98,99]. In the case of myotonic dystrophy 1, transcription of both the sense and antisense strands of the repetitive run led to the formation of a 21-nucleotide duplex RNA, pointing to the involvement of RNAi mechanisms. This antisense transcription was also linked to methylation of the lysine residue at position 9 of histone H3 and to recruitment of heterochromatin protein 1 (that is, to local formation of heterochromatin). Such RNA-induced initiation of transcriptional gene silencing could affect neighbouring gene expression, accounting for the diverse clinical manifestations of myotonic dystrophy 1 and perhaps of other diseases. More studies are needed to establish whether, and to what extent, different RNAi pathways are involved in repeat-expansion-associated diseases.

Future directions

Although the mechanisms that are responsible for repeat expansions in DNA are generally understood, many important questions remain unanswered. For example, will we find more expandable repeats and associated diseases in the future? If the structural concept presented in this review is correct, one would expect other structure-forming repeats to expand. But which factors are responsible for the initial expansions in human pedigrees? Mapping the replication origins in the vicinity of expandable repeats and studying in detail the mode of replication fork progression through the repeats in various cells and tissues should shed light on this matter. For example, recent mapping of the replication origin at the fragile X locus (*FRAXA*) provides additional support for the replication model of repeat expansions by showing that the structure-prone strand of the repetitive run functions as the lagging-strand template[100].

Another question is whether mechanisms of repeat expansion are conserved among various organisms or even among different cell types in the same organism. There are marked differences in expansion and contraction biases between mammals and unicellular organisms: repeats readily expand in humans but mostly contract in bacteria and yeast. In addition, repeats differ in their propensity to expand between various human cell types. We need to understand the reasons for these differences if we are to develop therapeutic approaches that could induce repeat contractions in humans. Could the trend for expansion of repeats in humans be reversed by pharmaceutical intervention? On the basis of our current knowledge, studies of drugs that affect DNA replication, repair or cell-cycle checkpoints are warranted.

Even more questions remain when it comes to the molecular pathways that lead from repeat expansions to disease. How widespread is the involvement of RNA toxicity in disease development? What is the role of repeat-containing RNA in the pathogenesis of polyglutamine-mediated disorders? What is the role of RNAi, broadly defined, in repeat-expansion-associated diseases? RNAi might have a considerable role, given that it provides an elegant explanation for the observed genetic dominance. And does chromatin silencing have a role in repeat expansion diseases? This emerging research direction seems to hold promise. Finally, is it possible to inhibit the aberrant processing of repeat-containing RNAs? If so, this would be of prime medical importance. Studies aimed at addressing these issues and related questions are underway in many laboratories worldwide. ■

1. Fleischer, B. Uber myotonische Dystrophie mit Katarakt. *Albrecht Von Graefes Arch. Klin. Exp. Ophthalmol.* **96,** 91–133 (1918).
2. Sherman, S. L. *et al.* Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum. Genet.* **69,** 289–299 (1985).
3. Verkerk, A. J. *et al.* Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65,** 905–914 (1991).
4. Kremer, E. J. *et al.* Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence $p(CCG)_n$. *Science* **252,** 1711–1714 (1991).
5. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352,** 77–79 (1991).
6. Brook, J. D. *et al.* 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 (1992).
7. Mahadevan, M. *et al.* Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. *Science* **255,** 1253–1255 (1992).
8. Mirkin, S. M. Molecular models for repeat expansions. *Chemtracts Biochem. Mol. Biol.* **17,** 639–662 (2004).
9. Pearson, C. E., Nichol Edamura, K. & Cleary, J. D. Repeat instability: mechanisms of dynamic mutations. *Nature Rev. Genet.* **6,** 729–742 (2005).
10. Kunst, C. B. & Warren, S. T. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* **77,** 853–861 (1994).
11. Jodice, C. *et al.* Effect of trinucleotide repeat length and parental sex on phenotypic variation in spinocerebellar ataxia I. *Am. J. Hum. Genet.* **54,** 959–965 (1994).
12. Brown, L. Y. & Brown, S. A. Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet.* **20,** 51–58 (2004).
13. Wells, R. D., Dere, R., Hebert, M. L., Napierala, M. & Son, L.S. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* **33,** 3785–3798 (2005).
14. Mirkin, S. M. DNA structures, repeat expansions and human hereditary disorders. *Curr. Opin. Struct. Biol.* **16,** 351–358 (2006).
15. Ranum, L. P. & Cooper, T. A. RNA-mediated neuromuscular disorders. *Annu. Rev. Neurosci.* **29,** 259–277 (2006).
16. Kunkel, T. A. Slippery DNA and diseases. *Nature* **365,** 207–208 (1993).
17. McMurray, C. T. DNA secondary structure: a common and causative factor for expansion in human disease. *Proc. Natl Acad. Sci. USA* **96,** 1823–1825 (1999).
18. Gacy, A. M., Goellner, G., Juranic, N., Macura, S. & McMurray, C. T. Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell* **81,** 533–540 (1995).
19. Dere, R., Napierala, M., Ranum, L. P. & Wells, R. D. Hairpin structure-forming propensity of the (CCTG•CAGG) tetranucleotide repeats contributes to the genetic instability associated with myotonic dystrophy type 2. *J. Biol. Chem.* **279,** 41715–41726 (2004).
20. Usdin, K. & Woodford, K. J. CGG repeats associated with DNA instability and chromosome fragility from structures that block DNA synthesis *in vitro*. *Nucleic Acids Res.* **23,** 4202–4209 (1995).
21. Fry, M. & Loeb, L. A. The fragile X syndrome $d(CGG)_n$ nucleotide repeats form a stable tetrahelical structure. *Proc. Natl Acad. Sci. USA* **91,** 4950–4954 (1994).
22. Pearson, C. E. & Sinden, R. R. Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry* **35,** 5041–5053 (1996).
23. Pearson, C. E. *et al.* Slipped-strand DNAs formed by long (CAG)•(CTG) repeats: slipped-out repeats and slip-out junctions. *Nucleic Acids Res.* **30,** 4534–4547 (2002).
24. Gacy, A. M. *et al.* GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol. Cell* **1,** 583–593 (1998).
25. Sakamoto, N. *et al.* Sticky DNA: self association properties of long (GAA)•(TTC) repeats in R•R•Y triplex structures from Friedreich's ataxia. *Mol. Cell* **3,** 465–475 (1999).
26. Vetcher, A. A. *et al.* Sticky DNA, a long (GAA•GAA•TTC) triplex that is formed intramolecularly, in the sequence of intron 1 of the frataxin gene. *J. Biol. Chem.* **277,** 39217–39227 (2002).
27. Potaman, V. N. *et al.* Unpaired structures in SCA10 $(ATTCT)_n$•$(AGAAT)_n$ repeats. *J. Mol. Biol.* **326,** 1095–1111 (2003).
28. Moore, H., Greenwell, P. W., Liu, C. P., Arnheim, N. & Petes, T. D. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc. Natl Acad. Sci. USA* **96,** 1504–1509 (1999).
29. Sakamoto, N. *et al.* GGA•TCC-interrupted triplets in long GAA•TTC repeats inhibit the formation of triplex and sticky DNA structures, alleviate transcription inhibition, and reduce genetic instabilities. *J. Biol. Chem.* **276,** 27178–27187 (2001).
30. Kang, S., Jaworski, A., Ohshima, K. & Wells, R. D. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nature Genet.* **10,** 213–218 (1995).
31. Ohshima, K. & Wells, R. D. Hairpin formation during DNA synthesis primer realignment *in vitro* in triplet repeat sequences from human hereditary disease genes. *J. Biol. Chem.* **272,** 16798–16806 (1997).
32. Freudenreich, C. H., Stavenhagen, J. B. & Zakian, V. A. Stability of a CTG•CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell. Biol.* **17,** 2090–2098 (1997).
33. Miret, J. J., Pessoa-Brandao, L. & Lahue, R. S. Orientation-dependent and sequence-specific expansions of CTG•CAG trinucleotide repeats in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **95,** 12438–12443 (1998).
34. Cleary, J. D., Nichol, K., Wang, Y. H. & Pearson, C. E. Evidence of *cis*-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nature Genet.* **31,** 37–46 (2002).
35. Rindler, M. P., Clark, R. M., Pollard, L. M., De Biase, I. & Bidichandani, S. I. Replication in mammalian cells recapitulates the locus-specific differences in somatic instability of genomic GAA triplet-repeats. *Nucleic Acids Res.* **34,** 6352–6361 (2006).
36. Bhattacharyya, S. & Lahue, R. S. *Saccharomyces cerevisiae* Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. *Mol. Cell. Biol.* **24,** 7324–7330 (2004).
37. Daee, D. L., Mertz, T., Collins, N. & Lahue, R. S. Post-replication repair inhibits CAG•CTG repeat expansions in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27,** 102–110 (2007).
38. Samadashwily, G. M., Raca, G. & Mirkin, S. M. Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.* **17,** 298–304 (1997).
39. Krasilnikova, M. M. & Mirkin, S. M. Replication stalling at Friedreich's ataxia $(GAA)_n$ repeats *in vivo*. *Mol. Cell. Biol.* **24,** 2286–2295 (2004).

40. Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R. S. & Mirkin, S. M. Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell. Biol.* **23,** 1349–1357 (2003).
41. Fouche, N., Ozgur, S., Roy, D. & Griffith, J. D. Replication fork regression in repetitive DNAs. *Nucleic Acids Res.* **34,** 6044–6050 (2006).
42. Manley, K., Shirley, T. L., Flaherty, L. & Messer, A. *Msh2* deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genet.* **23,** 471–473 (1999).
43. Savouret, C. *et al.* CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J.* **22,** 2264–2273 (2003).
44. van den Broek, W. J. *et al.* Somatic expansion behaviour of the $(CTG)_n$ repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum. Mol. Genet.* **11,** 191–198 (2002).
45. Kovtun, I. V. & McMurray, C. T. Trinucleotide expansion in haploid germ cells by gap repair. *Nature Genet.* **27,** 407–411 (2001).
46. Owen, B. A. *et al.* $(CAG)_n$-hairpin DNA binds to Msh2–Msh3 and changes properties of mismatch recognition. *Nature Struct. Mol. Biol.* **12,** 663–670 (2005).
47. Savouret, C. *et al.* MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Mol. Cell. Biol.* **24,** 629–637 (2004).
48. Yoon, S. R., Dubeau, L., de Young, M., Wexler, N. S. & Arnheim, N. Huntington disease expansion mutations in humans can occur before meiosis is completed. *Proc. Natl Acad. Sci. USA* **100,** 8834–8838 (2003).
49. Anvret, M. *et al.* Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum. Mol. Genet.* **2,** 1397–1400 (1993).
50. Kennedy, L. *et al.* Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.* **12,** 3359–3367 (2003).
51. Lia, A. S. *et al.* 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 (1998).
52. Fortune, M. T., Vassilopoulos, C., Coolbaugh, M. I., Siciliano, M. J. & Monckton, D. G. Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum. Mol. Genet.* **9,** 439–445 (2000).
53. Kovtun, I. V. *et al.* OGG1 initiates age-dependent CAG expansion in somatic cells during base excision repair of oxidized bases *in vitro* and *in vivo*. *Nature* **447,** 447–452 (2007).
54. Spiro, C. *et al.* Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol. Cell* **4,** 1079–1085 (1999).
55. Henricksen, L. A., Tom, S., Liu, Y. & Bambara, R. A. Inhibition of flap endonuclease 1 by flap secondary structure and relevance to repeat sequence expansion. *J. Biol. Chem.* **275,** 16420–16427 (2000).
56. Panigrahi, G. B., Lau, R., Montgomery, S. E., Leonard, M. R. & Pearson, C. E. Slipped (CTG)•(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. *Nature Struct. Mol. Biol.* **12,** 654–662 (2005).
57. Liu, Y., Zhang, H., Veeraraghavan, J., Bambara, R. A. & Freudenreich, C. H. *Saccharomyces cerevisiae* flap endonuclease 1 uses flap equilibration to maintain triplet repeat stability. *Mol. Cell. Biol.* **24,** 4049–4064 (2004).
58. van den Broek, W. J., Nelen, M. R., van der Heijden, G. W., Wansink, D. G. & Wieringa, B. Fen1 does not control somatic hypermutability of the $(CTG)_n$•$(CAG)_n$ repeat in a knock-in mouse model for DM1. *FEBS Lett.* **580,** 5208–5214 (2006).
59. Warren, S. T. Polyalanine expansion in synpolydactyly might result from unequal crossing-over of *HOXD13*. *Science* **275,** 408–409 (1997).
60. Richards, R. I. *et al.* Evidence of founder chromosomes in fragile X syndrome. *Nature Genet.* **1,** 257–260 (1992).
61. Jakupciak, J. P. & Wells, R. D. Genetic instabilities in (CTG)•(CAG) repeats occur by recombination. *J. Biol. Chem.* **274,** 23468–23479 (1999).
62. Napierala, M., Dere, R., Vetcher, A. & Wells, R. D. Structure-dependent recombination hot spot activity of GAA•TTC sequences from intron 1 of the Friedreich's ataxia gene. *J. Biol. Chem.* **279,** 6444–6454 (2004).
63. Dere, R. & Wells, R. D. DM2 CCTG•CAGG repeats are crossover hotspots that are more prone to expansions than the DM1 CTG•CAG repeats in *Escherichia coli*. *J. Mol. Biol.* **360,** 21–36 (2006).
64. Freudenreich, C. H., Kantrow, S. M. & Zakian, V. A. Expansion and length-dependent fragility of CTG repeats in yeast. *Science* **279,** 853–856 (1998).
65. Nag, D. K., Suri, M. & Stenson, E. K. Both CAG repeats and inverted DNA repeats stimulate spontaneous unequal sister-chromatid exchange in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **32,** 5677–5684 (2004).
66. Richard, G.-F., Goellner, G. M., McMurray, C. T. & Haber, J. E. Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11–RAD50–XRS2 complex. *EMBO J.* **19,** 2381–2390 (2000).
67. Richard, G.-F., Cyncynatus, C. & Dujon, B. Contractions and expansions of CAG•CTG trinucleotide repeats occur during ectopic gene conversion in yeast, by a *MUS81*-independent mechanism. *J. Mol. Biol.* **326,** 769–782 (2003).
68. Meservy, J. L. *et al.* Long CTG tracts from the myotonic dystrophy gene induce deletions and rearrangements during recombination at the *APRT* locus in CHO cells. *Mol. Cell. Biol.* **23,** 3152–3162 (2003).
69. Bacolla, A., Wojciechowska, M., Kosmider, B., Larson, J. E. & Wells, R. D. The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA Repair (Amst.)* **5,** 1161–1170 (2006).
70. Rolfsmeier, M. L., Dixon, M. J. & Lahue, R. S. Mismatch repair blocks expansions of interrupted trinucleotide repeats in yeast. *Mol. Cell* **6,** 1501–1507 (2000).
71. Mirkin, S. M. & Smirnova, E. V. Positioned to expand. *Nature Genet.* **31,** 5–6 (2002).
72. Cleary, J. D. & Pearson, C. E. Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. *Trends Genet.* **21,** 272–280 (2005).
73. Abu-Baker, A. & Rouleau, G. A. in *Genetic Instabilities and Neurological Diseases* (eds Wells, R. D. & Ashizawa, T.) 487–513 (Elsevier, Amsterdam, 2006).
74. Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H. & Housman, D. E. Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl Acad. Sci. USA* **94,** 7388–7393 (1997).
75. Amack, J. D. & Mahadevan, M. S. The myotonic dystrophy expanded CUG repeat tract is necessary but not sufficient to disrupt C2C12 myoblast differentiation. *Hum. Mol. Genet.* **10,** 1879–1887 (2001).
76. Mankodi, A. *et al.* Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* **289,** 1769–1772 (2000).
77. Tassone, F. *et al.* Elevated levels of *FMR1* mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.* **66,** 6–15 (2000).
78. Jin, P. *et al.* RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in *Drosophila*. *Neuron* **39,** 739–747 (2003).
79. Mutsuddi, M., Marshall, C. M., Benzow, K. A., Koob, M. D. & Rebay, I. The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with staufen in *Drosophila*. *Curr. Biol.* **14,** 302–308 (2004).
80. Lin, X. & Ashizawa, T. Recent progress in spinocerebellar ataxia type-10 (SCA10). *Cerebellum* **4,** 37–42 (2005).
81. Napierala, M. & Krzyzosiak, W. J. CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J. Biol. Chem.* **272,** 31079–31085 (1997).
82. Sobczak, K., de Mezer, M., Michlewski, G., Krol, J. & Krzyzosiak, W. J. RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.* **31,** 5469–5482 (2003).
83. Handa, V., Yeh, H. J., McPhie, P. & Usdin, K. The AUUCU repeats responsible for spinocerebellar ataxia type 10 form unusual RNA hairpins. *J. Biol. Chem.* **280,** 29340–29345 (2005).
84. Napierala, M., Michalowski, D., de Mezer, M. & Krzyzosiak, W. J. Facile *FMR1* mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res.* **33,** 451–463 (2005).
85. Fardaei, M. *et al.* Three proteins, MBNL, MBLL and MBXL, co-localize *in vivo* with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* **11,** 805–814 (2002).
86. Miller, J. W. *et al.* Recruitment of human muscleblind proteins to $(CUG)_n$ expansions associated with myotonic dystrophy. *EMBO J.* **19,** 4439–4448 (2000).
87. Lu, X., Timchenko, N. A. & Timchenko, L. T. Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy. *Hum. Mol. Genet.* **8,** 53–60 (1999).
88. Philips, A. V., Timchenko, L. T. & Cooper, T. A. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280,** 737–741 (1998).
89. Thornton, C. A., Swanson, M. S. & Cooper, T. A. in *Genetic Instabilities and Neurological Diseases* (eds Wells, R. D. & Ashizawa, T.) 37–54 (Elsevier, Amsterdam, 2006).
90. Ho, T. H. *et al.* Muscleblind proteins regulate alternative splicing. *EMBO J.* **23,** 3103–3112 (2004).
91. Kino, Y. *et al.* Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.* **13,** 495–507 (2004).
92. Kanadia, R. N. *et al.* A muscleblind knockout model for myotonic dystrophy. *Science* **302,** 1978–1980 (2003).
93. Kanadia, R. N. *et al.* Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. *Proc. Natl Acad. Sci. USA* **103,** 11748–11753 (2006).
94. Iwahashi, C. K. *et al.* Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129,** 256–271 (2006).
95. Malinina, L. Possible involvement of the RNAi pathway in trinucleotide repeat expansion diseases. *J. Biomol. Struct. Dyn.* **23,** 233–235 (2005).
96. Handa, V., Saha, T. & Usdin, K. The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. *Nucleic Acids Res.* **31,** 6243–6248 (2003).
97. Krol, J. *et al.* Ribonuclease Dicer cleaves triplet repeat hairpins into shorter repeats which silence specific targets. *Mol. Cell* **25,** 575–586 (2007).
98. Cho, D. H. *et al.* Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell* **20,** 483–489 (2005).
99. Moseley, M. L. *et al.* Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nature Genet.* **38,** 758–769 (2006).
100. Gray, S. J., Gerhardt, J., Doerfler, W., Small, L. E. & Fanning, E. An origin of DNA replication in the promoter region of the human fragile X mental retardation (*FMR1*) gene. *Mol. Cell. Biol.* **27,** 426–437 (2007).

Acknowledgements I thank W. Krzyzosiak, R. Lahue, K. Lobachev, C. McMurray, D. Monckton, C. Pearson, M. Swanson, K. Usdin and R. Wells for sharing their ideas and unpublished results. I also extend my gratitude to all the participants of the 5th International Conference on Unstable Microsatellites and Human Disease (Granada, Spain, 2006) for their intense and productive discussions, which helped to shape this review. I am indebted to my wife, Kate, for her invaluable critical comments. I thank J. White and P. White for their generous support. This work was supported by the National Institutes of Health.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (sergei.mirkin@tufts.edu).