REVIEWS

REPEAT INSTABILITY: MECHANISMS OF DYNAMIC MUTATIONS

Christopher E. Pearson‡, Kerrie Nichol Edamura* and John D. Cleary*‡*

Abstract | Disease-causing repeat instability is an important and unique form of mutation that is linked to more than 40 neurological, neurodegenerative and neuromuscular disorders. DNA repeat expansion mutations are dynamic and ongoing within tissues and across generations. The patterns of inherited and tissue-specific instability are determined by both gene-specific *cis*-elements and *trans*-acting DNA metabolic proteins. Repeat instability probably involves the formation of unusual DNA structures during DNA replication, repair and recombination. Experimental advances towards explaining the mechanisms of repeat instability have broadened our understanding of this mutational process. They have revealed surprising ways in which metabolic pathways can drive or protect from repeat instability.

GENETIC ANTICIPATION
A phenomenon in which disease severity increases and/or age of onset of disease decreases from one generation to the next.

**Program of Genetics and Genomic Biology, The Hospital for Sick Children, 15-312, TMDT, 101 College Street, East Tower, Toronto, Ontario M5G 1L7, Canada.*
‡Department of Molecular and Medical Genetics, University of Toronto, Canada.
Correspondence to C.E.P.
e-mail: cepearson@genet.sickkids.on.ca
doi:10.1038/nrg1689

Repeat instability is an important and unique form of mutation that is linked to more than 40 neurological, neurodegenerative and neuromuscular disorders (see supplementary information S1 (table)). Unlike static mutations, which are retained in somatic tissues and stably transmitted to offspring, the repeat mutation process is dynamic, with products that continue to mutate within tissues and across generations. Longer tracts are more likely to undergo an expansion mutation than shorter tracts. As repeat tract length correlates with disease severity and age of onset, this phenomenon leads to GENETIC ANTICIPATION, a hallmark of most repeat disorders. Trinucleotide repeats (TNRs) form the largest component of a broader category of repeat-associated disorders that also includes tetranucleotides (dystrophia myotonica 2, DM2), pentanucleotides (spinocerebellar ataxia 10, SCA10), minisatellites (epilepsy, progressive myoclonic 1, EPM1; insulin, INS) and megasatellites (facioscapulohumeral muscular dystrophy 1A, FSHMD1A) (FIG. 1a; TABLE 1). Repeat instability shows complex patterns between and within tissues that vary with developmental, epigenetic, proliferative and possibly environmental cues. In this article we focus on the mechanisms of repeat instability during DNA replication, repair and recombination, covering those studies that most closely relate to and/or explain events that occur in humans.

TNR instability: when, where and how?

Repeat instability probably arises through multiple processes that occur individually or in combination, depending on the tissue, proliferative status and developmental stage of the cell (FIG. 1b). Assigning any one DNA metabolic process as the cause of TNR instability is difficult owing to their overlapping relationships; nevertheless, individual contributions can be delimited. For example, tract-length heterogeneity (mosaicism) within a tissue indicates that instability has occurred but does not indicate when. However, age-dependent instability, which actively accumulates in post-mitotic tissues (neurons), is a clear indication of instability that arises independently of DNA replication (genome duplication) and so must be the result of genome-maintenance repair. Therefore, the assignment of specific metabolic processes to repeat instability must consider the cell's metabolic history. A proper deconstruction of repeat instability (minimizing complexity to maximize conclusions) by an experimental system must take into account the complexity of the human data. Determining the 'when' and 'where' will lead to a better understanding of 'how' repeat instability occurs.

Germline instability. All TNR diseases involve mutations during parent-to-offspring transmission, implicating germline mutations in TNR instability (FIG. 2).

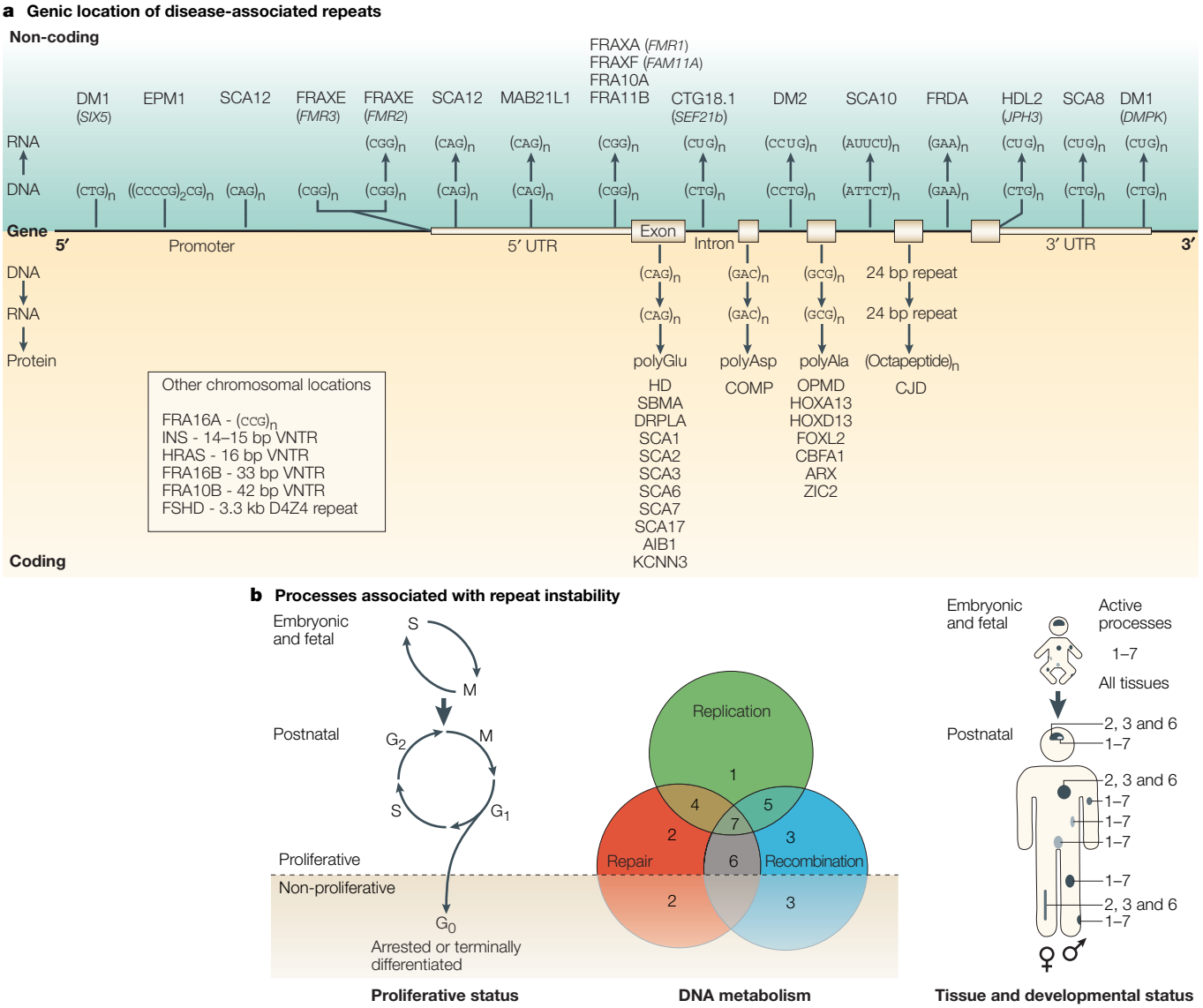


Figure 1 | **Unstable repeat tracts and the processes associated with repeat instability. a** | A schematic representation of the genic location of non-coding (top) and coding (bottom) disease-associated repeats (promoter, 5′ UTR, exons, introns, 3′ UTR or other chromosomal locations (includes undetermined/unclassified locations)). The DNA, RNA and amino-acid sequence for each repeat is noted. For dystrophia myotonica 1 (DM1), the CTG tract is located in the 3′ UTR of dystrophia myotonica protein kinase (DMPK), as well as the promoter of sine oculis homeobox homologue 5 (SIX5). **b** | The processes associated with repeat instability. Instability occurs in proliferative (above the dotted line) and non-proliferative (below the dotted line) tissues. The DNA metabolic processes of DNA replication, repair and recombination are associated with repeat instability, either independently (1–3) or in conjunction with other processes (4–7). The involvement of various metabolic processes varies depending on tissue and developmental stage (bottom right). The numbers in the tissue and developmental status section correspond generally to those in the DNA metabolism section. AIB1, amplified in breast cancer 1; ARX, aristaless-related homeobox; CBFA1, core binding factor α1; CJD, Creutzfeld–Jakob disease; COMP, cartilage oligomeric matrix protein; DM2, dystrophia myotonica 2; DRPLA, dentatorubral-pallidoluysian atrophy; EPM1, epilepsy, progressive myoclonic 1; FOXL2; forkhead box L2; FRAXA/FRAXE/FRAXF/FRA10A/FRA10B/FRA11B/FRA16A/FRA16B, rare fragile sites; FRDA, Friedreich ataxia; FSHD, facioscapulohumeral muscular dystrophy; HD, Huntington disease; HDL2, Huntington disease-like 2; HOXA13/HOXD13, homeobox A13 and D13; HRAS, v-Ha-ras Harvey rat sarcoma viral oncogene homologue; INS, insulin; JPH3, junctophilin 3; KCNN3, small conductance calcium-activated potassium channel protein 3; MAB21L1, mab-21-like 1; OPMD, oculopharyngeal muscular dystrophy; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia proteins; VNTR, variable number tandem repeat; ZIC2, zic family member 2.

Paternal and maternal expansion biases are evident and might be driven by processes that are specific to sperm or oocyte development. Highly specialized DNA metabolic activities that involve stage-specific expression of replication, repair or recombination factors, replication programmes, expression profiles and epigenetic modifications might contribute to germline instability during gametogenesis (reviewed in REF. 1). Germline mutations can occur in proliferating, arrested, meiotic or dormant haploid germ cells, implicating not

Table 1 | **Nucleotide repeat lengths within the human genome**

Repeat type	Unit size	Array size (in a normal population)
Megasatellite	Several kb	Up to several hundred kb
Satellite	5–171 bp	100 kb to several Mb
Minisatellite	6–64 bp	0.1–20 kb range
Microsatellite (includes trinucleotide repeats)	1–4 bp	Hundreds of bp

The data are from REF. 162 and the Tandemly Repeated (Satellite) DNA web site.

only meiotic recombination but also DNA replication and repair in germline repeat instability[1].

Paternal expansion bias. The paternal expansion bias that is characteristic of most CAG (polyGlu) disorders (FIG. 2a; see also supplementary information S1 (table)) might result from pre-meiotic mutations during mitotic cycles of spermatogenesis. In a study of patients with Huntington disease (HD), CAG expansions had occurred before meiosis, as they were already present in mitotic diploid germ cells[2] and might have arisen at any time from primordial germ cell segregation *in utero* through to the life-long, post-pubertal, SPERMATOGONIAL stem cell divisions[1]. Furthermore, the increase in expansions with increasing paternal age for numerous transgenic CAG/CTG mice supports continuing instability in the post-pubertal male germ line, probably through errors of replication and/or repair during the pre-meiotic proliferative stages of spermatogonial divisions[3–6]. Although paternal expansion bias is often attributed to the greater number of mitotic divisions for male gametogenesis, other spermatogenic-specific processes, such as stage-specific alterations in the levels and activity of various repair proteins, might also contribute.

Maternal expansion bias. In contrast to male germ cells, oogenic meiosis begins *in utero*, arrests for years (through birth and puberty), resumes only minutes before ovulation and is not completed until after fertilization. Also in contrast to post-meiotic male germ cells that are virtually void of proteins, arrested oocytes are primed with DNA repair and recombination activities, which can lead to age-related TNR instability, as observed in spinocerebellar ataxia 1 (SCA1) transgenic mice[7]. Therefore, the maternal expansion bias that is observed in fragile X syndrome (FRAXA; also known as fragile site mental retardation 1, FMR1) (FIG. 2b) and dystrophia myotonica 1 (DM1) might be linked to the highly extended time for oogenic meiosis.

CGG expansions are present in the germ cells of female fetuses with FULL MUTATIONS from FRAXA PREMUTATION mothers. These mutations probably arose by meiotic (recombination) events, rather than post-zygotic cellular divisions before fetal primordial germ cell segregation[8], as length mosaicism is rare in many premutation-to-full mutation transmissions (which are expected to arise through cell divisions)[9]. So, the FRAXA expansion mutation might occur in the grand maternal uterus during oogenesis within the developing maternal fetus, and is only revealed on transmission of the oocyte to her progeny (FIG. 2b).

Both pre-meiotic (in the grand maternal uterus) and post-zygotic events probably contribute to the large DM1 CTG expansions (>1,000 repeats) that arise by maternal transmission. Expansions that are present in prophase I immature oocytes might have arisen at any time after maternal zygote formation, which implicates either genome-duplication errors, genome-maintenance errors during quiescence or recombination during prophase I in the expansion process[10,11]. Post-zygotic expansions are not evident in 3-day embryo blastocysts (100–120 cells) that are derived from DM1 oocytes or DM1 sperm, but high levels of inter-tissue CTG instability arise in the second trimester[10,12]. So, the DM1 expansion mutation might occur in the grand maternal uterus during oogenesis (before completion of meiosis I) within the developing maternal fetus and is revealed on transmission of the oocyte to the offspring. In addition to germline instability there is also somatic instability that begins after the thirteenth week of fetal development of that offspring and during their adult life. Together these two processes contribute to the large maternal expansions.

Paternal contraction bias. A strong paternal contraction bias is evident for expanded repeats in patients with spinocerebellar ataxia 8 (SCA8), Friedreich ataxia (FRDA) and FRAXA, and, in the third case, might be related to CpG methylation. In fact, all males with full-mutation FRAXA have only CGG premutations in their testis and sperm[13]. These males inherited a single maternally expanded allele that experiences germ-cell-specific CGG contractions between 13 and 17 weeks of fetal development[8]. Premutation repeats in testes and sperm are unmethylated, which is in contrast with methylated full expansions in all other tissues[13], suggesting that the contraction process is coupled with an escape from aberrant CpG methylation and selective growth of the FMRP-expressing male germ cells[8]. The recently demonstrated protection against replication-mediated CGG contractions by methylation is consistent with the active contractions that arise in proliferating paternal primordial germ cells[14]. Male germline-specific, methylation-associated CGG contractions might also apply to the FRAXE, FRAXF, FRA10B, FRA11A and FRA16B loci, each of which is associated with aberrantly methylated CGG expansions[15]. Other TNRs that show paternal contractions[16–18], such as CAG, CTG or GAA repeats, cannot be methylated, but these repeats are embedded within CpG ISLANDS. Whether CpG methylation (or its absence) in the flanking regions of the SCA8, DM1 or FRDA repeats contributes to contractions in the male germ line is not known. It has been suggested that disease-causing contractions of the CpG-rich D4Z4 megasatellite is linked with aberrant CpG methylation in the germ line[19]. Further germline-specific epigenetic modifications might define both mutational bias and PARENT-OF-ORIGIN EFFECTS.

SPERMATOGONIA
The mitotically dividing stem cells of the male germ line, the descendants of which ultimately become mature sperm.

FULL MUTATION
The expanded repeat tract that is typically associated with disease. The term is often used to distinguish this event from an individual who has the shorter premutation or proto-mutation expansions that are not associated with disease.

PREMUTATION
A repeat tract of a length that rarely leads to disease symptoms. However, the possibility for further repeat-length expansion to occur on transmission is high as a result of the longer repeat length. The term applies to disorders such as Huntington disease and fragile X syndrome. The term proto-mutation applies to dystrophia myotonica 1, for which individuals might (or might not) eventually develop symptoms.

CpG ISLAND
A sequence of at least 200 bp with a greater number of CpG sites than expected for its GC content. These regions are often GC rich, associated with genes and typically undermethylated.

PARENT-OF-ORIGIN EFFECT
The increased proportion of paternal or maternal disease-causing transmission to offspring. This effect is molecularly explained by a paternal or maternal repeat-expansion bias in the germ line.

Tissue-specific somatic repeat instability

Tissue-specific somatic repeat instability has been observed in many but not all repeat diseases and is evident as tract length variations within or between tissues (FIGS 1b,2c; see also supplementary information S1 (table)). The timing, pattern and tissue selectivity of somatic instability vary between repeat disorders. For example, instability can be detected in adult but not fetal HD and spinal and bulbar muscular atrophy (SBMA) tissues[20,21], in contrast to FRAXA instability, which is evident only in early fetal tissues and not postnatally[13,22,23]. Contractions of the expanded FRDA GAA tract accumulate in lymphocytes throughout a patient's life[24]. In DM1 inter-tissue instability becomes apparent in fetuses at 13–16 weeks[12,25] during the rapid fetal growth of the second trimester[12], and for

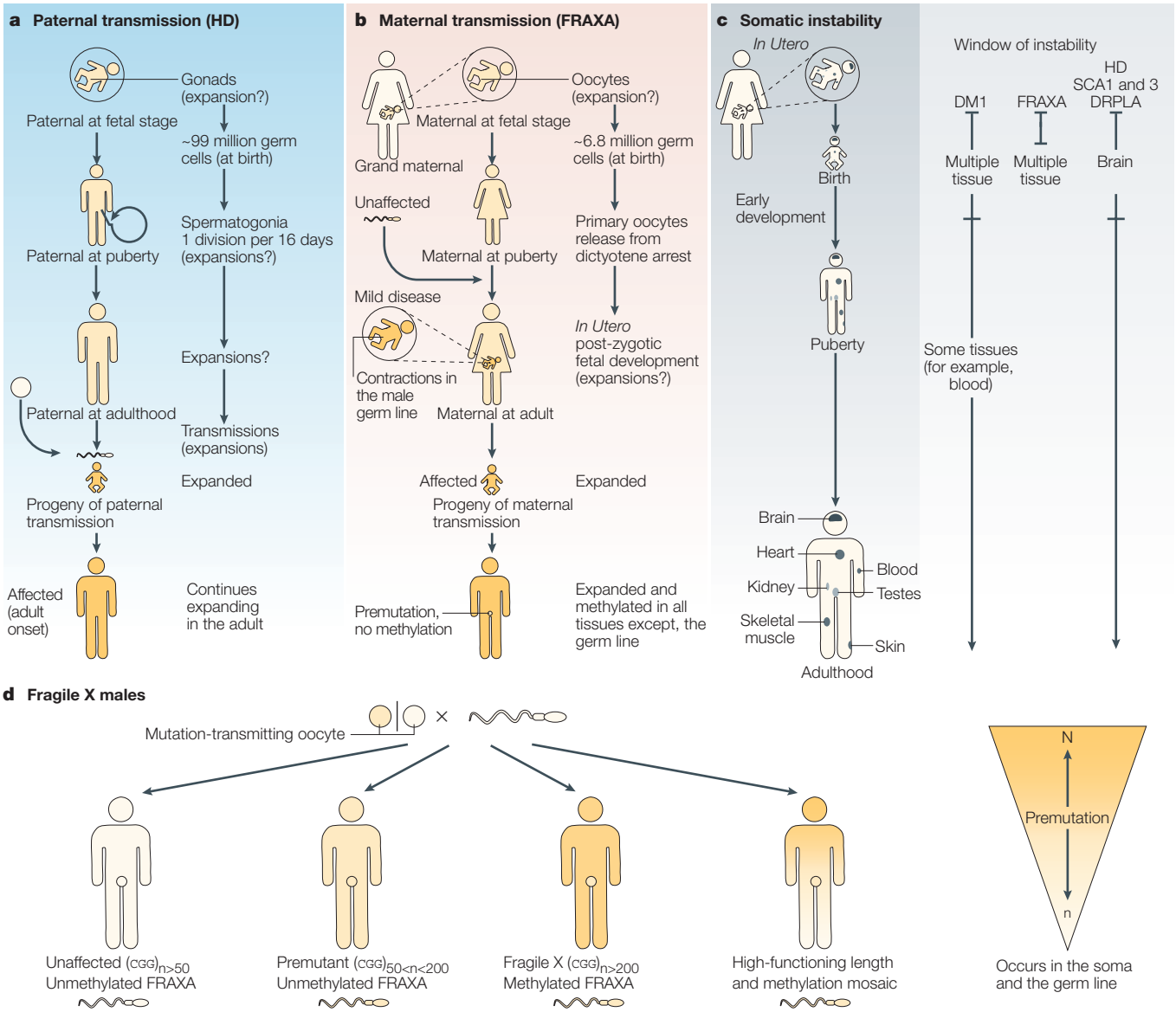


Figure 2 | **Germline and somatic instability. a** | Paternal transmission. Transmission of a Huntington disease (HD) premutation (intermediate allele) to the male fetus is shown. Expansions in the sperm could be transmitted, resulting in the onset (in adulthood) of HD in offspring of a premutation father. **b** | Maternal transmission. Transmission of a fragile X syndrome (FRAXA) premutation to full mutation might occur during either oogenesis or early post-zygotic events. FRAXA males have fully methylated expanded CGG in all tissues except sperm, where only unmethylated premutation lengths exist. **c** | Somatic instability. Variations in repeat lengths between tissues can occur during multiple stages of development. For most trinucleotide repeat disorders, a high degree of somatic instability occurs early in development, which for dystrophia myotonica 1 (DM1) continues to occur during postnatal growth in blood. Instability in the brain might also continue to occur throughout the lifetime of the individual for HD, spinocerebellar ataxia 1 and 3 (SCA1/SCA3) and dentatorubral-pallidoluysian atrophy (DRPLA). **d** | Transmission of a FRAXA CGG premutation through the female germ line (yellow circle) can result in males having unmethylated premutation lengths, methylated full-mutation FRAXA males or 'high-functioning' males who are mosaic for methylation and CGG-repeat length. Despite the differences in repeat length and methylation status within the somatic tissue of these individuals, only unmethylated premutation lengths are present in their testes (pale yellow circles) and sperm.

some tissues expansions continue through a patient's life. Furthermore, the very large CTG expansions in DM1 muscle seem to occur during the proliferative cell divisions of muscle, and might cease following terminal differentiation and cell quiescence[26,27]. The expansions in highly proliferative lymphocytes over the lifetime of patients who have DM1[28] and the larger expansions in tumours of those patients and DM1 mice provide a further link between expansion and cell proliferation, which indicates a role for genome-duplication processes[29–31]. As there seems to be no simple association between cell division and instability[32,33], tissue-specific *cis*-elements and/or *trans*-factors probably determine the stability of a given locus. In fact, tissue-specific repeat instability might be linked to the activity of repair systems, replication programme, epigenetic marks, chromatin packaging or transcription levels of the disease gene — all of which can differ between genetic loci, vary between tissues, and change throughout development and ageing[34–36].

Instability in the CNS. Extensive analysis of repeat instability in the CNS for various disease loci has revealed crucial insights into the roles of *cis*-elements, *trans*-factors and genome-maintenance repair. A contribution of DNA replication to CNS repeat instability is possible, but unlikely. Non-replication processes must be involved because repeat length varies in brains of numerous CAG transgenic mice, and repeat instability was seen in one GAA mouse after the postnatal 'brain growth spurt' (in mice replication peaks 6–17 days after birth; whereas in humans the brain growth spurt ends at birth)[31,32,37–42]. In fact, many but not all CAG diseases show age-dependent instability in the brains of patients[43–45], further supporting a role for genome-maintenance repair in repeat instability over the course of a patient's life.

Most of the polyGlu diseases in humans and in mouse models show some overlap in their pattern of instability across brain subregions (for example, the longest tracts are in the striatum or cerebrum, the shortest are in the cerebellum), which possibly reflects the developmental history of each subdivision. Although inter-region length variations are common, there is only limited CAG-length heterogeneity within a single brain region (see also BOX 1). Analysis of individual cell types in the brains of patients who have dentatorubral-pallidoluysian atrophy (DRPLA) revealed different degrees of length heterogeneity between neuronal subtypes, with tracts that are shorter and longer than those in the blood[46,47]. Cerebellar granular cells had shorter CAG tracts and the lengths were less variable than in Purkinje or cerebral neuronal cells. Glial cells showed more CAG instability than neurons, which indicates that proliferation, at least in glia, might contribute to instability. Cell-death-induced neurogenesis and proliferation could enhance CAG expansions[48]. The brain region-specific and cell type-specific CAG instability observed in humans and polyGlu mouse models might have arisen through a shared developmental history for each subdivision; because the tract length of the precursor cell(s) that seeded a given neural or glial region determined the ultimate stability pattern in that region[49]; and/or because tissue-specific or cell-specific *trans*-factor(s), such as DNA repair, determine instability patterns[37,50,51].

Expansion of CAG tracts in transgenic mouse brains seems to require the mismatch repair (MMR) proteins MSH2, MSH3 and PMS2 (REFS 5,6,31,41,52). Recent *in vitro* evidence revealed that human neuron-like cell extracts can process slipped CAG/CTG DNA mutagenic intermediates in an error-prone manner that can lead to expansions[53]. Deciphering which factors are responsible for inter-tissue length heterogeneity is a challenge given the complex cellular make-up of neural and muscle tissues and because the capacity for repair changes on differentiation, development and ageing[34,35].

Box 1 | **Role of instability in disease pathogenesis?**

In contrast to dystrophia myotonica 1 — in which differences that are as great as 1,000 repeats can occur between the affected tissue (muscle) and blood — the degree to which somatic instability contributes to the neurodegenerative polyGlu diseases is not obvious. In attempting to correlate repeat instability in the CNS with vulnerable brain regions[37,43,51,146], a low frequency of minor repeat length changes (±1–23 repeats) has been reported, with only rare examples of individual cells having incurred large expansions (≥1,000 repeats)[44]. With the exception of Huntington disease (HD)[147] and Machado–Joseph disease[148], all studies revealed a poor correlation of CAG instability with cell degeneration, which indicates that somatic instability does not contribute to pathogenesis. Alternatively, cells that incurred the largest expansions might have been those that degenerated and are no longer available for CAG-length analysis[44].

Is the correlation of somatic instability with susceptibility to degeneration the best measure of a potential contribution to pathogenesis? Neural degeneration (cell loss) is probably the principal cause of disease, but is probably not the only cause. For example, both the gain of function and partial loss of function of the expanded polyGlu proteins have been proposed as sources of disease[149,150]. Many of the polyGlu genes are ubiquitously expressed within and outside the CNS, including cells that do not degenerate. It is possible that some of the clinical symptoms arise from aberrant function(s) of the mutant polyGlu protein in tissues that do not degenerate. In fact, the ubiquitin–proteasome pathway is perturbed in both skin and the brain of patients with HD both before and after the onset of neurodegeneration[151]. Moreover, non-neural HD cells have a progressive pathology, including dysmorphic cells with aberrant nuclear morphology, multiple micronuclei, large vacuoles, numerous centrosomes, reduced mitotic index and increased aneuploidy[152]. Considering the prolonged disease course of many of the illnesses, it is likely that the neurons experience long periods of neuronal dysfunction before cell death. PolyGlu expansions can selectively induce transcriptome alterations that can lead to neural dysfunction before degeneration[153]. In fact, some patients with HD and transgenic mice with expanded polyGlus show neurological symptoms in the absence of neurodegeneration[154–157]. Biochemical functions of polyGlu proteins can be modulated by minor increases in the length of the polyGlu tract, which are likely to have a graded effect on natural and aberrant protein functions[158,159]. Proteins that have variant repeat numbers, even within the 'normal' ranges, might lead to clinical symptoms (reviewed in REFS 160,161). However, although minor CAG/polyGlu length alterations might have clinical effects, it is unlikely that the ablation of somatic instability will eliminate disease. In mouse models for HD somatic CAG expansions in the brain are MSH2-dependent, with no instability in its absence — an abrogation that delays neural pathology (nuclear mutant protein accumulation) by ~5 months[41]. It should be considered that these mice contain 109 repeats at birth, whereas most human patients with HD inherit 40–50 repeats and the relative contribution of somatic instability in disease progression might be very different. Better appreciation of the normal functions of the non-expanded protein products for each disease will enhance our understanding of the pathogenesis and of the potential contribution of somatic instability to pathogenesis.

Not all CAG/CTG disease loci share tissue-specific patterns of instability. For example, distinct brain region instability patterns occur between DRPLA, MJD (Machado–Joseph disease)/SCA3 and SCA1 (REF. 54). Although the SBMA tract is stable in the CNS it is unstable in muscle[55]; this is in contrast to neuronal instability of some TNR disorders, such as DRPLA. Similarly, the high degree of instability in muscle of patients who have DM1 and SBMA contrasts with the muscle that shows the lowest instability in HD and DRPLA[45,55,56]. In contrast to CAG/CTG diseases (HD, SCA1, SCA2, SCA3, DRPLA and DM1), in which repeat tracts are shorter in the cerebellum, the FRDA GAA is recalcitrant to contractions in the cerebellum relative to other brain regions[57], and might even be subject to expansions, as has been shown in transgenic mice[40]. There is no length mosaicism in the CNS of FRAXA patients — apart from 'high functioning' FRAXA methylation mosaics — who show dramatic levels of CGG brain-region heterogeneity[13,23,58]. Together, these variations in patterns of repeat instabilities indicate that in addition to tissue-specific or cell-specific *trans*-factors, gene-specific *cis*-elements also modulate the degree and pattern of repeat instability in the CNS.

Cis-elements and TNR instability

Cis-elements that might affect instability can be both internal (repeat sequence, tract length and purity) and external (flanking sequence elements, nucleosomes, CpG methylation and replication origins) to the repeat tract (reviewed in REF 36). That only certain repeat sequences become unstable might relate to sequence-specific DNA structures that are thought to be crucial for instability. TNR sequences can form slipped-stranded, quadruplex, triplex, sticky DNAs and intra-strand hairpins. Slipped-stranded structures are widely proposed to be mutagenic intermediates of repeat instability. Although they are likely to occur during mutation events, there is so far no evidence that these exist *in vivo*.

Instability is intrinsically connected to repeat length, as only tracts above a stability threshold (~34 repeats) become unstable. Interruptions can significantly alter this process; for example, SCA1 CAG tracts that are as large as 39 repeats with a single CAT interruption are somatically stable, whereas alleles that have 40 'pure' repeats are unstable[59]. The influence of repeat sequence, tract length and purity on the propensity of slipped-structure formation correlates with their effect on genetic instability, further implicating slipped DNAs as mutagenic intermediates of instability[60] during DNA metabolism.

In many repeat disorders, the inherited expansion mutations are in LINKAGE DISEQUILIBRIUM with closely linked DNA markers, which indicates that these chromosomal backgrounds might contain external (to the repeat) intralocus *cis*-elements that predispose, or drive, TNR expansions. The existence of interlocus-specific external *cis*-elements is supported by the fact that identical repeat sequences can show different levels of instability at different loci[36,61,62]. Further evidence for the contribution of *cis*-elements comes from mouse models in which mice generally require larger repeat tracts (>100 repeats) than humans to show instability[3,63]. However, the inclusion of large amounts of flanking human genomic sequence allows instability for moderate repeat lengths in some mouse models (45–78 repeats)[38,64–66], which indicates that elements within these flanking sequences drive instability. Furthermore, mice that carry a single integration of a large CTG repeat with minimal human flanking sequence had remarkably different patterns of instability[4] depending on the insertion site. Mice with larger amounts of patient-derived flanking sequence (45 kb) show high levels of instability for all insertion sites[64,67]. Selective exclusion of flank sequences has highlighted their importance for instability[66]. Together these observations demonstrate that *cis*-elements in humans and mouse models can influence instability.

CpG methylation, a *cis*-element, might stabilize the expanded CGG tracts of FRAXA. The period of somatic CGG instability in FRAXA is restricted to early embryonic and fetal development, and ends at a time that coincides with the aberrant methylation of the expanded CGG alleles[13,22,23,68,69]. The FRAXA male gonads that escape aberrant methylation continue to contract the CGG tract. Similarly, methylation status is tightly correlated with active somatic contractions in FRAXA methylation mosaics[23,70–72] (FIG. 2d). The effects of methylation on repeat instability might extend to other loci that also show aberrant CpG methylation of expanded CGG tracts[15].

A mechanistic link between CpG methylation and protection from replication-mediated CGG-repeat contractions has been demonstrated in primate cells[14]. Expression of a CpG methylase protected various human repeat sequences from contractions during replication in bacteria[73], including many repeats that are void of methylatable CpG sites, indicating the involvement of CpG sites that flank the repeats. CpG methylation regulation, through exposure to a methylation inhibitor, might directly or indirectly affect CTG instability in patients with DM1 (expansions) and rodent cells (contractions) that harbour expanded CTG tracts[74]; the significance of the direction of the effect is unclear.

DNA methylation alters DNA structure, protein binding, cellular activity and differentiation, implicating a complex mechanism through which CpG methylation might alter repeat instability. Locus-specific epigenetic CpG methylation status might be a signal that defines the timing, tissue and pattern of repeat instability of a particular genomic site.

Other external *cis*-elements include nucleosome and replication origin positioning. Expanded CTG and CGG repeats favour and disfavour nucleosome assembly, respectively[75,76], and the similarity between the stability threshold (35 repeats) and the length of DNA in a nucleosome (146 bp) also favours the involvement of nucleosomes in repeat instability. The loss of a DNaseI hypersensitive site near the expanded CTG tract that is associated with DM1 also supports the contribution of altered chromatin to instability[77]. Although the potential *cis*-elements of chromatin structure[77,78], GC content,

LINKAGE DISEQUILIBRIUM
A measure of genetic associations between alleles at different loci, which indicates whether allelic or marker associations on the same chromosome are more common than expected.

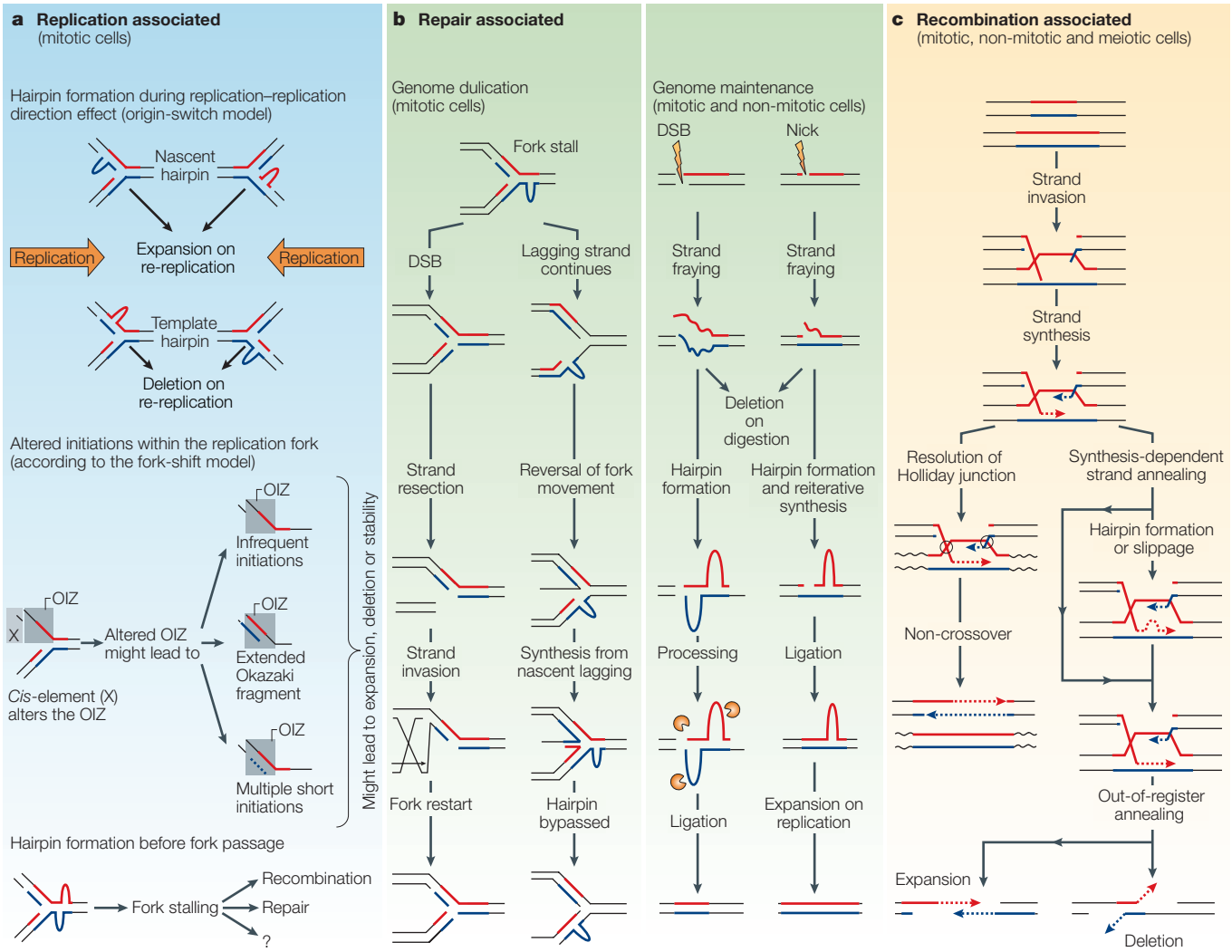


Figure 3 | **DNA metabolic processes associated with repeat instability. a**| Replication-associated instability. Replication across the hairpin might result in expansions or deletions for nascent or template hairpins, respectively. Alterations in the Okazaki initiation zone (OIZ) relative to repeat and Okazaki initiations might influence the formation of hairpins. An advancing replication fork that is stalled by hairpin formation will require another process for restart. **b** | Repair-associated instability can be associated with genome duplication and maintenance. Following replication forks stalling, the induction of a double-strand break (DSB) (first column) or fork reversal (second column) might result in length alterations being maintained during proceeding rounds of replication. During genome maintenance, the presence of a DSB or a nick within the repeat tract might lead to strand fraying and trinucleotide-repeat-specific structures. Failure of repair to correct the alteration will result in length differences. **c** | Recombination-mediated instability. Homologous recombination between allelic repeats might occur with or without the exchange of flanks. Gene conversion might occur after DSB or replication-fork blockage. Single strands from the broken repeat invade the sister tract, allowing for completion or restart of synthesis. During this period, hairpins or reiterative synthesis might occur, prompting repeat-length changes. Circles represent Holliday junctions.

CpG density[61,62], transcription levels[55] and replication alterations[14,79] have been suggested to affect TNR instability, the exact nature and interaction between *cis*-elements and TNR instability remains unclear.

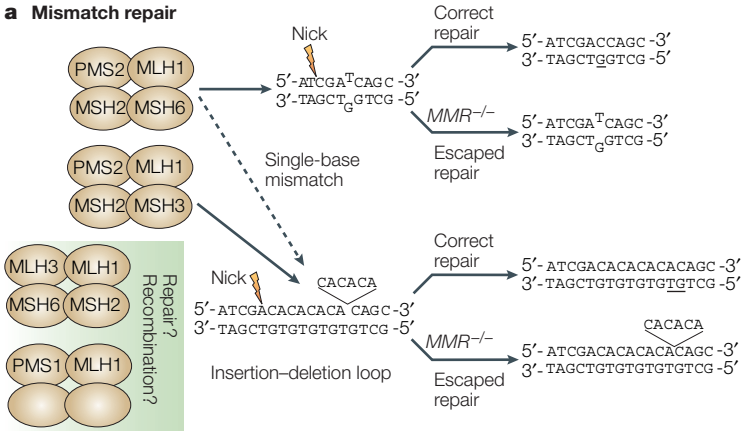
Replication

The association of TNR instability with proliferation, strongly supports a role for replication in repeat instability[25,27,80]; in fact, proliferation and/or replication are required for instability in several model systems[81–83]. The formation of unusual DNA structures and DNA slippage during lagging-strand synthesis might facilitate instability (FIG. 3a). Support for the involvement of the lagging strand in TNR instability comes from the similarity between the instability threshold and OKAZAKI FRAGMENT length, the altered instability of repeat tracts in yeast *rad27/fen1* mutants (flap exo- or endonucleases that are involved in Okazaki processing), and from the propensity of replication-forks to pause at repeat tracts. Treatment of DM1 patient cells with drugs that alter replication-fork progression affects the ongoing CTG expansion, leaving the normal *DM1* allele and other repeat loci unaltered[82]. These data strongly support, at least in DM1, a role for the perturbation of replication-fork dynamics (reviewed in REF. 79) in TNR instability.

OKAZAKI FRAGMENT
Short DNA fragments (~140 nucleotides for primates) that are produced on the lagging strand of the replication fork during discontinuous DNA replication; these fragments are eventually processed (to remove the RNA primer) and ligated to form the mature full-length lagging nascent strand.

The location of replication origins relative to the repeat tract might be crucial for instability in proliferating cells. Switching the location of the origin (origin switch) determines the direction of replication through the repeat and in turn defines the lagging-strand template; factors that affect repeat tract stability in several model systems. An unclear association of TNRs with regions of active replication[84] indicates a complex relationship that is beyond simple replication direction.

Shifting the location of the origin, but not the direction of replication, might also affect instability[81] (origin shift) by altering the location of the repeat relative to the Okazaki initiation zone (OIZ) — the single-stranded region of the lagging-strand template. The influence of this position on the formation of mutagenic intermediates[85] might influence instability. However, given the stochastic nature of Okazaki initiation sites that are large distances from the origin, this effect would only apply to origin shifts at locations that are proximal to the repeat



Repair status	Microsatellites*	Point mutations*	Transgenic mice CTG/CAG
MMR$^{+/+}$	Stable	Stable	Expansion bias
MSH2$^{-/-}$	Destabilized‡	Increased	Stabilized or deletion bias§
MSH6$^{-/-}$	Stable	Increased	Enhanced expansions or stabilized§
MSH3$^{-/-}$	Weakly destabilized	Stable	Stabilized
PMS2$^{-/-}$	Destabilized	Increased	Mildly stabilized and very rare large deletions (>20 repeats)

*For most organisms. ‡Small changes (± 1–3 repeats). §Different observations for different mice.

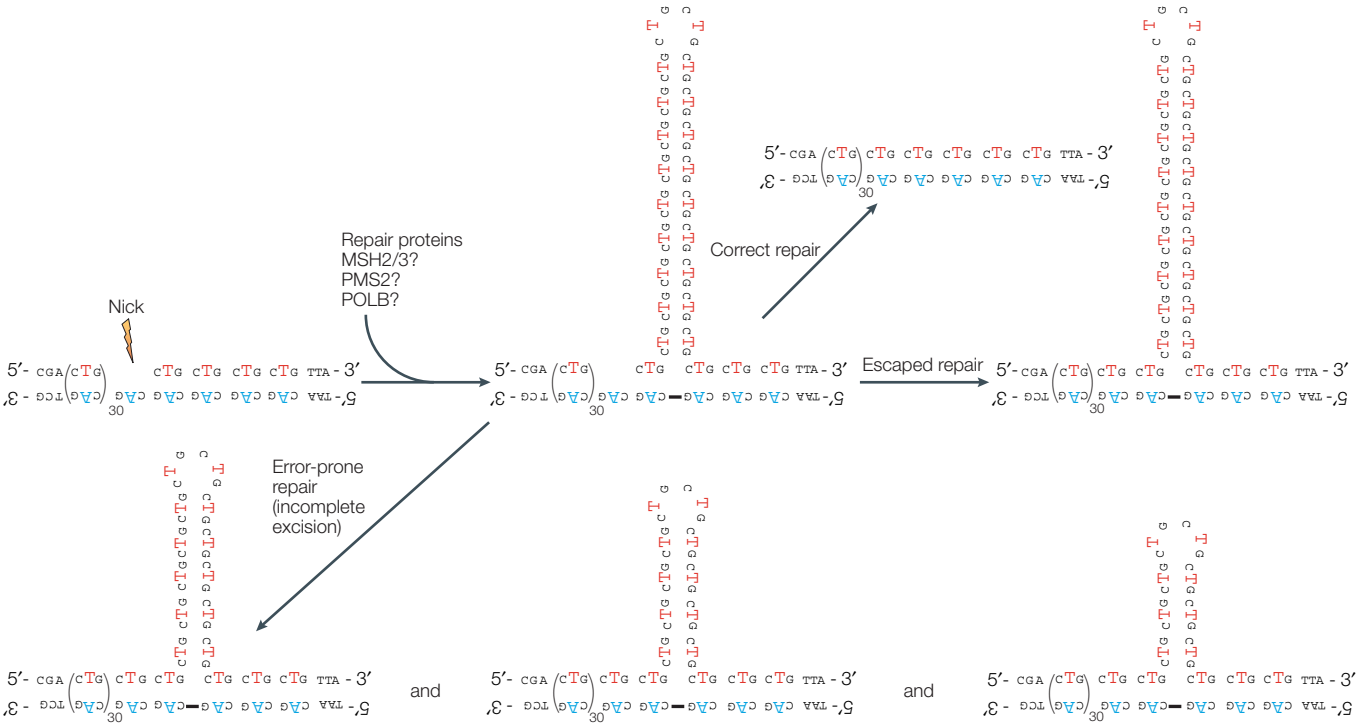


Figure 4 | **Mismatch repair and trinucleotide repeat instability. a** | Processing single-base mismatches is accomplished with a mismatch repair (MMR) complex that involves MSH2 and MSH6. Insertion–deletion loops might also involve MSH2 and MSH6, but are more likely to use an MMR complex that contains MSH2 and MSH3. Correct and escaped repair scenarios are shown for both events. **b** | Repair of damage-induced nicks or nicks that occur during replication (Okazaki fragments) of trinucleotide repeats. MMR proteins that are involved might be able to correctly process the structures that are formed. Alternatively, the presence of the structure might result in escaped repair or error-prone repair, both ending in instability. Repair processes (correct, escaped and error-prone) occur independently of MSH2, MSH3, MLH1, XPF and XPG. PMS1/PMS2, post-meiotic segregation 1 and 2; POLB, polymerase-β; XPF/XPG, xeroderma pigmentosum, complementation group F and G.

tract (<0.35 kb). Alternatively, *cis*-elements that mediate a change in the location of the repeat relative to the OIZ would influence repeat instability independent of origin location (fork shift). These models can be influenced by repeat sequence, tract size, flanking *cis*-elements, the epigenetic environment and by the accompanying cellular and biochemical processes (that is, expression of repair proteins), all of which might vary between and within proliferative tissues and developmental stages.

The pausing of the replication fork within the repeat tract, either as a cause or result of mutagenic intermediate formation, has also been linked to instability. Biochemical[86], bacterial[87] and yeast[88,89] studies reveal that pausing is influenced by repeat length, purity and replication direction. The recovery of paused replication forks might involve slippage events and/or double-strand breaks (DSBs) within or proximal to the repeat tract. Induced DSBs lead to repeat instability[90] and to deletions that extend into the flanking sequence[91,92]. Restarting replication forks that are paused at repeats might require repair and/or recombination processes, both of which might result in further TNR instability.

The ability of proteins, such as FEN1, to process replication-induced errors is crucial to expansions. FEN1 is a structure-specific nuclease that is required to process single-strand overhangs (flaps) of Okazaki fragments on the lagging strands of replication forks. TNR secondary structures inhibit FEN1 processing in a length-dependent manner[93–95], such that ligation of these products without removal of the extra strand-displaced repeats would lead to expansions. The increased frequency of CAG/CTG expansions in *fen1* null yeast and defective endonuclease (not exonuclease) activities supports the idea that FEN1 normally protects against CAG/CTG expansions[96,97]. The degree to which FEN1 influences CAG/CTG instability varies greatly between species: *Fen1*$^{+/-}$ mice show only a minor effect on instability[98] with changes of ±1 to 2 repeat units[97], which is in contrast to the near doubling of a $(CTG)_{85}$ tract in yeast[95,96]; whereas in flies, *Fen1* deficiency has no effect on CAG/CTG instability[99].

Other replication-associated proteins might also participate in TNR instability (see supplementary information S2 (table)). For example, the BLM or WRN helicase can enhance FEN1 cleavage of CTG flaps[100]. Furthermore, Dna2, a helicase, tracks along free ends and has flap-cleaving activity that is blocked by CTG structures[101]. Like Fen1, Dna2 might protect against CAG/CTG expansions, although repeat instability is unaffected in *dna2* yeast mutants[96,102]. Other replication-associated proteins might contribute to repeat instability, but could do so more in line with a role in repair, such as DNA polymerases or ligase to complete the processing of aberrant repair tracts.

Repair

The accumulation of CAG instability in non-mitotic patient tissues, such as brain tissues and the altered instability in various repair-deficient mouse tissues (mitotic and non-mitotic)[5,33,41,98] supports the involvement of various DNA repair processes in CAG instability (see supplementary information S2 (table)). Repair-mediated TNR instability might involve the processing (aberrant repair) of DNA mutagenic intermediates that are formed at replication forks, sites of double-strand or single-strand breaks, or endogenous DNA damage (for example, oxidative) within or near the repeat tract (FIG. 3b).

Although repair of chemically or environmentally induced damage has also been implicated in TNR instability[33,103], these insults also activate several other cellular pathways. The contrast between the consequence of genome-wide instability that is due to repair deficiency and locus-specific repeat instability implies that the cell's ability to process repeat-specific mutagenic intermediates, rather than a repair-pathway defect, drives locus-specific instability[104–107].

Mismatch repair. MMR proteins are implicated in CAG/CTG instability (FIG. 4b). This mutagenic role contrasts with the corrective repair role of these proteins at almost all other sequences, including non-CNG repeats. Strangely, the expansion mutations of CAG/CTG transgenes in numerous mouse models requires the functional MMR proteins MSH2, MSH3 and PMS2 (REFS 5,41,52,108). A role for a MSH2:MSH3 complex in CAG/CTG instability was presumed, but not proved, on the basis of the increased instability in an MSH6-deficient background — an absence of MSH6 might increase amounts of MSH2:MSH3 complexes to enhance expansions[31]. In a recent report, an independent CAG transgenic mouse was used to demonstrate stabilization in the absence of MSH3, but failed to detect increased instability in the absence of MSH6 (REF. 109). It has been shown that MSH2 (REF. 110) and MSH2:MSH3 complexes[109] bind CAG hairpins, although the significance of this binding to either repair or instability remains unclear — particularly as the correct or escaped processing of slipped repeats occurs independently of either protein[53] (see below). The MSH2:MSH3 complex directs the repair of insertion–deletion loops of up to 13 nucleotides. This process is tightly linked with DNA-binding-dependent ATPase activity of MSH2:MSH3, which is regulated by a mismatch-stimulated ADP-to-ATP exchange[111]. Recently it was suggested that the MSH2:MSH3 complex binds to $(CAG)_{13}$ hairpins but with reduced ATPase activity relative to a preferred short $(CA)_4$ loop[109]. One interpretation of this is that the ATPase activity of MSH2 is not required for its role in CAG expansions. If true, crossing CAG transgenic mice to mice that harbour the ATPase-defective MSH2 protein that retains full mismatch-binding activity[112] would produce mismatch repair-deficient mice that show spontaneous CAG expansions, unlike MSH2 null mice[52]. A deficiency of PMS2, which in complex with MLH1 can bind ssDNA[113], was also required for CAG/CTG expansions, but to a lesser degree than MSH2 or MSH3 (REF. 33). So PMS2 might be functionally redundant with PMS1 and/or MLH3. Accumulation of CAG/CTG expansions in mitotic, non-mitotic (brain) and germline tissues depends on MSH2, MSH3 and PMS2 (REFS 5,6,31,33,41,52). In the absence of MMR, CAG/CTG tracts were either mildly or completely stabilized[31,41,52,114] or in some mice tissues

tended to delete[5,6]. The variable degrees of instability and effects of MMR-deficiencies observed in these mouse models might arise as a result of specific differences such as tract length, integration site, replication direction, transcription direction and mouse strain differences, and/or because of the interdependence of the MMR proteins. Although the *in vivo* data clearly indicate a requirement for MMR proteins in CAG instability, they do not indicate a mechanism or pathway through which these proteins might function.

MSH2:MSH3 might be required for correct repair of slipped CTG or CAG repeats, but many mutagenic intermediates escape this process and lead to expansions. In this case the absence of functional MMR would be expected to increase expansions. MSH2:MSH3 could bind to slipped CTG or CAG repeats that protect them from being repaired and in this way contribute to repeat expansion. In this case the absence of MMR would lead to correct repair and a reduction in expansions. And finally MSH2:MSH3 might be required to process the slip-outs in an error-prone manner, thereby giving rise to expansions. None of these scenarios is consistent with the mouse data, particularly the repeat contractions that arise in the absence of MSH2 or MSH3 (REFS 5,6).

Mechanistic insights have recently been provided by *in vitro* analysis of the fidelity by which slipped CTG or CAG intermediates are processed by human cell proteins[53]. Correct repair, escaped repair and error-prone repair were observed. Repair path depended on slip-out composition (CAG or CTG) and nick location (nicks in the same or opposite strand as the slip-outs or nicks upstream or downstream of the slip-outs). Importantly, all outcomes were independent of human MSH2, MSH3, MLH1, XPF or XPG proteins (FIG. 4b). Correct repair would stably maintain the repeat; although slippage mutation events might arise they are corrected and so not detected. Escaped repair of slip-outs could lead to expansion or deletion products. Although escaped repair might involve masking of slipped DNAs from repair by bound proteins, MSH2, MSH3 or PMS2 are not involved in this process. Interestingly, a new form of error-prone repair was detected in which excess repeats in expansion intermediates (slip-outs in a nicked strand) were incompletely excised — presenting a new way to generate expansions. Error-prone repair might be a source for the expansion bias that is observed in patients, as deletion intermediates (slip-outs that are opposite a nicked strand) do not yield error-prone products. Although neither repair outcome requires MMR proteins, these proteins might contribute to CAG/CTG expansions by driving the formation of slipped DNAs, possibly in conjunction with strand displacement from a nick by polymerase-β[115] or with DNA helicases during the unwinding or annealing of DNA strands[116–120] (FIG. 4b). Once formed, these expansion intermediates would be processed in an error-prone manner, the excess retained repeats would be incorporated as expansions. Interestingly, human neuron-like cell extracts yielded each repair outcome, supporting a role for these processes in CTG/CAG instability in post-mitotic brain cells of patients[53].

MMR and the direction of replication. Several MMR proteins are associated with proliferating cell nuclear antigen (PCNA), which probably brings these proteins in close contact with the replication machinery at replication forks. In some MMR-deficient bacteria and yeast, the contraction frequency is altered for CAG but not CTG lagging-strand templates, which indicates that the contribution of MMR to instability might depend on replication direction[121–123] (see supplementary information S2 (table)). Similar replication-direction-dependent effects were observed for deficiencies of *rad27*/*fen1* and DNA ligase[102,124]. Differences in the ability of repair proteins to process the different DNA intermediates that are formed when replication proceeds in one direction versus the other (CTG versus CAG) might explain the sensitivity of replication to direction.

Double-strand break repair. Repair of DSBs is essential for genomic stability and cell survival. In yeast, TNR tracts are susceptible to DSBs, possibly through aberrant endonucleolytic cleavage of DNA intermediates[90,125]. The subsequent repair of these breaks might lead to instability and in some cases yields large contractions of the repeat and flanking sequences[90,125,126]. Studies in bacteria, primate and rodent cells revealed that the repair of a DSB within a CTG/CAG-repeat tract can lead to large contractions that frequently include all or most of the repeat and some flanking non-repetitive sequence[91,92,127]. Such rare deletion events are known to occur only at the FRAXA locus, and might be mediated by methylation and replication[163]. Repair of DSBs that are induced by replication defects associated with TNRs, such as slippage and fork pausing[87–89], might lead to instability[90,126,128], possibly through GENE CONVERSION and SINGLE-STRAND ANNEALING.

Nucleotide excision repair, transcription and trinucleotide repeat instability. Nucleotide excision repair (NER) is specific for DNA damage that is caused by ultraviolet light and by certain DNA-distorting lesions. Bacterial models indicate a role for NER in either protecting or enhancing large CTG/CAG contractions, phenomena that might depend on transcription, replication direction and repeat tract[129,130]. However human XERODERMA PIGMENTOSUM cells with defective NER proteins did not show repeat instability at the *DM1*, *SBMA*, *MJD* or *HD* loci[107]. Moreover, processing slipped-CTG or slipped-CAG repeats by human proteins did not require either of the structure-specific nucleases XPF or XPG[53].

Given that each disease-associated repeat tract is transcribed, transcription levels and transcription-coupled repair (TCR, which is preferential repair of the transcribed strand) and differentiation-altered TCR might contribute to repeat instability[34]. Bacterial models revealed that transcription increased the levels of large TNR contractions. These effects were sensitive to replication direction[131–133] — transcription progression might collide head-to-head or head-to-tail with replication forks. The positive correlation between transcription levels and tissue-specific instability in

GENE CONVERSION
A meiotic process that involves the non-reciprocal transfer of genetic information, in which one allele directs the conversion of its sister allele.

SINGLE-STRAND ANNEALING
A process that is typically initiated after a double-strand break, in which regions of ssDNA are created to allow for complementary strands of sister chromatids or homologues to anneal to each other.

XERODERMA PIGMENTOSUM
A human disease that is characterized by extraordinary sensitivity to sunlight, and is caused by a defect in the ultraviolet-mutation-repair system.

patients who have SBMA [55] and the absence of such a correlation in DM1 mice[32] could suggest that the effects of transcription on instability varies between loci.

Protein functional diversity and repair pathways. Given the unorthodox implication that DNA repair proteins might contribute to repeat instability, elucidating the mechanistic role of these factors in this mutagenic processes is challenging, not least owing to the functional diversity of many repair proteins. For example, in addition to correcting mismatches, MMR proteins participate in NER, base excision repair, DSB repair, transcription-coupled repair, meiotic recombination, blocking recombination between non-identical sequences, single-strand annealing and the apoptotic response to certain DNA-damaging agents[134]. Interestingly, MMR proteins also contribute to locus-specific mutations that naturally occur at immunoglobin genes[135]. Whether there are mechanistic parallels between the mutagenic roles of MMR proteins in these processes and repeat instability has still to be determined. When looking for their mechanistic contribution to TNR instability, one should avoid restricting their functional scope to the name that they hold (that is, mismatch), or to the functions that they have originally been attributed (that is, repair).

Recombination

Recombination between sequences that flank the repeat tract is rare in most TNR disorders[136,137]. Although the loss of many recombination proteins in many systems[5,90,138] does not alter instability, limited data from patients, and bacterial[139], yeast[126] and rodent[127] models do support a contribution of recombination to TNR instability. In yeast, meiotic recombination might be more important than mitotic events[125,140,141]: instability might arise through DSBs that are induced within repeats by the meiotic-specific endonuclease SPO11 (REF. 142). In humans, the massive germline expansions at the *DM1*, *DM2*, *FRAXA*, *SCA8*, *SCA10* and *FRDA* loci might involve recombination within the tract, probably by intrachromosomal strand annealing (between sister chromatids) as a result of DNA synthesis-dependent strand annealing (FIG. 3b). The precise role of recombination in instability is difficult to determine because it is difficult to detect recombination events that have break points within the repeat tracts.

Recombination has a more definitive role in the instability of repeats with unusual and/or interrupted configurations[143]. Deletions in the INS variable number tandem repeat minisatellite arose through a simple intra-allelic deletion, probably as a result of slippage or unequal crossover between sister chromatids with misaligned repeats[144]. By contrast, INS pre-meiotic germline expansions are complex, involving intra-allelic rearrangements with extensive remodelling of the repeat and inter-allelic repeat exchanges. Interestingly, half the *de novo* inherited cases of facioscapulohumeral muscular dystrophy (FSHD) can be attributed to high levels of post-fertilization somatic (recombinogenic) mutations in the associated 3.3 kb D4Z4 megasatellite tract[19,145]. Patients who have FSHD with D4Z4 length heterogeneity harbour cells that have at least two distinct disease allele-contraction sizes, which supports the involvement of mitotic gene conversion with or without crossover. In addition, both intrachromosomal (between sister chromatids, probably through DNA synthesis-dependent strand annealing) and interchromosomal rearrangements (between chromosome 4q and 10q D4Z4 repeats) can occur. Although it is often speculated that recombination has a greater role in polyalanine expansion[143], there is no experimental evidence to favour this pathway over others, such as replicative slippage. In fact, the unusual configuration of these repeat tracts, absent in most TNR disorders, allows for an easier identification, analysis and tracking of recombination to repeat instability.

Conclusions

Deconstructing the mechanism(s) of repeat instability has proved exceedingly complex; this is in part due to the overlapping contributions of each DNA metabolic process and in part due to difficulties in identifying the precise timing of instability. The contribution of these processes to instability seems to be unique not only to the specific developmental stage, tissue or proliferation status of the affected cell but also to the specific disorder.

Disease-associated repeat expansion is a phenomenon that seems to be uniquely human. A wide range of model systems, each with its own inherent strengths and limitations have been used to understand what 'drives' and/or maintains repeat instability. A comparison of the combined data that are obtained through these various model systems — taking into account their limitations — with observations from human studies will allow a greater understanding of the underlying mechanism(s) of repeat instability. Given the numerous differences between each repeat sequence and each disease locus, as well as the differences between models, generalizations of mechanism(s) should be made with caution.

Many challenging and important questions remain. Factors that contribute to tissue-specific instability, specifically in primary affected tissues of patients and/or appropriate non-human models (muscle, CNS and germ line) need to be identified. Locus-specific *cis*-elements and *trans*-factors that drive instability need to be identified and the timing, pattern and tissue selectivity of repeat instability determined. How naturally occurring contractions of expanded repeat tracts arise is not understood, neither in the germ line nor in the soma. In the long term, analysis of these contractions might reveal a means to harness this process for clinical benefits.

Exogenous agents (genetic, RNAi, chemical, nutritional, environmental or others) that specifically target the expansion or contraction processes in somatic and germline lineages remain to be identified. A long-term goal here is to modulate instability (levels and directions) in the hope of clinical benefits. It is likely that the answers to these questions will reveal surprising genetic and biological phenomena, and probably more questions.

1. Pearson, C. E. Slipping while sleeping? Trinucleotide repeat expansions in germ cells. *Trends Mol. Med.* **9**, 490–495 (2003).
2. 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).
By using laser-capture microdissection and repeat-length analysis by single-cell PCR the authors advanced our understanding of the timing (pre-meiotic) and mechanism of CAG expansions and parent-of-origin (paternal) expansion biases.
3. Mangiarini, L. *et al.* Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nature Genet.* **15**, 197–200 (1997).
4. Zhang, Y., Monckton, D. G., Siciliano, M. J., Connor, T. H. & Meistrich, M. L. Age and insertion site dependence of repeat number instability of a human *DM1* transgene in individual mouse sperm. *Hum. Mol. Genet.* **11**, 791–798 (2002).
5. Savouret, C. *et al.* CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J.* **22**, 2264–2273 (2003).
6. 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).
An important study on the timing (pre-meiotic) of male germline expansions/contractions and the involvement of the mismatch repair protein MSH2.
7. Kaytor, M. D., Burright, E. N., Duvick, L. A., Zoghbi, H. Y. & Orr, H. T. Increased trinucleotide repeat instability with advanced maternal age. *Hum. Mol. Genet.* **6**, 2135–2139 (1997).
8. Malter, H. E. *et al.* Characterization of the full fragile X syndrome mutation in fetal gametes. *Nature Genet.* **15**, 165–169 (1997).
The authors provide an insight into the timing of the parent-of-origin effect, including maternal expansion bias and paternal contraction bias, and the potential role of CpG methylation in CGG instability.
9. Moutou, C., Vincent, M. C., Biancalana, V. & Mandel, J. L. Transition from premutation to full mutation in fragile X syndrome is likely to be prezygotic. *Hum. Mol. Genet.* **6**, 971–979 (1997).
10. De Temmerman, N. *et al.* Intergenerational instability of the expanded CTG repeat in the *DMPK* gene: studies in human gametes and preimplantation embryos. *Am. J. Hum. Genet.* **75**, 325–329 (2004).
This paper gives an insight into the timing and mechanism of parent-of-origin maternal expansion bias of a CTG repeat and the timing of post-fertilization CTG instability.
11. Lenzi, M. L. *et al.* Extreme heterogeneity in the molecular events leading to the establishment of chiasmata during meiosis I in human oocytes. *Am. J. Hum. Genet.* **76**, 112–127 (2005).
12. Martorell, L., Johnson, K., Boucher, C. A. & Baiget, M. Somatic instability of the myotonic dystrophy $(CTG)_n$ repeat during human fetal development. *Hum. Mol. Genet.* **6**, 877–880 (1997).
An important insight into the timing of somatic CTG instability and the tissue specificity of this instability, some of which can continue throughout the lifetime of the adult.
13. Reyniers, E. *et al.* Postmortem examination of two fragile X brothers with an *FMR1* full mutation. *Am. J. Med. Genet.* **84**, 245–249 (1999).
14. Nichol Edamura, K., Leonard, M. R. & Pearson, C. E. Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells. *Am. J. Hum. Genet.* **76**, 302–311 (2005).
15. Carbonell, P. *et al.* FRAXE mutation analysis in three Spanish families. *Am. J. Med. Genet.* **64**, 434–440 (1996).
16. Moseley, M. L. *et al.* SCA8 CTG repeat: *en masse* contractions in sperm and intergenerational sequence changes may play a role in reduced penetrance. *Hum. Mol. Genet.* **9**, 2125–2130 (2000).
17. Monckton, D. G., Wong, L. J., Ashizawa, T. & Caskey, C. T. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum. Mol. Genet.* **4**, 1–8 (1995).
This study uses the ultra-sensitive technique of small pool PCR to qualitatively and quantitatively assess CTG-length heterogeneity, allowing the detection of the products of rare events. It also provides an insight into germline contraction events.
18. De Michele, G. *et al.* Parental gender, age at birth and expansion length influence GAA repeat intergenerational instability in the *X25* gene: pedigree studies and analysis of sperm from patients with Friedreich's ataxia. *Hum. Mol. Genet.* **7**, 1901–1906 (1998).
19. Lemmers, R. J. *et al.* Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **75**, 44–53 (2004).
20. Benitez, J. *et al.* Somatic stability in chorionic villi samples and other Huntington fetal tissues. *Hum. Genet.* **96**, 229–232 (1995).
21. Jedele, K. B. *et al.* Spinal and bulbar muscular atrophy (SBMA): somatic stability of an expanded CAG repeat in fetal tissues. *Clin. Genet.* **54**, 148–151 (1998).
22. Devys, D. *et al.* Analysis of full fragile X mutations in fetal tissues and monozygotic twins indicate that abnormal methylation and somatic heterogeneity are established early in development. *Am. J. Med. Genet.* **43**, 208–216 (1992).
23. Taylor, A. K. *et al.* Tissue heterogeneity of the *FMR1* mutation in a high-functioning male with fragile X syndrome. *Am. J. Med. Genet.* **84**, 233–239 (1999).
Together with reference 13, this study contributed to the understanding of the timing and tissue-specificity of germline and somatic mosaicism and CpG methylation mosaicism. It indicates a role of CpG methylation in CGG instability.
24. Pollard, L. M. *et al.* Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. *Nucleic Acids Res.* **32**, 5962–5971 (2004).
25. Wohrle, D. *et al.* Heterogeneity of DM kinase repeat expansion in different fetal tissues and further expansion during cell proliferation *in vitro*: evidence for a casual involvement of methyl-directed DNA mismatch repair in triplet repeat stability. *Hum. Mol. Genet.* **4**, 1147–1153 (1995).
26. Thornton, C. A., Johnson, K. & Moxley, R. T. 3rd. Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann. Neurol.* **35**, 104–107 (1994).
27. Zatz, M. *et al.* Analysis of the CTG repeat in skeletal muscle of young and adult myotonic dystrophy patients: when does the expansion occur? *Hum. Mol. Genet.* **4**, 401–406 (1995).
28. Martorell, L., Martinez, J. M., Carey, N., Johnson, K. & Baiget, M. Comparison of CTG repeat length expansion and clinical progression of myotonic dystrophy over a five year period. *J. Med. Genet.* **32**, 593–596 (1995).
29. Kinoshita, M. *et al.* A patient with myotonic dystrophy type 1 (DM1) accompanied by laryngeal and renal cell carcinomas had a small CTG triplet repeat expansion but no somatic instability in normal tissues. *Intern. Med.* **41**, 312–318 (2002).
30. Osanai, R., Kinoshita, M., Hirose, K., Homma, T. & Kawabata, I. CTG triplet repeat expansion in a laryngeal carcinoma from a patient with myotonic dystrophy. *Muscle Nerve* **23**, 804–806 (2000).
31. 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).
32. 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).
33. Gomes-Pereira, M., Fortune, M. T. & Monckton, D. G. 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 (2001).
34. Nouspikel, T. & Hanawalt, P. C. DNA repair in terminally differenetiated cells. *DNA Repair* **1**, 59–75 (2002).
35. Vinson, R. K. & Hales, B. F. DNA repair during organogenesis. *Mutat. Res.* **509**, 79–91 (2002).
36. Cleary, J. D. & Pearson, C. E. The contribution of *cis*-elements to disease-associated repeat instability: Clinical and experimental evidence. *Cytogenet. Genome Res.* **100**, 25–55 (2003).
37. Watase, K., Venken, K. J., Sun, Y., Orr, H. T. & Zoghbi, H. Y. Regional differences of somatic CAG repeat instability do not account for selective neuronal vulnerability in a knock-in mouse model of SCA1. *Hum. Mol. Genet.* **12**, 2789–2795 (2003).
38. Sato, T. *et al.* 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 (1999).
39. 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).
40. Al-Mahdawi, S. *et al.* GAA repeat instability in Friedreich ataxia YAC transgenic mice. *Genomics* **84**, 301–310 (2004).
41. Wheeler, V. C. *et al.* Mismatch repair gene *Msh2* modifies the timing of early disease in Hdh^{Q111} striatum. *Hum. Mol. Genet.* **12**, 273–281 (2003).
42. Dobbing, J. & Sands, J. Comparative aspects of the brain growth spurt. *Early Hum. Dev.* **3**, 79–83 (1979).
43. Maciel, P., Lopes-Cendes, I., Kish, S., Sequeiros, J. & Rouleau, G. A. Mosaicism of the CAG repeat in CNS tissue in relation to age at death in spinocerebellar ataxia type 1 and Machado–Joseph disease patients. *Am. J. Hum. Genet.* **60**, 993–996 (1997).
44. 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).
45. Takano, H. *et al.* Somatic mosaicism of expanded CAG repeats in brains of patients with dentatorubral-pallidoluysian atrophy: cellular population-dependent dynamics of mitotic instability. *Am. J. Hum. Genet.* **58**, 1212–1222 (1996).
46. Hashida, H. *et al.* Single cell analysis of CAG repeat in brains of dentatorubral-pallidoluysian atrophy (DRPLA). *J. Neurol. Sci.* **190**, 87–93 (2001).
47. Watanabe, H. *et al.* Differential somatic CAG repeat instability in variable brain cell lineage in dentatorubral pallidoluysian atrophy (DRPLA): a laser-captured microdissection (LCM)-based analysis. *Hum. Genet.* **107**, 452–457 (2000).
In references 46 and 47 the authors used high-selective laser-capture microdissection to study neuronal and non-neuronal somatic instability in patients' brains, and the timing and tissue-specificity of repeat expansions and contractions.
48. Curtis, M. A. *et al.* Increased cell proliferation and neurogenesis in the adult human Huntington's disease brain. *Proc. Natl Acad. Sci. USA* **100**, 9023–9027 (2003).
49. Kahlem, P. & Djian, P. The expanded CAG repeat associated with juvenile Huntington disease shows a common origin of most or all neurons and glia in human cerebrum. *Neurosci. Lett.* **286**, 203–207 (2000).
50. Kennedy, L. & Shelbourne, P. F. Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Hum. Mol. Genet.* **9**, 2539–2544 (2000).
51. Lopes-Cendes, I. *et al.* Somatic mosaicism in the central nervous system in spinocerebellar ataxia type 1 and Machado–Joseph disease. *Ann. Neurol.* **40**, 199–206 (1996).
52. 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).
The authors propose the active contribution of the MSH2 mismatch repair protein to the mutation of repeat sequences — a model that contrasts with this protein's role in maintaining genome integrity. They also reveal the effects of tissue-specificity of MSH2 on somatic CAG instability.
53. 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).
This paper provides direct evidence for the processing of slipped DNAs, through a potential mutagenic intermediate of instability, and reveals mechanistic pathways through which this processing might lead to repeat stability, expansion and contraction.
54. Hashida, H., Goto, J., Kurisaki, H., Mizusawa, H. & Kanazawa, I. Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: dentatorubral-pallidoluysian atrophy, Machado–Joseph disease, and spinocerebellar ataxia type 1. *Ann. Neurol.* **41**, 505–511 (1997).
55. Tanaka, F. *et al.* Tissue-specific somatic mosaicism in spinal and bulbar muscular atrophy is dependent on CAG-repeat length and androgen receptor-gene expression level. *Am. J. Hum. Genet.* **65**, 966–973 (1999).
56. Ansved, T., Lundin, A. & Anvret, M. Larger CAG expansions in skeletal muscle compared with lymphocytes in Kennedy disease but not in Huntington disease. *Neurology* **51**, 1442–1444 (1998).

57. Montermini, L., Kish, S. J., Jiralerspong, S., Lamarche, J. B. & Pandolfo, M. Somatic mosaicism for Friedreich's ataxia GAA triplet repeat expansions in the central nervous system. *Neurology* **49**, 606–610 (1997).
58. Tassone, F., Hagerman, R. J., Gane, L. W. & Taylor, A. K. Strong similarities of the *FMR1* mutation in multiple tissues: postmortem studies of a male with a full mutation and a male carrier of a premutation. *Am. J. Med. Genet.* **84**, 240–244 (1999).
59. Chong, S. S. *et al.* Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nature Genet.* **10**, 344–350 (1995).
60. Pearson, C. E. *et al.* Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry* **37**, 2701–2708 (1998).
61. Brock, G. J., Anderson, N. H. & Monckton, D. G. *Cis*-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum. Mol. Genet.* **8**, 1061–1067 (1999).
62. Gourdon, G., Dessen, P., Lia, A. S., Junien, C. & Hofmann-Radvanyi, H. Intriguing association between disease associated unstable trinucleotide repeat and CpG island. *Ann. Genet.* **40**, 73–77 (1997).
63. Monckton, D. G., Coolbaugh, M. I., Ashizawa, K. T., Siciliano, M. J. & Caskey, C. T. Hypermutable myotonic dystrophy CTG repeats in transgenic mice. *Nature Genet.* **15**, 193–196 (1997).
64. Gourdon, G. *et al.* Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice. *Nature Genet.* **15**, 190–192 (1997).
65. La Spada, A. R. *et al.* Androgen receptor YAC transgenic mice carrying CAG 45 alleles show trinucleotide repeat instability. *Hum. Mol. Genet.* **7**, 959–967 (1998).
66. Libby, R. T. *et al.* Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. *Hum. Mol. Genet.* **12**, 41–50 (2003).
This article demonstrates the need for flanking sequence for the active instability to occur in both germline and somatic tissues and confirms the implication (from haplotype analysis) that *cis*-elements contribute to repeat instability.
67. Seznec, H. *et al.* 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 (2000).
68. Sutcliffe, J. S. *et al.* DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum. Mol. Genet.* **1**, 397–400 (1992).
69. Willemsen, R., Bontekoe, C. J., Severijnen, L. A. & Oostra, B. A. Timing of the absence of *FMR1* expression in full mutation chorionic villi. *Hum. Genet.* **110**, 601–605 (2002).
70. Wohrle, D. *et al.* Unusual mutations in high functioning fragile X males: apparent instability of expanded unmethylated CGG repeats. *J. Med. Genet.* **35**, 103–111 (1998).
71. Burman, R. W., Popovich, B. W., Jacky, P. B. & Turker, M. S. Fully expanded *FMR1* CGG repeats exhibit a length- and differentiation-dependent instability in cell hybrids that is independent of DNA methylation. *Hum. Mol. Genet.* **8**, 2293–2302 (1999).
72. Salat, U., Bardoni, B., Wohrle, D. & Steinbach, P. Increase of FMRP expression, raised levels of *FMR1* mRNA, and clonal selection in proliferating cells with unmethylated fragile X repeat expansions: a clue to the sex bias in the transmission of full mutations? *J. Med. Genet.* **37**, 842–850 (2000).
73. Nichol, K. & Pearson, C. E. CpG methylation modifies the genetic stability of cloned repeat sequences. *Genome Res.* **12**, 1246–1256 (2002).
74. Gorbunova, V., Seluanov, A., Mittelman, D. & Wilson, J. H. Genome-wide demethylation destabilizes CTG. CAG trinucleotide repeats in mammalian cells. *Hum. Mol. Genet.* **13**, 2979–2989 (2004).
75. Wang, Y. H., Amirhaeri, S., Kang, S., Wells, R. D. & Griffith, J. D. Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science* **265**, 669–671 (1994).
76. Mulvihill, D. J., Edamura, K. N., Hagerman, K. A., Pearson, C. E. & Wang, Y. H. Effect of CAT or AGG interruptions and CpG methylation on nucleosome assembly upon trinucleotide repeats on spinocerebellar ataxia type 1 and fragile X syndrome. *J. Biol. Chem.* **280**, 4498–4503 (2005).
77. Otten, A. D. & Tapscott, S. J. Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc. Natl Acad. Sci. USA* **92**, 5465–5469 (1995).
78. Filippova, G. N. *et al.* CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nature Genet.* **28**, 335–343 (2001).
79. Cleary J. D., Nichol Edamura, K. & Pearson C. E. The complex nature of trinucleotide repeat instability. *Chemtracts Biochem. Mol. Biol.* **17**, 663–676 (2004).
80. Martorell, L. *et al.* Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Hum. Mol. Genet.* **7**, 307–312 (1998).
81. 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).
This work reveals the complex influence of DNA replication on CTG instability. Along with works such as reference 87, it highlights the importance of replication-fork perturbation.
82. Yang, Z., Lau, R., Marcadier, J. L., Chitayat, D. & Pearson, C. E. Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells. *Am. J. Hum. Genet.* **73**, 1092–1105 (2003).
83. Freudenreich, C. H. & Lahiri, M. Structure-forming CAG/CTG repeat sequences are sensitive to breakage in the absence of Mrc1 checkpoint function and S-phase checkpoint signaling: implications for trinucleotide repeat expansion diseases. *Cell Cycle* **3**, 1370–1374 (2004).
84. Nenguke, T., Aladjem, M. I., Gusella, J. F., Wexler, N. S. & Arnheim, N. Candidate DNA replication initiation regions at human trinucleotide repeat disease loci. *Hum. Mol. Genet.* **12**, 1021–1028 (2003).
85. 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).
86. Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S. & Wells, R. D. Pausing of DNA synthesis *in vitro* at specific loci in CTG and CGG triplet repeats from human hereditary disease genes. *J. Biol. Chem.* **270**, 27014–27021 (1995).
87. Samadashwily, G. M., Raca, G. & Mirkin, S. M. Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.* **17**, 298–304 (1997).
88. Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R. & Mirkin, S. M. Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell. Biol.* **23**, 1349–1357 (2003).
89. Krasilnikova, M. M. & Mirkin, S. M. Replication stalling at Friedreich's ataxia (GAA)n repeats *in vivo*. *Mol. Cell. Biol.* **24**, 2286–2295 (2004).
90. Freudenreich, C. H., Kantrow, S. M. & Zakian, V. A. Expansion and length-dependent fragility of CTG repeats in yeast. *Science* **279**, 853–856 (1998).
91. Marcadier, J. L. & Pearson, C. E. Fidelity of primate cell repair of a double-strand break within a (CTG)•(CAG) tract. Effect of slipped DNA structures. *J. Biol. Chem.* **278**, 33848–33856 (2003).
92. Hebert, M. L., Spitz, L. A. & Wells, R. D. DNA double-strand breaks induce deletion of CTG. CAG repeats in an orientation-dependent manner in *Escherichia coli*. *J. Mol. Biol.* **336**, 655–672 (2004).
93. Spiro, C. *et al.* Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol. Cell* **4**, 1079–1085 (1999).
94. 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).
95. Lee, S. & Park, M. S. Human FEN-1 can process the 5′-flap DNA of CTG/CAG triplet repeat derived from human genetic diseases by length and sequence dependent manner. *Exp. Mol. Med.* **34**, 313–317 (2002).
96. Callahan, J. L., Andrews, K. J., Zakian, V. A. & Freudenreich, C. H. Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. *Mol. Cell. Biol.* **23**, 7849–7860 (2003).
97. 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).
98. Spiro, C. & McMurray, C. T. Nuclease-deficient FEN-1 blocks Rad51/BRCA1-mediated repair and causes trinucleotide repeat instability. *Mol. Cell. Biol.* **23**, 6063–6074 (2003).
99. Jackson, S. M. *et al.* A SCA7 CAG/CTG repeat expansion is stable in *Drosophila melanogaster* despite modulation of genomic context and gene dosage. *Gene* **347**, 35–41 (2005).
100. Wang, W. & Bambara, R. A. Human Bloom protein stimulates flap endonuclease 1 activity by resolving DNA secondary structure. *J. Biol. Chem.* **280**, 5391–5399 (2005).
101. Kao, H. I., Veeraraghavan, J., Polaczek, P., Campbell, J. L. & Bambara, R. A. On the roles of *Saccharomyces cerevisiae* Dna2p and Flap endonuclease 1 in Okazaki fragment processing. *J Biol Chem* **279**, 15014–15024 (2004).
102. Ireland, M. J., Reinke, S. S. & Livingston, D. M. The impact of lagging strand replication mutations on the stability of CAG repeat tracts in yeast. *Genetics* **155**, 1657–1665 (2000).
103. Hashem, V. I. *et al.* Chemotherapeutic deletion of CTG repeats in lymphoblast cells from DM1 patients. *Nucleic Acids Res.* **32**, 6334–6346 (2004).
104. Otto, C. J., Almqvist, E., Hayden, M. R. & Andrew, S. E. The 'flap' endonuclease gene *FEN1* is excluded as a candidate gene implicated in the CAG repeat expansion underlying Huntington disease. *Clin. Genet.* **59**, 122–127 (2001).
105. Kramer, P. R., Pearson, C. E. & Sinden, R. R. Stability of triplet repeats of myotonic dystrophy and fragile X loci in human mutator mismatch repair cell lines. *Hum. Genet.* **98**, 151–157 (1996).
106. Goellner, G. M. *et al.* Different mechanisms underlie DNA instability in Huntington disease and colorectal cancer. *Am. J. Hum. Genet.* **60**, 879–890 (1997).
107. Grewal, R. P. Neurodegeneration in xeroderma pigmentosum: a trinucleotide repeat mutation analysis. *J. Neurol. Sci.* **163**, 183–186 (1999).
108. Kovtun, I. V. & McMurray, C. T. Trinucleotide expansion in haploid germ cells by gap repair. *Nature Genet.* **27**, 407–411 (2001).
109. 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).
110. Pearson, C. E., Ewel, A., Acharya, S., Fishel, R. A. & Sinden, R. R. Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum. Mol. Genet.* **6**, 1117–1123 (1997).
111. Wilson, T. E. & Lieber, M. R. Efficient processing of DNA ends during yeast nonhomologous end joining. Evidence for a DNA polymeraseβ (*Pol4*)-dependent pathway. *J. Biol. Chem.* **274**, 23599–23609 (1999).
112. Lin, D. P. *et al.* An Msh2 point mutation uncouples DNA mismatch repair and apoptosis. *Cancer Res.* **64**, 517–522 (2004).
113. Plotz, G., Raedle, J., Brieger, A., Trojan, J. & Zeuzem, S. hMutSα forms an ATP-dependent complex with hMutLα and hMutLβ on DNA. *Nucleic Acids Res.* **30**, 711–718 (2002).
114. Gomes-Pereira, M., Fortune, M. T., Ingram, L., McAbney, J. P. & Monckton, D. G. *Pms2* is a genetic enhancer of trinucleotide CAG•CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum. Mol. Genet.* **13**, 1815–1825 (2004).
115. Hartenstine, M. J., Goodman, M. F. & Petruska, J. Base stacking and even/odd behavior of hairpin loops in DNA triplet repeat slippage and expansion with DNA polymerase. *J. Biol. Chem.* **275**, 18382–18390 (2000).
116. Gaudreault, I., Guay, D. & Lebel, M. YB-1 promotes strand separation *in vitro* of duplex DNA containing either mispaired bases or cisplatin modifications, exhibits endonucleolytic activities and binds several DNA repair proteins. *Nucleic Acids Res.* **32**, 316–327 (2004).
117. Yang, Q. *et al.* The mismatch DNA repair heterodimer, hMSH2/6, regulates BLM helicase. *Oncogene* **23**, 3749–3756 (2004).
118. Cheok, C. F., Wu, L., Garcia, P. L., Janscak, P. & Hickson, I. D. The Bloom's syndrome helicase promotes the annealing of complementary single-stranded DNA. *Nucleic Acids Res.* **33**, 3932–3941 (2005).
119. Doherty, K. M. *et al.* RECQ1 helicase interacts with human mismatch repair factors that regulate genetic recombination. *J. Biol. Chem.* **280**, 28085–28094 (2005).
120. Sharma, S. *et al.* Biochemical analysis of the DNA unwinding and strand annealing activities catalyzed by human RECQ1. *J. Biol. Chem.* **280**, 28072–28084 (2005).
121. Schmidt, K. H., Abbott, C. M. & Leach, D. R. Two opposing effects of mismatch repair on CTG repeat instability in *Escherichia coli*. *Mol. Microbiol.* **35**, 463–471 (2000).
122. Schweitzer, J. K. & Livingston, D. M. Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. *Hum. Mol. Genet.* **6**, 349–355 (1997).
123. Jaworski, A. *et al.* Mismatch repair in *Escherichia coli* enhances instability of $(CTG)_n$ triplet repeats from human hereditary diseases. *Proc. Natl Acad. Sci. USA* **92**, 11019–11023 (1995).
124. Schweitzer, J. K. & Livingston, D. M. Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. *Hum. Mol. Genet.* **7**, 69–74 (1998).

125. Jankowski, C., Nasar, F. & Nag, D. K. Meiotic instability of CAG repeat tracts occurs by double-strand break repair in yeast. *Proc. Natl Acad. Sci. USA* **97**, 2134–2139 (2000).
126. 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).
127. 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).
128. Paques, F., Leung, W. Y. & Haber, J. E. Expansions and contractions in a tandem repeat induced by double-strand break repair. *Mol. Cell. Biol.* **18**, 2045–2054 (1998).
129. Parniewski, P., Bacolla, A., Jaworski, A. & Wells, R. D. Nucleotide excision repair affects the stability of long transcribed (CTG*CAG) tracts in an orientation-dependent manner in *Escherichia coli*. *Nucleic Acids Res.* **27**, 616–623 (1999).
130. Oussatcheva, E. A., Hashem, V. I., Zou, Y., Sinden, R. R. & Potaman, V. N. Involvement of the nucleotide excision repair protein UvrA in instability of CAG*CTG repeat sequences in *Escherichia coli*. *J. Biol. Chem.* **276**, 30878–30884 (2001).
131. Bowater, R. P., Jaworski, A., Larson, J. E., Parniewski, P. & Wells, R. D. Transcription increases the deletion frequency of long CTG•CAG triplet repeats from plasmids in *Escherichia coli*. *Nucleic Acids Res.* **25**, 2861–2868 (1997).
132. Parniewski, P., Jaworski, A., Wells, R. D. & Bowater, R. P. Length of CTG•CAG repeats determines the influence of mismatch repair on genetic instability. *J. Mol. Biol.* **299**, 865–874 (2000).
133. Schumacher, S., Pinet, I. & Bichara, M. Modulation of transcription reveals a new mechanism of triplet repeat instability in *Escherichia coli*. *J. Mol. Biol.* **307**, 39–49 (2001).
134. Stojic, L., Brun, R. & Jiricny, J. Mismatch repair and DNA damage signalling. *DNA Repair (Amst.)* **3**, 1091–1101 (2004).
135. Lee, G. S., Brandt, V. L. & Roth, D. B. B cell development leads off with a base hit: dU:dG mismatches in class switching and hypermutation. *Mol. Cell* **16**, 505–508 (2004).
136. O'Hoy, K. L. *et al.* Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. *Science* **259**, 809–812 (1993).
137. van den Ouweland, A. M. *et al.* Loss of mutation at the *FMR1* locus through multiple exchanges between maternal X chromosomes. *Hum. Mol. Genet.* **3**, 1823–1827 (1994).
138. 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).
139. Jakupciak, J. P. & Wells, R. D. Genetic instabilities in (CTG•CAG) repeats occur by recombination. *J. Biol. Chem.* **274**, 23468–23479 (1999).
140. Cohen, H., Sears, D. D., Zenvirth, D., Hieter, P. & Simchen, G. Increased instability of human CTG repeat tracts on yeast artificial chromosomes during gametogenesis. *Mol. Cell. Biol.* **19**, 4153–4158 (1999).
141. 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).
142. Jankowski, C. & Nag, D. K. Most meiotic CAG repeat tract-length alterations in yeast are *SPO11* dependent. *Mol. Genet. Genomics* **267**, 64–70 (2002).
143. Brown, L. Y. & Brown, S. A. Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet.* **20**, 51–58 (2004).
144. Stead, J. D. & Jeffreys, A. J. Allele diversity and germline mutation at the insulin minisatellite. *Hum. Mol. Genet.* **9**, 713–723 (2000).
145. van der Maarel, S. M. *et al. De novo* facioscapulohumeral muscular dystrophy: frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. *Am. J. Hum. Genet.* **66**, 26–35 (2000).
146. Cancel, G. *et al.* Somatic mosaicism of the CAG repeat expansion in spinocerebellar ataxia type 3/Machado–Joseph disease. *Hum. Mutat.* **11**, 23–27 (1998).
147. Telenius, H. *et al.* Molecular analysis of juvenile Huntington disease: the major influence on $(CAG)_n$ repeat length is the sex of the affected parent. *Hum. Mol. Genet.* **2**, 1535–1540 (1993).
148. Onodera, O. *et al.* Progressive atrophy of cerebellum and brainstem as a function of age and the size of the expanded CAG repeats in the *MJD1* gene in Machado–Joseph disease. *Ann. Neurol.* **43**, 288–296 (1998).
149. Li, J. Y., Plomann, M. & Brundin, P. Huntington's disease: a synaptopathy? *Trends Mol. Med.* **9**, 414–420 (2003).
150. Mhatre, A. N. *et al.* Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nature Genet.* **5**, 184–188 (1993).
151. Seo, H., Sonntag, K. C. & Isacson, O. Generalized brain and skin proteasome inhibition in Huntington's disease. *Ann. Neurol.* **56**, 319–328 (2004).
152. Sathasivam, K. *et al.* Centrosome disorganization in fibroblast cultures derived from R6/2 Huntington's disease (HD) transgenic mice and HD patients. *Hum. Mol. Genet.* **10**, 2425–2435 (2001).
153. Lin, X., Antalffy, B., Kang, D., Orr, H. T. & Zoghbi, H. Y. Polyglutamine expansion down-regulates specific neuronal genes before pathologic changes in SCA1. *Nature Neurosci.* **3**, 157–163 (2000).
154. Caramins, M., Halliday, G., McCusker, E. & Trent, R. J. Genetically confirmed clinical Huntington's disease with no observable cell loss. *J. Neurol. Neurosurg. Psychiatry* **74**, 968–970 (2003).
155. Adachi, H. *et al.* Transgenic mice with an expanded CAG repeat controlled by the human AR promoter show polyglutamine nuclear inclusions and neuronal dysfunction without neuronal cell death. *Hum. Mol. Genet.* **10**, 1039–1048 (2001).
156. Xuereb, J. H., MacMillan, J. C., Snell, R., Davies, P. & Harper, P. S. Neuropathological diagnosis and CAG repeat expansion in Huntington's disease. *J. Neurol. Neurosurg. Psychiatry* **60**, 78–81 (1996).
157. Mizuno, H. *et al.* An autopsy case with clinically and molecular genetically diagnosed Huntington's disease with only minimal non-specific neuropathological findings. *Clin. Neuropathol.* **19**, 94–103 (2000).
158. Chamberlain, N. L., Driver, E. D. & Miesfeld, R. L. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res.* **22**, 3181–3186 (1994).
159. Beilin, J., Ball, E. M., Favaloro, J. M. & Zajac, J. D. Effect of the androgen receptor CAG repeat polymorphism on transcriptional activity: specificity in prostate and non-prostate cell lines. *J. Mol. Endocrinol.* **25**, 85–96 (2000).
160. Linja, M. J. & Visakorpi, T. Alterations of androgen receptor in prostate cancer. *J. Steroid Biochem. Mol. Biol.* **92**, 255–264 (2004).
161. Tsujimoto, Y. *et al. In situ* shortening of CAG repeat length within the androgen receptor gene in prostatic cancer and its possible precursors. *Prostate* **58**, 283–290 (2004).
162. Strachan, T. & Read, A. P. *Human Molecular Genetics 2* 2nd edn (John Wiley & Sons, New York, 1999).
163. Nichol Edamura, K. & Pearson, C. E. DNA methylation and replication: implications for the 'deletion hotspot' region of FMR1. *Hum. Genet.* (in the press).

Acknowledgements

The authors wish to thank the many scientists, too numerous to list individually, for discussions and comments on this manuscript. In addition, we wish to express our gratitude to the members of the Pearson laboratory (K. Hagerman, R. Lau, M. Leonard and G. Panigrahi) for their input into the manuscript. We apologize to those authors whose work we could not cite owing to space limitations. The Pearson laboratory is supported by funding from the Muscular Dystrophy Association, USA, the Canadian Institutes of Health Research (CIHR), the Fragile X Research Foundation of Canada and the University of Toronto Deans fund. J.D.C. is supported by a CIHR doctoral research award. C.E.P. is a CIHR Scholar and Canadian Genetic Disease Network Scholar.

Competing interests statement

The authors declare no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
FRAXE | FRAXF | FRA10B | FRA11A | FRA16B

OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
dentatorubral-pallidoluysian atrophy | dystrophia myotonica 1 | dystrophia myotonica 2 | facioscapulohumeral muscular dystrophy 1A | fragile X syndrome | Friedreich ataxia | Huntington disease | insulin | Machado–Joseph disease | myoclonic epilepsy of Unverricht and Lundborg | spinal and bulbar muscular atrophy | spinocerebellar ataxia 8 | spinocerebellar ataxia 10

FURTHER INFORMATION

Huntington's Disease Society of America: www.hdsa.org
Muscular Dystrophy Association: www.mdausa.org
Neuromuscular Disease Centre DNA Repeat Sequence and Disease page: www.neuro.wustl.edu/neuromuscular/mother/dnarep.htm
Tandemly Repeated (Satellite) DNA: http://arbl.cvmbs.colostate.edu/hbooks/genetics/medgen/dnatesting/satellites.html
The National Fragile X Foundation: www.fragilex.org
The Pearson Laboratory home page: www.cepearsonlab.com
Trinucleotide Repeats at OMIM: www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=omim&term=trinucleotide+repeat

SUPPLEMENTARY INFORMATION

See online article: S1 (table) | S2 (table)
Access to this links box is available online.