Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
Crit Rev Biochem Mol Biol. Author manuscript; available in PMC 2016 March 01.

Published in final edited form as:
Crit Rev Biochem Mol Biol. 2015 ; 50(2): 142–167. doi:10.3109/10409238.2014.999192.

Repeat instability during DNA repair: Insights from model systems

Karen Usdin[1], **Nealia C. M. House**[2], and **Catherine H. Freudenreich**[2,3]
[1]Laboratory of Cell and Molecular Biology, NIDDK, NIH, Bethesda, MD, USA
[2]Department of Biology, Tufts University, Medford, MA, USA
[3]Program in Genetics, Tufts University, Boston, MA, USA

Abstract

The expansion of repeated sequences is the cause of over 30 inherited genetic diseases, including Huntington disease, myotonic dystrophy (types 1 and 2), fragile X syndrome, many spinocerebellar ataxias, and some cases of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Repeat expansions are dynamic, and disease inheritance and progression are influenced by the size and the rate of expansion. Thus, an understanding of the various cellular mechanisms that cooperate to control or promote repeat expansions is of interest to human health. In addition, the study of repeat expansion and contraction mechanisms has provided insight into how repair pathways operate in the context of structure-forming DNA, as well as insights into non-canonical roles for repair proteins. Here we review the mechanisms of repeat instability, with a special emphasis on the knowledge gained from the various model systems that have been developed to study this topic. We cover the repair pathways and proteins that operate to maintain genome stability, or in some cases cause instability, and the cross-talk and interactions between them.

Keywords

Chromosome fragility; DNA damage checkpoint; DNA structure; recombination; repair; replication; structure-specific helicases; trinucleotide repeat expansion

---

Repeat expansions cause human disease

The expansion of trinucleotide repeats (TNRs) was first recognized as a cause of human disease in 1991 when the expansion of a CGG/CCG repeat in the 5′ UTR of the Fragile X mental retardation 1 (*FMR1*) gene was shown to be the cause of Fragile X syndrome (FXS) (Verkerk *et al*., 1991; Yu *et al*., 1991) and the expansion of a CAG/CTG repeat in the coding region of the androgen receptor gene was shown to be the cause of spinal-bulbar muscular atrophy (SMBA) (La Spada *et al*., 1991). Both these diseases exhibited a

---

Address for correspondence: Catherine H. Freudenreich, Department of Biology, Tufts University, Medford, MA 02155, USA. Tel: +1 617 627 4037. catherine.freudenreich@tufts.edu.

Declaration of interest
The authors report no declarations of interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

phenomenon described as anticipation, a worsening of the disease in each subsequent generation. Identification of several other TNR expansion diseases followed shortly, including Huntington disease (HD) caused by the expansion of a CAG repeat within the coding region of the Huntingtin gene, several spinocerebellar ataxias (SCAs) caused by CAG repeats in coding regions, myotonic dystrophy type I (DM1), caused by a CTG expansion in the 3′ UTR of the dystrophia myotonica protein kinase (*DMPK*) gene, and Friedreich ataxia (FRDA), caused by a GAA expansion in the first intron of the frataxin (*FXN*) gene (Table 1; note: the repeat on the coding strand is used to refer to the disease-causing repeat for a particular locus) (reviewed in Gatchel & Zoghbi, 2005; Lopez *et al*., 2010; McMurray, 2010; Mirkin, 2006; Pearson *et al*., 2005). In each case, a more severe disease phenotype and/or an earlier age of onset is caused by inheritance of a longer repeat allele in the offspring compared with the parent, confirming the more deleterious nature of longer repeat tracts and explaining the molecular basis of anticipation. Since the initial discoveries, many more diseases have been identified that are caused by repeat expansions, most of which are trinucleotide repeats, although several diseases are caused by repeats of larger units. Examples include SCA10 caused by an ATTCT expansion (Matsuura *et al*., 2000) and the recently identified G_4C_2 repeat expansion in the *c9orf72* gene, which is the most common cause of familial amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (DeJesus-Hernandez *et al*., 2011; Renton *et al*., 2011) (Table 1). In addition, it has recently been recognized that different sized expansions can in some cases result in different disease phenotypes. In particular, CGG expansions in the *FMR1* gene result in a group of disorders known collectively as the Fragile X-related disorders (FXDs). Expansions that result in alleles with 55–200 repeats (the so-called premutation or PM alleles) confer elevated risk of Fragile X-associated primary ovarian insufficiency (FXPOI) and Fragile X-associated tremor/ataxia syndrome (FXTAS), while expansions that generate alleles with >200 repeats are associated with FXS (Loesch & Hagerman, 2012; Table 1).

This review will focus primarily on mechanisms that cause repeat expansion (gain of units) and contraction (loss of units) in the CAG/CTG, CGG/CCG, and GAA/TTC trinucleotide repeats, which cause the majority of diseases and have been most intensively studied in model systems. However, the principles are likely to apply to other repeat expansion diseases.

Disease-associated repeats form non-B DNA structures

The expandable repeats (both trinucleotide and larger units) are able to form stable non-B form DNA structures that can interfere with normal cellular processes. For the CAG/CTG repeat, both the CAG and the CTG strands can form hairpin structures, where intrastrand rather than interstrand base-pairing occurs, with the CTG strand forming a more stable hairpin than the CAG strand (Figure 1A) (Gacy *et al*., 1995; Mitas *et al*., 1995b). Evidence for *in vivo* cleavage by synthetic zinc-finger nucleases specific for these hairpins demonstrates that these structures form *in vivo* in human cells (Liu *et al*., 2010). Similarly, both strands of the CGG/CCG TNR can form stable hairpin structures *in vitro* with the CGG strand forming the more stable structure (Mariappan *et al*., 1998; Mitas *et al*., 1995a; Nadel *et al*., 1995; Usdin & Woodford, 1995; Yu *et al*., 1997). Melting of both strands in a repeat tract followed by reannealing can lead to hairpins on both strands, termed a slipped strand

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

structure (Figure 1B) (Pearson *et al*., 1998). The CGG strand also has the potential to form G-quadruplex structures (Figure 1D) and the CCG strand to form i-tetraplexes (Fojtik & Vorlickova, 2001; Fry & Loeb, 1994; Kettani *et al*., 1995; Usdin & Woodford, 1995). Recent results indicate that it is the CGG hairpin structure that interferes with DNA replication in yeast cells (Anand *et al*., 2012). While quadruplexes are formed by many sequences *in vitro* (Biffi *et al*., 2013; Lam *et al*., 2013), whether G-quadruplex or i-tetraplex formation is relevant for CGG expansion *in vivo* is currently unknown; if they do form, it is likely that many of their biological properties would overlap with hairpins. Of note, the repeats responsible for progressive myoclonus epilepsy and the ALS/FTD-associated G_4C_2 expansion also form quadruplexes (Grigg *et al*., 2014; Potaman *et al*., 2004; Saha & Usdin, 2001). While hairpins have also been reported to be formed by very long GAA repeats (Heidenfelder *et al*., 2003), the major structures formed by GAA/TTC repeats are triplexes in which the third strand is either derived from the pyrimidine strand to form a pyrimidine:purine:pyrimidine triplex or derived from the purine strand to form a purine:purine:pyrimidine triplex (Figure 1E) (Grabczyk & Usdin, 2000; Potaman *et al*., 2004; Sakamoto *et al*., 2001). The ATTCT repeat expansion is not predicted to form structures, but rather acts as a DNA-unwinding element (Figure 1F) (Potaman *et al*., 2003). Interruptions to the purity of the repeat tract reduces the risk of expansion for many of these diseases (Chung *et al*., 1993; Eichler *et al*., 1994; Gao *et al*., 2008; Imbert *et al*., 1996; Liquori *et al*., 2003; Nolin *et al*., 2013; Yrigollen *et al*., 2012), likely because the interruption reduces the stability or length of the secondary structures that can be formed.

In addition to unusual DNA:DNA pairing, the G-rich repeats are especially prone to forming R-loops (Figure 1C). R-loops are RNA:DNA hybrids that form during transcription. While all sequences form transient R-loops behind RNA polymerase II (RNAPII), those that are formed in regions with a strand asymmetry with respect to purines and pyrimidines are particularly prone to form stable and persistent hybrids (Belotserkovskii *et al*., 2013). R-loops preferentially form when G-rich sequences are on the non-template strand (Belotserkovskii *et al*., 2010). Persistent RNA–DNA hybrids have been demonstrated at a CAG/CTG repeat in human cells, where their levels were correlated with instability (Lin *et al*., 2010a,b). CGG/CCG, GAA/TTC, and GGGGCC/GGCCCC repeats also form stable RNA:DNA hybrids (Grabczyk *et al*., 2007; Groh *et al*., 2014; Loomis *et al*., 2014; Reddy *et al*., 2014), and bidirectional transcription can further increase R-loop formation (Reddy *et al*., 2014, 2011). These hybrids could favor formation of intrastrand-folded structures by the non-template strand that remains unpaired. It is also possible that the ssDNA on the non-template strand is vulnerable to nicking. Both these events could facilitate repeat instability. The common ability to form these secondary structures of various flavors and forms are thought to be at the basis of repeat instability in the genome.

Developmental timing of TNR expansions

At expandable TNR loci in the human genome, there is a normal-size (non-disease-causing) repeat tract of variable length in the population, generally 5–30 repeat units. There is a fairly sharp size threshold where the repeat tract becomes much more unstable and likely to expand inter-generationally: around 35 uninterrupted repeat units for most loci (Table 1). The existence of this threshold indicates that an inherent property of DNA is likely involved

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in the switch to an unstable state. This could be a threshold for forming stable DNA structures *in vivo*. Coincidentally, this threshold corresponds to the size of an Okazaki fragment as well as the length of DNA that is wrapped around one nucleosome. Nonetheless, in systems where rare expansions can be measured, CAG/CTG repeats as small as 13 units are more likely to expand than a non-structure-forming control (Rolfsmeier *et al*., 2001). Additionally, *in vitro* studies suggest that even a small number of repeats produce expansions (see, for example, Henricksen *et al*., 2000; Liu *et al*., 2009; Pluciennik *et al*., 2013). In yeast, measurement of expansion frequencies shows that there is a non-linear increase with repeat size: for a CTG repeat on the lagging daughter strand (Okazaki fragment), expansions occur in about 1 in 10^5 cells for $(CTG)_{25}$ (Miret *et al*., 1998) but in about 1 in 10^2 cells for $(CTG)_{70}$ (Callahan *et al*., 2003), a 1000× increase. In humans, CAG expansion frequencies also increase dramatically once above the threshold length of ~35 repeats (reviewed in McMurray, 2010). In the case of Fragile X syndrome, the CGG expansion frequency approaches 100% when alleles with 85 repeats are maternally transmitted (Nolin *et al*., 2011; Yrigollen *et al*., 2014). The increased risk of expansion with longer sizes suggests that therapies that slow expansion rate or increase contractions could have an impact on disease inheritance or progression.

Intergenerational expansions of TNRs, which are inherited in all tissues, could occur at several developmental stages: in the dividing germ cell precursors, during germ cell maturation, during meiosis, during post-meiotic sperm maturation, or even during very early embryonic development. For HD, which has a paternal bias for inheritance of expanded alleles, an illuminating study of repeat size in human male spermatocytes concluded that the majority of expansions occur as a pre-meiotic event, although some additional expansions were observed to occur post-meiotically (Yoon *et al*., 2003). Yet in a mouse model of HD, it was concluded that most expansions arise after meiosis in the haploid gamete (Kovtun & McMurray, 2001), and recent results indicate that these post-meiotic expansions arise during remodeling of spermatid chromatin (Simard *et al*., 2014). A factor that could explain the difference between the mouse and human data is that the human spermatogonial cells studied were estimated to have undergone many more cell divisions (~600–800 over 40 years) compared with those in the mice (~35 over <3 months). Thus the post-meiotic expansion phase may be more prominent in the mouse model (Yoon *et al*., 2003). A study of male germ line instability in a DM1 mouse model with >300 CTG repeats also showed expansions occurring in pre-meiotic spermatogonia (Savouret *et al*., 2004). Together, these studies show that paternally inherited expansions can occur in dividing spermatogonia, and thus potentially could be coupled with replication. Large intergenerational expansions such as those that cause FXS and a congenital form of DM1 are seen almost exclusively when the allele is maternally transmitted (De Temmerman *et al*., 2004; Oberle *et al*., 1991; Yu *et al*., 1991). Expansions can be seen in oocytes from human DM1 patients (De Temmerman *et al*., 2004; Dean *et al*., 2006). In the case of both FXS and DM1, an increase in expansions is seen with maternal age, consistent with the idea that expansions occur in the post-natal oocyte in humans (Morales *et al*., 2014; Yrigollen *et al*., 2014). Furthermore, in a mouse model of the FXDs, the likelihood of expansion in the progeny of mice heterozygous for mutations that eliminate expansions depends on the parental genotype rather than the offspring genotype, consistent with expansion occurring prezygotically (Lokanga *et al*.,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2014b). Since oocytes are arrested in prophase I of meiosis until ovulation, these data would be compatible with an expansion mechanism that occurs independently of chromosomal replication.

In addition to intergenerational expansions, significant instability in somatic tissues can be seen in some TNR diseases, which can contribute to disease progression (for reviews see Dion, 2014; Lopez Castel *et al*., 2010; Pearson *et al*., 2005)). For example, in human HD patients, dramatic expansions (gains of up to 1000 repeats) are observed in striatal cells, the brain region most affected by the disease (Kennedy *et al*., 2003). Many mouse models also show extensive repeat expansion in the brain (for example, Libby *et al*., 2003; Lokanga *et al*., 2013; Mangiarini *et al*., 1997) and specifically in post-mitotic neurons (Gonitel *et al*., 2008). Since even expansion-prone somatic cells that are not post-mitotic proliferate very slowly, non-replicative expansion mechanisms are likely to explain many somatic expansions. Supporting this idea, inhibition of cell proliferation using a number of different approaches does not reduce expansions in cultured transgenic DM1 mouse cells (Gomes-Pereira *et al*., 2014). However, there may be replication-coupled somatic expansions in some tissues. A study of CTG repeat sizes in DM1 patient tissues, where the largest expansions are observed in muscle, pointed toward expansions occurring in a window during early development when muscle cells are proliferating prior to terminal differentiation (Zatz *et al*., 1995). Even in differentiated tissues, there are replicating cell populations that could provide a source of continued expansions, such as stem cells, where some triplet repeats are highly expandable (Du *et al*., 2013; Ku *et al*., 2010). In the brain, dividing glial cells show evidence of ongoing CAG/CTG instability *in situ* and in culture (Claassen & Lahue, 2007; Watanabe *et al*., 2000). Because there is a relationship between repeat length and disease severity, understanding the mechanisms of both intergenerational and somatic instability in replicating and non-replicating cells is relevant for understanding disease progression, and may provide an entry point for therapies.

Cis and *trans*-acting factors that modulate repeat instability

Even with the same repeat length, some cell types are more expansion prone than others. For example, pluripotent stem cells from individuals with FRDA and DM1 show expansion while the fibroblasts from which they are derived do not (Du *et al*., 2012, 2013). In mouse models of the TNR diseases, some tissues are more expansion prone than others and these differ between different disease models (Clark *et al*., 2007; Fortune *et al*., 2000; Goula *et al*., 2009; Kennedy *et al*., 2003; Lokanga *et al*., 2013), suggesting that a combination of cell-type-specific factors and locus-specific factors must play a role in the determination of expansion frequency.

A role for *cis*-acting factors in expansion is suggested by the fact that in many of the repeat expansion diseases some haplotypes are more likely to expand than others (Ennis *et al*., 2007; Martins *et al*., 2008; Murray *et al*., 2000; Richards *et al*., 1992; Takiyama *et al*., 1995; Warby *et al*., 2009). In addition, a comparison of different CAG/CTG triplet repeat loci indicated that the flanking CG content and proximity to CpG islands modifies the repeat expandability (Brock *et al*., 1999; Nestor & Monckton, 2011). Another *cis*-acting factor-influencing repeat instability is orientation with respect to replication direction; in bacterial

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

cells, yeast cells, and HeLa cell extract contractions were more frequent with the CTG repeat was on the lagging template strand, while CTG on the nascent lagging strand (Okazaki fragment) favored expansions (Cleary *et al*., 2002; Freudenreich *et al*., 1997; Kang *et al*., 1995; Maurer *et al*., 1996; Miret *et al*., 1998; Panigrahi *et al*., 2002). These data fit with the known areas of single-stranded character at the replication fork and the greater stability of the CTG hairpin: on the exposed lagging strand template, bypass of a CTG hairpin would lead to contraction, whereas a CTG hairpin on the 5′ flap of the nascent lagging strand would lead to an expansion (see Figure 2 and section ''General mechanisms of TNR expansion and contraction''). It also predicts that a switch in origin location with respect to the repeat could change the balance between expansions and contractions, the ''ori switch model'' (Cleary & Pearson, 2005; Mirkin & Smirnova, 2002). Indeed, recent data on the replication patterns at the *DM1* and *FMR1* loci in human cells shows evidence for changes in origin usage at unstable alleles (Cleary *et al*., 2010; Gerhardt *et al*., 2014a; Mirkin & Mirkin, 2014). At the *FMR1* locus, a common SNP identified in one of the flanking origins has been shown to be associated with an expansion-prone haplotype (Gerhardt *et al*., 2014b).

Another *cis*-acting factor shown to influence repeat instability is transcription. Interestingly, all expandable repeats (with the possible exception of the EPM1-causing dodecamer) are located in the transcribed part of their associated genes. A role for transcription in repeat expansion of CAG/CTG repeats is suggested by the comparison of two different transgenic mouse lines that exhibited a higher level of instability in the line showing the higher level of transcription (Goula *et al*., 2012b). Using controllable systems, it has been observed that induction of transcription in *E. coli*, Drosophila, and human tissue culture, all increase CAG repeat instability (Bowater *et al*., 1997; Jung & Bonini, 2007; Lin *et al*., 2006). In humans, silenced FX alleles are stable (Wohrle *et al*., 2001), and in mice with ~135 repeats in the endogenous FMR1 gene on the X chromosome, expansion is only seen in females when this allele is on the active X chromosome (Lokanga *et al*., 2014a). In a further twist, some repeat loci, such as the those that cause SCA8, FXS, SCA7, DM1, and ALS-FTD, are transcribed bi-directionally in humans (Ladd *et al*., 2007; Moseley *et al*., 2006; Nakamori *et al*., 2011; Sopher *et al*., 2011; Zu *et al*., 2013), and induction of bidirectional transcription in a human cell model further increased instability over unidirectional transcription (Lin & Wilson, 2012; Lin *et al*., 2010a,b). Transcription may be important because it creates an opportunity for the DNA hairpins, which are thought to be the substrates for expansion, to form. During passage through the repeat, the transcription complex would temporarily occlude the template strand, thus leaving the non-template strand transiently unpaired and able to form a structure, providing the trigger for expansion.

On the other hand, some animal models do not show a good correlation between the *levels* of transcription and the extent of expansion. Specifically, there is no correlation between the level of transcription of the disease locus and the extent of expansion in different tissues in either the DM1 or the FXS mouse model (Entezam *et al*., 2010; Lia *et al*., 1998) or between the amount of expansion seen in the fly model of SCA3 and the level of transcription induction (Jung & Bonini, 2007). In HD patients and an HD mouse model, expansion is more extensive in the striatum than the cerebellum despite the similar levels of HTT

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

transcription (Dure *et al*., 1994; Goula *et al*., 2012b). In the case of the HD mouse, it has been suggested that tissue-dependent regulation of promoter proximal pausing at the *HTT* gene might underlie tissue-selective instability (Goula *et al*., 2013). A recent study in yeast suggests another model for the effect of transcription. Shah *et al*. (2014) showed that a GAA repeat located in a non-transcribed region upstream of an inducible promoter exhibited an increased rate of expansions when transcription was induced nearby. Further investigation revealed that induction of transcription resulted in lower nucleosome occupancy in the upstream region containing the GAA tract. The authors conclude that transcriptional state, rather than transcription elongation through the repeat, was responsible for the 10-fold increase in expansion rate that occurred under induced conditions (Shah *et al*., 2014). They speculate that the lower nucleosome occupancy occurring due to transcriptional activation allows more template switching during replication.

A less well understood *cis*-acting factor shown to influence repeat instability is the local chromatin structure. Alterations in the level of packing of the chromatin fiber, histone modifications, DNA methylation, chromatin binding proteins, and chromatin looping could all affect repeat instability via the ability either to influence transcription levels or facilitate the formation of secondary structures, or promote accessibility to DNA repair/recombination processes. See Dion & Wilson (2009) for a comprehensive review of this topic. For example, SCA7 mouse models only show instability of a $(CAG)_{92}$ tract if the ataxin-7 gene is flanked by human genomic DNA containing functional-binding sites for CTCF, a protein implicated in large-scale chromatin structure that has binding sites next to several expansion prone repeat loci (Libby *et al*., 2008). Many expandable repeats are near CpG islands, and several studies correlate changes in DNA methylation with repeat stability. For example, inhibition of DNA methylation increased expansions at the CTG tract in the *DMPK* locus in DM1 patient cells (Gorbunova *et al*., 2004), and deficiency of DNA methyltransferase Dnmt1 increased germ-line but not somatic expansions in a SCA1 mouse model (Dion *et al*., 2008). The effect of the Dnmt1 deficiency was not mediated by an increase in transcription through the repeat or by many genes thought to be important for expansion; therefore, a different aspect of the chromatin structure influenced by methylation may contribute to expansion propensity. Histone modifications may also affect repeat instability through changes in chromatin structure. In flies, lower histone acetylation levels (due to haploinsufficency of the histone acetyltransferase CBP) increase CAG repeat instability, while a higher level of acetylation (due to treatment with trichostatin A (TSA), a histone deacetylase inhibitor) decreases instability (Jung & Bonini, 2007). Both CpG methylation and histone acetylation influence nucleosome assembly on CGG repeats (Mulvihill *et al*., 2005), perhaps linking these two modifications. Acetylation of histone H3 at lysine 56 is important for nucleosome deposition during DNA replication and repair, and defects in H3-K56 acetylation were found to increase CAG contractions and fragility in yeast (Yang & Freudenreich, 2010). H3-K56 acetylation was hypothesized to facilitate proper nucleosome assembly at the replication fork to prevent DNA structure formation and subsequent slippage events on the template strand or fork breakage. An example of an epigenetic mark affecting repair within a repeat is provided by our recent study of factors that prevent CAG expansions in yeast, that revealed that acetylation of the histone H4 N-terminal tail by the NuA4 and Hat1 enzymes, and chromatin remodeling by RSC2, were important for the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

fidelity of repair within the repetitive DNA tract (House *et al*., 2014b) (see section ''Post-replication gap repair and fork restart'' of this review). In this case, the modification was shown to occur directly at the expanded repeat coincident with damage, but not at a no-repeat control in the same location. However, the effects of some epigenetic modifiers may be indirect. For example, knockdown of some histone deacetylases that suppress CAG expansions in yeast and human cells may do so by altering the stability of the Sae2 nuclease (human CtIP) (Debacker *et al*., 2012).

There is also evidence for *trans*-acting factors influencing expansion rate, and thereby affecting the risk of disease or age of onset. For example, the risk of expansion of some FMR1 alleles is lower in the general population than it is in families with a family history of FXS (Nolin *et al*., 2013). In myotonic dystrophy patients, the level of somatic instability is highly heritable, implying a role for individual-specific *trans*-acting genetic modifiers (Morales *et al*., 2012; Neville *et al*., 1994). Potential *trans*-acting factors include the absolute and relative levels of different proteins involved in the expansions as well as the levels of proteins involved in protecting the genome against these events. Recently, transcript and protein level analysis has revealed that there are significant expression differences in repair protein levels in tissues with different expansion propensities, including Msh2, Msh3, Msh6, PCNA, Rpa1, Fen1, Lig1, HMGB1, and Polβ (Du *et al*., 2012; Goula *et al*., 2009; Mason *et al*., 2014; Seriola *et al*., 2011; Tome *et al*., 2013), all proteins shown to influence TNR instability (see sections ''The role of mismatch repair proteins in repeat instability'', ''The role of base excision repair proteins in repeat instability'', ''The role of nucleotide excision repair proteins in repeat instability'', ''Homology-dependent recombinational repair of gaps and stalled forks'', and ''The role of DSB repair in preventing repeat fragility and instability''). Also, different alleles or expression levels of mismatch repair proteins have been identified in mouse strains that exhibit different expansion frequencies (Pinto *et al*., 2013; Tome *et al*., 2013).

General mechanisms of TNR expansion and contraction

Addition of repeat units by definition involves some sort of DNA synthesis. Early models of repeat expansion envisioned that the incorporation of additional bases might arise as a result of strand slippage during replication (Richards & Sutherland, 1994) (Figure 2A). It is now understood that strand slippage can occur in many different contexts, including DNA replication in dividing cells, as initially envisioned, or during various types of repair synthesis. In addition, three other mechanisms can create expansions: failure to remove displaced 5′ DNA flaps and their subsequent incorporation into DNA (Figure 2B), misalignment during recombinational repair (Figure 2C), and synthesis during double-strand break (DSB) repair (Figure 2D). These events can occur both during the duplication of genomes and during repair processes. In addition to expansions, these unstable elements can also lose repeat units, a process termed contraction. Contractions can occur by three basic mechanisms: replication over a template hairpin either during replication or during repair synthesis (Figure 2A), misalignment during recombinational repair (Figure 2C), or resection during double-strand break repair (Figure 2D). Again, each of these events can occur during several types of replication or repair processes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Although the basic idea of how repeat units can be gained and lost is fairly straightforward, a more detailed understanding of how instability occurs in different repair pathways, and the identification of which mechanisms are operating in different cell types and developmental windows, has been surprisingly difficult to determine. The use of model organisms has been invaluable in this regard, as proteins can be manipulated to perturb particular pathways and measure any resulting repeat stabilization or destabilization. The picture that has emerged to date is complex and interesting, providing many potential targets for manipulating the process in human patients, with more certainly to be learned in the future. In addition, study of these structure-forming sequences has yielded valuable insights into basic mechanisms of replication and repair that have increased our understanding of these crucial cellular processes. The study of repeat instability and fragility mechanisms has been particularly valuable in providing insights into how normally genome-protective pathways can fail, leading to genomic changes, and ultimately to disease pathology.

DNA damage at structure-forming repeats

As illustrated in Figure 2, DNA structures can interfere with a number of DNA transactions. For example, GAA/TCC triplexes and GGC/CCG repeats strongly interfere with replication progression, acting as site-specific barriers (Anand *et al*., 2012; Krasilnikova & Mirkin, 2004; Voineagu *et al*., 2009b). CAG/CTG repeats are much weaker barriers, but replication bypass can still leave a gap that must be filled post-replicatively (Kerrest *et al*., 2009). Hairpins on 5′ flaps interfere with the ligation step of nick repair and Okazaki fragment processing, leaving an unligated nick (Liu & Bambara, 2003; Liu *et al*., 2004; Panigrahi *et al*., 2005; Spiro *et al*., 1999). Indeed, deletion of Fen1 (yeast Rad27), a key protein needed for 5′ flap processing, results in a large increase in CAG repeat fragility (Callahan *et al*., 2003). In addition to the difficulties in processing the secondary structures formed by the repeat, some of the disease-associated hairpins have been shown to be prone to oxidative damage (Jarem *et al*., 2009) and they may also be vulnerable to cleavage by single-stranded nucleases. Since even short TNRs can form DNA structures that interfere with DNA processes (see sections ‘‘Disease associated repeats form non-B DNA structures’’ and ‘‘Developmental timing of TNR expansions’’), cellular mechanisms that can facilitate repair of the nick, gap, or stalled fork are predicted to be important for tolerance of even normal-length TNRs in the genome, but particularly for expanded tracts, as well as other sites of structured DNA. Indeed, cellular checkpoint mechanisms are induced in cells with expanded TNRs, and checkpoint signaling proteins are important for TNR stability (see section ‘‘Role of the DNA damage checkpoint in sensing damage at TNRs and regulating repair’’).

Expanded repeats are fragile sites

Expanded TNRs are sites of increased chromosomal fragility, causing DNA breakage that can result in chromosome loss or rearrangement. A DSB at an expanded repeat can occur by multiple mechanisms: (1) processing of a stalled fork (e.g. by nucleases acting on the exposed ssDNA or a structured DNA, such as the four-way junction of a reversed fork or recombination intermediate); (2) conversion of an unligated nick to a DSB; (3) direct cleavage of the repeat, e.g. by structure-specific nuclease cleavage; (4) physical breakage or cleavage of an unreplicated region during cell division.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In humans, cells with expanded CGG repeats in the *FMR1* or *FMR2* genes exhibit a cytologically visible fragile site coincident with the repeat. The fragile site is seen as a gap, constriction, or break on metaphase spreads of chromosomes from cells grown with either very high or very low folate levels. This folate stress is thought to lead to a nucleotide pool imbalance that affects the rate of DNA replication through the repeat. Deletions and translocations at the site of CGG expansions have been documented, which provides indirect evidence that a double-strand break (DSB) occurred on the chromosome. Expanded repeats at the *FMR1* gene are associated with very late replication and problems with replication initiation and/or elongation (Gerhardt *et al*., 2014a; Hansen *et al*., 1993, 1997; Subramanian *et al*., 1996; Yudkin *et al*., 2014).

For other expanded TNRs, there is no cytological appearance of a fragile site in human cells (Jalal *et al*., 1993; Wenger *et al*., 1996); however, loss of the chromosome with an expanded CTG repeat at the DM1 locus is increased (Casella *et al*., 2003), suggesting that breakage is occurring *in vivo*. In addition, data from yeast provide strong evidence that these sequences cause increased frequencies of chromosome breakage. CAG/CTG tracts of 45 units or longer can induce fragility of a yeast chromosome as measured by sensitive genetic assays that detect large deletions or chromosome end loss (Callahan *et al*., 2003; Freudenreich *et al*., 1998). Additionally, breakage of longer CAG tracts (79–250 repeats) are visible as physical chromosome breaks in mitosis (Freudenreich *et al*., 1998) and meiosis (Jankowski *et al*., 2000). GAA/TCC tracts and other AT-rich elements predicted to form hairpins or cruciforms also cause chromosome fragility (Kim *et al*., 2008; Lobachev *et al*., 2002; Zhang *et al*., 2012b), as does the ATTCT repeat unwinding element (Cherng *et al*., 2011). Comparison of CAG and CGG fragility using the same yeast genetic assay reveals that CGG repeats are 3–10-fold more fragile than CAG repeats of equivalent size, offering one explanation of why CGG fragility is more evident in human cells (Balakumaran *et al*., 2000). Chromosome context or location may also play an important role in determining levels of fragility. For example, an impediment to replication generated by repeats that are located in genes that normally replicate late may be more likely to result in cells entering mitosis before replication of that region is complete. Expanded CGG repeats seem to escape the DNA damage checkpoint in yeast, whereas CAG repeats do not, providing another explanation for the apparent increased fragility of the CGG repeat (Voineagu *et al*., 2009a).

Protective and mutagenic roles of DNA repair at repetitive DNA

Chromosome breakage at an expanded TNR represents the failure of normal repair processes. However, most of the time that a nick, gap, or DSB occurs at a TNR tract the cell will succeed in repairing that damage without chromosome arm loss, deletion, or rearrangement. The repair may occur with fidelity, i.e. with no loss or gain of genetic material, or with a change in repeat number, but with no consequence to the rest of the genome. This outcome represents successful repair (no fragility), but with the possibility of a detrimental repeat expansion; e.g. the ‘‘protective’’ repair process has gone awry because of the challenges of replicating or aligning DNA across a repetitive region, resulting in repeat instability. Lastly, there is evidence that some ‘‘repair’’ pathways, such as mismatch repair (MMR) and transcription-coupled repair (TCR) may inappropriately recognize extruded TNR hairpins or triplex structures as DNA damage, and initiate an unnecessary or

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

''gratuitous'' repair event that is in no way protective and leads to repeat instability (Gomes-Pereira *et al*., 2004; Pluciennik *et al*., 2013; Salinas-Rios *et al*., 2011). In the remainder of the review, we summarize what is known about DNA repair pathways whose components have been implicated in repeat instability.

The role of mismatch repair proteins in repeat instability

Mismatch repair (MMR) is normally involved in the repair of base mismatches or insertions/deletions (IDLs) (see Modrich, 2006 for review). It is initiated by recognition of the mismatch by MSH2/MSH6 (MutSα), a heterodimer that binds to single base mismatches and small IDLs or MSH2/MSH3 (MutSβ), a complex that is involved in the repair of longer IDLs. A MutL complex is then recruited to form a ternary complex. Three different MutL complexes are found in mammals: MutLα a MLH1/PMS2 heterodimer and perhaps the most important of the MutL complexes for MMR outside of meiosis; MutLβ, a MLH1/PMS1 heterodimer; and MutLγ, a heterodimer of MLH1/MLH3. The MutS/MutL complex then interacts with the homotrimeric-proliferating cell nuclear antigen (PCNA). This process activates the latent endonuclease activity of the MutL complex, resulting in the generation of nicks nearby. MutL nuclease activity is strand directed if there is a pre-existing nick to signal the nascent strand, or it can cleave either strand if no such signal exists. The Exo1 exonuclease is then loaded at the nicks by the MutS/MutL complex and generates a single-stranded gap starting at the nick and extending ~150 nucleotides past the mismatch. Gap filling mediated by Polδ then occurs, with DNA ligase I (Lig1) sealing the nick.

MMR proteins are important for repeat expansion in a wide variety of model systems, including mammalian cells and mouse models. MMR proteins also play a role in CTG/CAG repeat expansions in yeast, although to a lesser extent than in some of the mouse models (Kantartzis *et al*., 2012; Schweitzer & Livingston, 1997). Importantly, hairpins or slip-stranded structures formed by CTG/CAG repeats bind MutSβ *in vitro* (Lang *et al*., 2011; Owen *et al*., 2005, 2009), and binding to these repeats (Gannon *et al*., 2012) and to GAA/TTC-repeats has been reported *in vivo* (Du *et al*., 2012; Ezzatizadeh *et al*., 2012).

As can be seen in Table 2, the preponderance of evidence supports a role for a MutS complex in repeat expansion. MutS complexes, and in particular MutSβ, are required for both intergenerational and somatic expansion in most mouse models of repeat expansion. Some differences are seen between specific models that may reflect the effect of sequence context, since *cis*-acting sequences affect expansion (Brock *et al*., 1999; Cleary *et al*., 2002; Libby *et al*., 2003), or genetic background, since mouse strains differ significantly in the levels of expression of key MMR proteins (Pinto *et al*., 2013; Tome *et al*., 2013). Even in the case of FRDA in which none of the MutS components seem to play a role in intergenerational expansion in the transgenic mouse model (Ezzatizadeh *et al*., 2012), loss of either of the MutSα components reduces somatic expansions (Bourn *et al*., 2012). Furthermore, knockdown of MutSα reduces expansions of the endogenous FRDA allele in patient derived induced pluripotent stem cells (iPSCs) (Du *et al*., 2012; Ku *et al*., 2010) and expression of MutSβ promotes GAA expansions in a human kidney cell line (Halabi *et al*., 2012). Thus far FRDA appears to be unique in the contribution of MutSα along with MutSβ

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

to expansion. This may reflect the fact the FRDA repeat forms a triplex while the other repeats examined do not.

The requirement for MMR proteins is not limited to the MutS complex. Mlh1, the protein common to all three MutL complexes found in mammals, and Mlh3, the Mlh1-binding partner in MutLγ, are required for all somatic expansions in an HD mouse model (Gomes-Pereira *et al*., 2004; Pinto *et al*., 2013). Since MutLγ is thought to interact with MutSβ not MutSα (Charbonneau *et al*., 2009), this supports a key role for MutSβ in expansion of CAG-repeats. However, in a DM1 mouse model, PMS2, the binding partner of Mlh1 in MutLα, is required for 50% of somatic expansions (Gomes-Pereira *et al*., 2004). Despite the lack of a role for MutSα or MutSβ in intergenerationally transmitted expansions in the FRDA mouse (Bourn *et al*., 2012; Ezzatizadeh *et al*., 2012), Mlh1 is required for both germ line and somatic expansions (Ezzatizadeh *et al*., 2014). Since PMS2 (MutLα), protects against expansions in this model (Bourn *et al*., 2012) this suggests that there is also a role for MutLγ in FRDA expansions.

A role of MutLγ in repeat expansion is intriguing since Mlh3 is present in mammalian cells at levels 60 times lower levels than PMS2 (Cannavo *et al*., 2005). While its role in meiotic recombination has been known for some time, the role of MutLγ in typical MMR is not well understood. MutLγ is able to carry out only low levels of MMR *in vitro*, and recent reports suggest that while MutLγ can localize to sites of DNA damage, it is unable to substitute for MutLα in MMR *in vivo* (Roesner *et al*., 2013). It is of interest to note that the MutLγ complex shows a marked preference for binding to Holliday junctions (Ranjha *et al*., 2014). It may be that the requirement of the MutLγ complex reflects a role in the recognition and processing of a similar DNA structure.

The requirement of a MutL complex suggests that the role of MMR proteins is not limited simply to lesion recognition. Since loading of the proliferating cell nuclear antigen (PCNA) clamp that is required for MutL activation during MMR requires a strand discontinuity, MMR was once thought to be confined to replication when leading and lagging strand termini are available for PCNA loading. However, it has recently been shown that both small CAG and CTG loopouts of 2–3 repeats support PCNA loading, and thus the MutLα activation that could allow MMR to occur even in post-mitotic cells (Pluciennik *et al*., 2013). Thus expansions might arise in such cells by a process like that outlined in Figure 3(A) in which PCNA loading would be facilitated by the presence of the hairpins. These hairpins could be preformed slipped strand structures as depicted in Figure 3(A), or formed by another process (e.g. gap filling, or transcription; see Figures 3B, 4, and 5). PCNA loading in turn results in activation of the latent endonuclease activity of MutLα and the generation of nicks that can be used by Exo1 to remove the region of DNA across from the template hairpin. Polδ would then fill in the gap. To avoid a contraction, the template hairpin would have to be unwound. We suggest the WRN DNA helicase (yeast Sgs1), which interacts with Polδ and Mlh1 and has been shown to prevent contractions (see section ‘‘The role of helicases in resolving hairpins during repair and fork restart’’), as a good candidate for acting at this step. In the absence of Exo1 digestion, repair could also occur via strand displacement synthesis by Polδ (Kadyrov *et al*., 2009) with or without strand slippage (dotted arrow to Figure 3B). After fill-in, Lig1 would then seal the nick generating a

heteroduplex with a single loopout. A repair reaction originating on the complementary strand would result in the resolution of the heteroduplex by fill-in, resulting in an expansion. Since MutLα nicking on a template without preexisting nicks occurs without strand bias (Pluciennik *et al*., 2013), nicking could occur on either strand. The location of the nicks would determine the order of events and whether expansion, contraction or error-free repair results. Although this pathway can occur in non-replicating cells, if it did occur in the context of replication, recognition of a template hairpin by MutSβ followed by MutL cleavage could create a DSB; expansions and contractions could then occur during fork restart or DSB repair (Pluciennik *et al*., 2013).

In addition to canonical MMR, a variety of non-canonical roles for MMR proteins are known, including their involvement in the repair of oxidative damage, homologous recombination (HR), double-strand break (DSB) repair, immunoglobulin class switching, and somatic hypermutation. It may thus be that the MutS and MutL complexes affect expansions via their participation in other repair pathways. Of interest in this regard is the observation that while MutSβ binding to a CAG-hairpin in bulk solution has been reported to be indistinguishable from binding to a mismatch that is efficiently repaired by MMR (Tian *et al*., 2009), single molecule FRET has shown that MutSβ binding results in a subset of molecules having a different conformation (Lang *et al*., 2011). This complex is proposed to trap MutSβ and inhibit its dissociation from the hairpin. This has led to the suggestion that hairpin processing by MMR would be inefficient and could result in the hairpin being repaired by an alternative DNA repair pathway (McMurray, 2008). On preformed loop/hairpin substrates *in vitro*, expansion can be observed even in the absence of MMR proteins (Panigrahi *et al*., 2005, 2010). This *in vitro* repair is also independent of proteins involved in base excision repair and nucleotide excision repair. Also, in a HeLa nuclear extract, excess MutSβ does not inhibit or promote CTG or CAG repair of preformed hairpin substrates (Tian *et al*., 2009; Zhang *et al*., 2012a). These results have led to the suggestion that MutSβ may act to promote the formation of the secondary structures rather than determining repair outcome.

The role of base excision repair (BER) proteins in repeat instability

BER is the major pathway by which oxidized bases are repaired in mammalian cells. This process is initiated either by the removal of the damaged base by a DNA glycosylase or by the spontaneous hydrolysis of the glycosidic bond connecting the damaged base to the sugar phosphate back-bone (see Liu & Wilson, 2012 for a review). The resultant abasic site is processed by the apurinic/apyrimidinic (AP) endonuclease 1 (APE1) or the AP lyase function of some glycosylases to generate a single nucleotide gap (Figure 3B). If the AP site or the 5′-sugar phosphate group is not oxidized, the gap is processed by the single-nucleotide BER (SN-BER) pathway. In SN-BER, a single nucleotide is inserted by DNA polymerase β (Polβ) and the nick is sealed by DNA ligase III/XRCC1 or DNA ligase 1 (Lig1). In the absence of a ligatable nick, a second BER pathway, the long patch (LP)-BER pathway, is used instead. The most efficient form of the LP-BER pathway has been suggested to involve the coordinated actions of Polβ and Fen1 acting via a ‘‘Hit and Run’’ mechanism that consists of stepwise and distributive gap-filling by Polβ and single-nucleotide gap formation by Fen1 to generate a ligatable nick that is then sealed by Lig1

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Liu *et al*., 2005). This process yields repair patches of two nucleotides or more without the generation of a long 5′ flap resulting from strand displacement synthesis. LP-BER can also occur by strand-displacement synthesis that involves Polβ or the replicative polymerases Polδ and/or Polε, and perhaps Polη or other TLS polymerases. In this case, Fen1 removes the resultant flap (Figure 3B).

A role for BER in repeat expansion is suggested by the fact that treatment of HD fibroblasts with H_2O_2 increases the expansion frequency (Kovtun *et al*., 2007), as does exposure of FX premutation mice to potassium bromate (Entezam *et al*., 2010). The fact that the loss of either OGG1, the major DNA glycosylase responsible for removing 7,8-dihydro-8-oxoguanine (8-oxoG), the most common DNA oxidation product, or the endonuclease VIII-like 1 (NEIL1) glycosylase, decreases somatic expansions in a transgenic HD mouse model (Kovtun *et al*., 2007; Mollersen *et al*., 2012) would be consistent with this idea.

Biochemical experiments on CAG/CTG-repeat substrates have led to a number of different models being proposed for expansion via the LP-BER pathway. One model is based on the observation that during repair of an 8-oxoG lesion in the vicinity of a CAG repeat tract, multinucleotide gap filling occurs that exhibits periodic pausing sites consistent with hairpin formation (Liu *et al*., 2009). According to this model, expansion occurs during Polβ-mediated LP-BER if the optimal coordination between Fen1 and Polβ is disrupted. Expansions would arise because strand displacement and hairpin formation is able to occur before ligation, and Fen1 cannot cleave at the 3′ base of the resultant hairpin (Figure 3B, right pathway). Polβ would then carry out multinucleotide gap-filling synthesis and Fen1 would be forced to use its alternate cleavage activity to process only a few nucleotides off of the 5′ flap to generate a substrate suitable for ligation (Liu *et al*., 2009). The net result would be that many of the flap bases would be incorporated into the nascent strand (Figure 3B, right pathway; Fen1 ''alternate'' cleavage of flap 5′ to hairpin). A second model invokes strand-slippage on the nascent strand in the repeat region during LP-BER in the presence of both Polβ and Polδ. This model is based on the observation that while in the absence of Polβ, Polδ is able to use its 3′–5′ proof-reading ability to excise hairpins that lack a 3′ tail, in the presence of Polβ excision of this hairpin does not take place. Instead Polβ adds a few nucleotides to the 3′ end of the hairpin thus generating an effective primer for Polδ-catalyzed DNA synthesis that occurs without hairpin excision. The net result is that the hairpin becomes fixed in the nascent strand and can result in a repeat expansion (Chan *et al*., 2013) (Figure 3B, left pathway). In either case, since the hairpins formed by CAG and CTG repeats are hot spots for oxidation, and since OGG1 has a reduced affinity for 8-oxoG in these hairpins and excises them at a significantly lower rate compared with duplexes (Jarem *et al*., 2011), hairpin formation could result in a ''toxic oxidation cycle'' in which the repair of one lesion would increase the opportunity for the generation of additional oxidized bases.

A key step in the LP-BER pathway is flap cleavage by FEN1. The formation of secondary structures inhibits FEN1 cleavage of 5′ flaps *in vitro* (Spiro *et al*., 1999) and FEN1 has been shown to be important in creating expansions in an *in vitro* system that models BER (Liu *et al*., 2009). Fen1 heterozygosity resulted in fewer intergenerational contractions in an HD mouse model (Spiro & McMurray, 2003). However, Fen1 heterozygosity did not affect the frequency of expansions or contractions in either a FX or a DM1 mouse model (Entezam *et*

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

al., 2010; van den Broek *et al*., 2006), or were expansions affected in DM1 *Fen1*$^{-/-}$ blastocysts that stained negatively for Fen1 protein (van den Broek *et al*., 2006). Since *Fen1*$^{-/-}$ blastocysts do not survive, and the level of Fen1 in heterozygotes may still be high enough to sustain the normal level of expansion, the question of a role for Fen1 in germ line expansions in mouse models remains open. In a yeast study, cells heterozygous for Fen1/Rad27 maintained $(CAG)_{70\text{-}130}$ repeat lengths; expansions increased only when the tract reached 155 repeats, and the frequency was still less than in *fen1* cells (Yang & Freudenreich, 2007). In a mouse model of DM1, a hypomorphic mutation of DNA ligase 1 (46BR), reduced intergenerational repeat expansions and increased contractions but only when the expansion prone allele was maternally transmitted (Lopez Castel *et al*., 2011). However, the same mutation had no effect on intergenerational and somatic expansions in the FX mouse model (Entezam *et al*., 2010). It may be that a delicate balance exists between Fen1 and Lig1 (and perhaps other factors) so that an effect on expansions is not always seen in heterozygous animals. In the HD mouse model, the stoichiometry of proteins involved in BER has been suggested to explain why the striatum is more prone to expansion than the cerebellum. Specifically, a low ratio of proteins like FEN1, APE1, and LIG1 relative to Polβ would favor strand displacement that would lead to expansions (Goula *et al*., 2009, 2012a). Whether this correlation holds up when a larger range of tissues are tested remains to be seen.

Expansions in the HD mouse model can still occur in the absence of OGG1 or NEIL1. However, this does not necessarily invalidate a BER model. It could be that other glycosylases are able to compensate for the loss of these enzymes or that other sorts of DNA damage also contribute to expansion. Alternatively, Polη has recently been shown to be recruited by the MutS complex to participate in BER of tandem lesions that are not recognized by DNA glycosylases (Zlatanou *et al*., 2011), but whether MutSβ is also able to recruit Polη in response to hairpin lesions has not, to our knowledge, been tested. One advantage of a mechanism such as this is that it could account for the requirement of MMR proteins for expansion *in vivo*, using a combination of MutS-mediated lesion recognition and the BER pathway (Figure 3). A role for Polη in expansion is intriguing given the demonstration that this polymerase is involved in replicating past DNA structures present at common fragile sites (Bergoglio *et al*., 2013). However, it is not necessary to invoke this particular pathway in order to account for the role of MMR proteins. For example, it may be that the MMR proteins act to facilitate the formation or to stabilize the hairpin intermediates arising during LP-BER.

In addition to a role in generating expansions, there is evidence to suggest that BER can result in contractions. Oxidative damage to human cells containing plasmids carrying CTG-repeats, or treatment of FRDA lymphoblasts with temozolomide, an alkylating agent, result in the generation of contractions (Lai *et al*., 2013, 2014). BER of the lesion caused by oxidative damage would involve the generation of an abasic site, and the presence of such a site within a synthetic $(CTG)_{20}$ or $(GAA)_{20}$ tract results in both deletions and expansions in *in vitro* reactions with reconstituted with BER proteins; the location of the damaged base has been proposed to govern whether BER results in expansion or contraction (Lai *et al*., 2013).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Contractions have been suggested to occur via hairpin bypass by Polβ (Xu *et al*., 2013). There is also evidence to suggest that BER can protect against expansions (Xu *et al*., 2014).

The role of nucleotide excision repair (NER) proteins in repeat instability

NER is involved in the repair of damaged bases including oxidized bases. NER is composed of two different pathways: one that occurs globally (global genome repair; GGR) and one that is confined to actively transcribed genes (transcription coupled repair; TCR) (see Hanawalt & Spivak, 2008 for review). In GGR, proteins like the XPC–Rad23B complex are responsible for DNA damage recognition, while in TCR, RNA polymerase stalls at a lesion and the blocked polymerase acts as the DNA damage signal, leading to the recruitment of proteins specific to TCR such as the Cockayne Syndrome B (CSB) protein (also known as ERCC6). The two pathways converge downstream of the DNA damage signal step and share a common lesion excision process in which an incision by XPF-ERCC1 5′ of the lesion and by XPG 3′ of the lesion lead to removal of 25–30 nucleotides. Repair is completed by Polδ, Polε, and/or Polκ and the nick is sealed by Lig1 or Lig3/XRCC1. A role for TCR in repeat expansion is appealing given the evidence for the importance of an open chromatin configuration and/or transcription for repeat instability (Goula *et al*., 2012b; Jung & Bonini, 2007; Lin & Wilson, 2012; Lin *et al*., 2006, 2010a,b; Lokanga *et al*., 2014a; Nakamori *et al*., 2011) and the fact that in HeLa cell extracts, slip-outs of $(CAG)_{20}$ or $(CTG)_{20}$ repeats on either strand cause stalling of RNA polymerase (Salinas-Rios *et al*., 2011).

Loss of XPC has little effect on repeat expansion in an HD mouse model (Dragileva *et al*., 2009) and siRNA knockdown of XPC had no effect in a human cell model for CAG/CTG contractions (Lin *et al*., 2006). This suggests that GGR is not responsible for generating either expansions or contractions, at least in these model systems. In contrast, there are a number of lines of evidence implicating TCR proteins in repeat instability. The loss of the fly XPG homolog, Mus201, reduced expansion in a Drosophila model of SCA3 (Jung & Bonini, 2007) while the loss of CSB in a mouse model for FXS led to a lower frequency of germ-line expansions in older females and a reduction in the extent of somatic expansions in some tissues (Zhao & Usdin, 2014). A role for CSB in expansions is appealing given the demonstration that a CSB variant is associated with increased expansion risk in humans with SCA3 (Martins *et al*., 2014). However, in a HD mouse model, the loss of CSB alone had no significant effect on somatic expansions in a wild-type background, although it did increase the somatic expansion frequency in *Ogg1* null mice (Kovtun *et al*., 2011). An increase in the germ-line expansion frequency was also seen in $Csb^{-/-}$ HD mice, although this increase did not reach statistical significance. Loss of XPA, a protein involved in stabilizing the denatured transcription bubble during TCR, also reduced somatic expansions in the neuronal tissue of a mouse model for SCA1, but had no effect on intergenerational expansion frequencies (Hubert *et al*., 2011). SiRNA knockdown of XPG, ERCC1, and CSB reduced repeat contractions in a human cell model of CAG repeat contractions (Lin & Wilson, 2007). A role for XPG in generating contractions would be consistent with the observation that XPG stimulates hairpin removal *in vitro* (Hou *et al*., 2011). Thus, evidence suggests that TCR proteins may contribute to both expansions and contractions in some model systems.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

It is possible to envision a model for TCR in the generation of both expansions and contractions, as illustrated in Figure 4. TCR could be induced during transcription stalling either as a result of the R loops that are known to be formed by many disease associated repeats (see section ‘‘Disease-associated repeats form non-B DNA structures’’) or the formation of secondary structures on the non-template strand (Belotserkovskii *et al*., 2013; Duquette *et al*., 2004). It is also possible that RNAPII stalls at hairpins formed during passage of a previous transcription complex. CSB and other factors are then recruited to the stalled transcription complex to initiate TCR, RNAPII is displaced, and the denatured bubble is stabilized by XPA and RPA. The RNAPII-blocking lesion is then removed by dual incision, the first carried out by XPF-ERCC1 that cleaves 5′ of the lesion, the second carried out by XPG occurring downstream of the lesion. Since XPG does not require a free 5′ single-stranded end (Hohl *et al*., 2003), cleavage may not be impeded by the structure in the flap. The result would be a 25–30 nt gap that is filled by Polδ, Polε, and/or Polκ. If the polymerase fills the gap faithfully, no tract length change occurs and the original allele is regenerated. If strand slippage occurs during gap filling DNA synthesis, an expansion would result if slippage involved the nascent strand and a contraction would result if it was the template strand that slipped (Figure 4).

The knockdown of MutSβ as well as various components of the TCR pathway both reduced the mutation rate in an *in vitro* assay for repeat contractions (Lin *et al*., 2006). It may be that MutSβ acts to facilitate repeat instability by binding to transcription-induced hairpins, thereby blocking RNAPII progression and inducing TCR. It is interesting that in yeast, Msh2 has physical and genetic interactions with multiple members of the TCR pathway (Saccharomyces Genome Database). Since MutLα knockdown increases contractions (Lin *et al*., 2009), it was suggested that engagement of MutLα may divert the repair away from the contraction-prone TCR pathway (Lin & Wilson, 2012), perhaps by favoring repair of the hairpin by MMR as illustrated in Figure 3.

There is also the intriguing possibility that some TCR components are acting outside of the TCR pathway to promote expansions (Figure 4). Since CSB is essential for TCR, the fact that loss of *Csb* in the FXD mouse model does not eliminate expansions completely, would be consistent with the idea that Csb plays an auxiliary role in expansion (Zhao & Usdin, 2014). For example, it could act in a BER-based pathway via its ability to increase the incision activity of NEIL1 and OGG1 (Muftuoglu *et al*., 2009; Tuo *et al*., 2002) and to positively regulate hOGG1 gene expression (Javeri *et al*., 2011; Tuo *et al*., 2002). CSB also facilitates annealing of single strands (Muftuoglu *et al*., 2006), which could promote hairpin formation.

Homology-dependent recombinational repair of gaps and stalled forks

Somatic expansions in differentiated cells are most likely occurring by one or more of the non-replicative mechanisms described in the previous sections. However, as summarized in section ‘‘Developmental timing of TNR expansions’’, there is evidence that CAG/CTG repeat expansions occur in replicating pre-meiotic cells, resulting in intergenerational transmission of disease, and that some replicating somatic tissues incur expansions. Instability in these cell types could arise either by a non-replicative repair pathway or in a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

replication-coupled repair process. Studies in model systems, summarized below, confirm that replication-coupled repair processes can be a source of large expansions.

As discussed earlier, DNA structures formed by repetitive DNA sequences are impediments for DNA synthesis (Kerrest *et al*., 2009; Krasilnikova *et al*., 2007; Liu & Leffak, 2012; Pelletier *et al*., 2003; Shah *et al*., 2012; Voineagu *et al*., 2009b) and can cause fork stalling or, if the lesion is bypassed, a gap behind the replication fork. Such ssDNA gaps occur when the replication fork encounters a lesion and leading and lagging strand synthesis become uncoupled (reviewed by Heller & Marians, 2006)). Also, nicks or small gaps that arise independently of DNA replication, for instance during BER or NER, can be enlarged or become DSBs if replicated (Byun *et al*., 2005; Hashimoto *et al*., 2010).

Post-replication gap repair and fork restart

DNA damage tolerance (DDT) pathways, also referred to as post-replication repair (PRR), are dependent on ubiquitination of PCNA, and can be subdivided into two categories: translesion synthesis (TLS) and error-free template switching (reviewed by Boiteux & Jinks-Robertson, 2013; Saugar *et al*., 2014). The TLS branch is dependent on the PCNA ubiquitin ligases, Rad6 and Rad18, which together monoubiquitinate Lys164 of PCNA. The monoubiquitin signal is required for the recruitment of translesion polymerases (e.g. Polζ or Polη) that synthesize across the lesion. TLS does not require invasion into a homologous template and can be mutagenic. However, mutations in the TLS polymerases had no effect on CAG repeat instability (Dixon & Lahue, 2002; Freudenreich *et al*., 1998) or GAA repeat stability (Shah *et al*., 2012) in budding yeast, indicating that the TLS pathway is not a significant source of expansions. Recent data suggests that when replicative polymerases are compromised, some GAA repeat expansions do occur by a Polζ-dependent mechanism (Shah *et al*., 2012), as do short duplications initiated by small hairpins (Northam *et al*., 2014), so this conclusion could be revisited.

The error-free branch of PRR is dependent on the mono-ubiquitination of PCNA by Rad6–Rad18 E2–E3 ubiquitin ligases but then requires additional ubiquitylation action by Ubc13–Mms2–Rad5 E2–E3 ubiquitin ligases (mammalian HLTF/SHPRH), leading to a poly-ubiquitinated PCNA molecule (Boiteux & Jinks-Robertson, 2013; Saugar *et al*., 2014). Error-free PRR further requires the action of the homologous recombination proteins Rad51, Rad52, Rad57, and Rad54 (Gangavarapu *et al*., 2007; Minca & Kowalski, 2010; Vanoli *et al*., 2010). The requirement for HR proteins is consistent with the use of the undamaged sister chromatid as a template for synthesis through a poorly understood template-switch mechanism that is similar to sister chromatid recombination (Boiteux & Jinks-Robertson, 2013; Saugar *et al*., 2014). Indeed, Rad5 is required for sister chromatid recombination at non-repetitive sequences (House *et al*., 2014b; Zhang & Lawrence, 2005). In Saccharomyces cerevisiae, spontaneous sister chromatid recombination (SCR) is proposed to occur because of gaps formed behind the replication fork (Mozlin *et al*., 2008), and is induced by both CAG repeats and inverted repeats (Nag *et al*., 2004). It is the propensity of these sequences to form secondary structures that is thought to impede replication and induce sister chromatid recombination, which is dependent on the presence of Rad52 (Nag *et al*., 2004).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Rad5, Rad18, HTLF, and PCNA ubiquitination have all been shown to inhibit $(CTG)_{13}$ or $(CAG)_{25}$ expansions in yeast and human cells (Daee *et al*., 2007; Frizzell *et al*., 2014). At a longer $(CAG)_{85}$ repeat, deletion of Rad5 also mildly increased expansions, 3-fold over wild-type (House *et al*., 2014b). The role of Rad5 at short CAG repeats was epistatic to the Srs2 protein, a 3′–5′ helicase and anti-recombinase, implicating a role for Srs2 in DNA damage tolerance as well, perhaps by unwinding hairpins (Daee *et al*., 2007; Dhar & Lahue, 2008). Yeast lacking Srs2 display an increase in CAG and CTG fragility and instability, which for longer repeats is dependent on the presence of Rad51 (Bhattacharyya & Lahue, 2004; Kerrest *et al*., 2009). In a haploid strain, two-dimensional (2D) gel electrophoresis analysis showed the formation of joint molecules at a $(CAG)_{45\text{-}55}$ tract (Kerrest *et al*., 2009). The joint molecules were dependent on the presence of Srs2 but not Rad51, indicating that they could be reversed forks or other template switch events between sister chromatids (Kerrest *et al*., 2009). Altogether these data indicate that CAG tracts induce both SCR and template switch events, and that, in yeast, the Srs2 helicase is important to prevent instability during this process. This template switch event must play an important role in repairing TNR-related gaps, because in its absence (e.g. in *rad5* or *rad18* strains) expansions occur, by an alternative unknown pathway.

The chromatin environment at gaps also contributes to efficient repair by error-free PRR (reviewed in House *et al*., 2014a). The absence of Anc1, a subunit of the chromatin modifying complexes INO80, SWI/SNF, and NuA3, leads to an increase in $(CAG)_{25}$ expansion frequency that is equivalent to the increase in *rad5* and *mms2* mutants (Daee *et al*., 2007; Erlich *et al*., 2008). However, Anc1 may play a role in an as yet undefined branch of PRR since it was not epistatic to all members of a single branch of error-free damage tolerance (Erlich *et al*., 2008). Indirect effects of chromatin modifiers also need to be considered. For example, knockdown of histone deacetylases (HDACs) Rpd3L and Hda1 (hHDAC3) suppresses CAG expansions in yeast and human cells, potentially by altering the stability of the Sae2 nuclease (Debacker *et al*., 2012). Our group recently demonstrated that histone H4-K16 acetylation by the chromatin-modifying complex NuA4 is specifically enriched at expanded CAG repeats in yeast, and is required for high-fidelity error-free PRR and $(CAG)_{85}$ repeat maintenance (House *et al*., 2014b). In the absence of Esa1, the catalytic subunit of the NuA4 complex, repeat expansions were significantly increased compared with wild-type cells and this increase was dependent on Rad52, Rad57, and Rad5, indicative of a template switching event that requires invasion into the sister chromatid (House *et al*., 2014b). Further, Rsc2, an acetyl-lysine-binding subunit of the RSC chromatin remodeler, was recruited to the CAG repeat coincident with the peak in H4-K16 acetylation, suggesting a possible recruitment mechanism for chromatin modifiers to promote template switching without expansions (House *et al*., 2014b). Thus, although template switch is a protective pathway, it can be a source of repeat length changes if it occurs without accessory factors such as helicases (Srs2) and chromatin remodelers or modifiers (Rsc2, H4-K16ac, and Anc1) that allow it to occur with fidelity.

In contrast to short $(CAG)_{13\text{-}25}$ repeats, ATTCT and GAA repeat expansions are promoted by the presence of Rad5 in yeast (Cherng *et al*., 2011; Shishkin *et al*., 2009), analogous to the situation described above for expansions of $(CAG)_{85}$ that occur in the absence of H4-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

K16ac. In addition, a *rad5* mutant displays decreased ATTCT fragility (Cherng *et al*., 2011), suggesting that template switching events can lead to chromosome breaks at this DNA unwinding element. Rad5-dependent expansions of the GAA repeat were proposed to occur by a template switching mechanism in which the GAA repeat expansions arise from dissociation of the leading strand from its normal template and aberrant copying from the newly synthesized Okazaki fragment (Shishkin *et al*., 2009). The model predicts that copying would not be dependent on DNA structure per se, but would be facilitated by pausing of the replication fork (Cherng *et al*., 2011; Shishkin *et al*., 2009).

How can one explain the different dependencies on Rad5 and Rad51/Rad52 observed for different types and sizes of repeats? We propose the following model to account for the somewhat contradictory roles of proteins in the PRR pathway on repeat instability (Figure 5). For longer or more ‘‘slippery’’ repeats (GAA, ATTCT, and longer CAGs), the fork stall could be strong enough to mediate a template switching event directly at the stalled fork, which could be equivalent to a fork reversal, hypothesized to be facilitated by Rad5 (Figure 5A). There is experimental evidence for fork reversal at both CAG and GAA repeats by direct visualization of replication intermediates by 2D gel electrophoresis and electron microscopy (EM) (Follonier *et al*., 2013; Fouche *et al*., 2006; Kerrest *et al*., 2009). For CAG repeats, the size needed to produce a fork stall stable enough to be visualized on a 2D gel is approximately 90–100 CAGs (G. -F. Richard, personal communication). Two models are drawn for the template switch: for hairpin-forming sequences, fold-back of the leading strand would allow DNA synthesis from the leading strand, resulting in a repeat expansion upon fork restart (Figure 5A, left) (first proposed by Mirkin, 2006). For non-hairpin-forming sequences, copying off of the lagging nascent strand provides the extra DNA synthesis, as proposed in Shishkin *et al*. (2009) for large-scale GAA expansions (Figure 5A, right; see Kim & Mirkin, 2013 for review). On the other hand, a single hairpin is more likely to be bypassed, leading to a post-replicative template switch that initiates from a gap, and looks more like SCR (Figure 5B; House *et al*., 2014b; Kerrest *et al*., 2009). This latter event may be more common for mid-length CAGs, at or above the expansion threshold of 35 repeats but still less than the size needed to produce a stable fork stall. For very short $(CAG)_{13–25}$ tracts, simple post-replicative hairpin unwinding by the Srs2 protein could be sufficient most of the time, with less engagement of the full recombination pathway (Figure 5C; Bhattacharyya & Lahue, 2004). This idea of length-dependent differences is supported by the fact that Tof1, a subunit of the replication-pausing complex, protects against instability for both ATTCT repeats (Cherng *et al*., 2011) and $(CAG)_{130–155}$ repeats, but has a lesser effect at $(CAG)_{85}$ (L. Gellon, M. Lahiri, C.H. Freudenreich, unpublished data). Thus, the Tof1 complex may act to limit the template-switching events at stalled forks that can allow for repeat expansions.

The role of helicases in resolving hairpins during repair and fork restart

As described above, the Srs2 protein plays a key role in preventing fragility and instability of expanded CAG repeats in yeast. Srs2 was shown to have a robust ability to unwind CTG and CCG hairpins *in vitro* (Anand *et al*., 2012; Bhattacharyya & Lahue, 2005). Additionally, Srs2 is needed to facilitate replication past a $(CGG)_{45}$ repeat that causes a barrier to replication via hairpin formation, determined using direct visualization of fork stalling *in*

vivo by 2D gel electrophoresis (Anand *et al*., 2012). Srs2 had no activity on replication barriers due to G-quadruplex structures or protein binding, thus it is specific to DNA hairpins. This Srs2 function was unique among the helicases tested (Sgs1, Pif1, and Rrm3), and was dependent on its helicase activity and its ability to interact with PCNA, but not on its Rad51 displacement activity. Altogether, at least three roles for the Srs2 protein at hairpin-forming sequences have been defined: (1) to facilitate fork progression through hairpin DNA structures (Anand *et al*., 2012) (Figure 6B), (2) to effect hairpin unwinding during PRR, independent of Rad51 displacement (Daee *et al*., 2007; Dhar & Lahue, 2008) (Figure 6A), and (3) to prevent CAG expansions and contractions by restraining Rad51-dependent recombination (Kerrest *et al*., 2009) (Figure 5B). In addition, deletion of Srs2 significantly increases chromosome fragility of both CAG and CGG repeats in yeast (Anand *et al*., 2012; Kerrest *et al*., 2009), likely because of increased fork stalling due to defective hairpin unwinding (Anand *et al*., 2012; J. Nguyen, R. Anand, C.H. Freudenreich, unpublished data).

Recent evidence indicates that the RTEL1 helicase performs at least some of these functions in human cells (Frizzell *et al*., 2014). *RTEL1* knockdown resulted in an increase in CAG expansions to a similar level as knockdown of Rad18 and HLTF, homologs of yeast Rad18 and Rad5, and RTEL1 could also partially substitute for Srs2 in yeast cells to prevent CAG expansions. Fbh1, another potential Srs2 ortholog, did not complement any of these phenotypes in yeast or protect against expansions in human cells. Although RTEL1 and Srs2 lack protein sequence homology and have opposite DNA unwinding polarities, these results indicate a functional conservation between the two enzymes with respect to CAG repeat expansions. Both helicases are able to unwind CAG and CTG hairpin structures *in vitro*, although RTEL1 additionally acts at G4 DNA and is important in telomere maintenance (Frizzell *et al*., 2014; Vannier *et al*., 2012, 2013). There may be other proteins in human cells that share in the ability to unwind hairpins and prevent repeat expansions. One candidate is PARI, which, like Srs2, interacts with sumoylated PCNA and prevents recombination, but since PARI is not an active helicase, it would have to work with another as yet unknown factor to actively unwind DNA structures (Moldovan *et al*., 2012). Another helicase that has been shown to prevent CAG instability and fragility in yeast is Sgs1 (Kerrest *et al*., 2009). Sgs1 can unwind hairpins and G4 DNA *in vitro*, as can its human homologs WRN and BLM (Anand *et al*., 2012; Bhattacharyya & Lahue, 2005; Huber *et al*., 2002; Shen & Loeb, 2001; Sun *et al*., 1998). Sgs1 and WRN interact with Polδ, and thus are well positioned to unwind structures on the lagging strand template. Indeed, deletion of Sgs1 led to a large increase in repeat contractions, but no increase in expansions, consistent with such a role (Bhattacharyya & Lahue, 2004; Kerrest *et al*., 2009). Interestingly, WRN was identified in a screen for activities in HeLa extracts that were able to stimulate repair of a CTG hairpin on the template strand (Chan *et al*., 2012). The WRN helicase efficiently unwound CTG hairpins in this system to promote Polδ-catalyzed DNA synthesis across the gap and prevent deletions (Chan *et al*., 2012). Thus Srs2/RTEL and Sgs1/WRN may work together to resolve hairpins on different strands (Figure 6). This mechanism could also be relevant during gap repair in non-dividing cells (Figure 3).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

The role of DSB repair in preventing repeat fragility and instability

Homologous recombination pathways of DSB repair

Recombination is a shared mechanism between gap repair and DSB repair that uses a homologous template, which can be a sister chromatid, homologous chromosome, or ectopic site, to copy and replace missing information (Symington & Gautier, 2011). Two-ended double-strand breaks (DSBs) are first bound by the MRX (yeast) or MRN (human) complex, consisting of Mre11, Rad50, and yeast Xrs2 or human Nbs1. MRN controls processing of the DNA ends by endonucleases Mre11 and Sae2 and exonucleases Exo1 and Dna2, to expose a single-stranded 3′ end and initiate HR (Mimitou & Symington, 2011). MRX/MRN also has a structural role in bridging DNA ends, as well as a role as a sensor and a mediator of the checkpoint response (Williams *et al*., 2007). In addition, Mre11 binds to stalled forks (Tittel-Elmer *et al*., 2009), where it could act to tether one-ended breaks to facilitate repair of the broken fork and fork restart.

Early studies in human patients showed no exchange of flanking markers at expanded repeat loci, and cases of conversion to the shorter allele were not documented, leading to the conclusion that meiotic crossovers or gene conversion between homologs were not involved in repeat expansions (reviewed in Monckton & Caskey, 1995). However mitotic recombination between sister chromatids does not lead to a detectable change in flanking sequence (Symington & Gautier, 2011). Furthermore, some human minisatellite loci show germ line repeat expansions that seem to arise from complex gene conversion events involving the homologous chromosome but that rarely are associated with the exchange of flanking markers (Jeffreys *et al*., 1994). Experiments in E. coli showed that long CAG/CTG repeats (67–175 repeat units) induce recombination, and the authors proposed that this occurs because the polymerase stalls within the repeat tract, which can cause nicks or DSBs that stimulate HR (Hebert & Wells, 2005; Napierala *et al*., 2002; Pluciennik *et al*., 2002). Expanded GAA/TTC repeats also stimulate mitotic crossovers in yeast (Tang *et al*., 2011). In yeast, the effect of HR on CAG repeat stability was assessed directly by induction of a DSB and selection for repair events that used an ectopically provided $(CAG)_{98}$ tract for repair (Richard *et al*., 2000). This experiment showed that repeat instability occurred during HR repair, resulting in a much higher percentage of expansions (13%) and contractions (30%), than a control that did not undergo repair (0% expansions and 10% contractions). Mre11 was required for efficient repair, and the proportion of expansions increased in its absence (Richard *et al*., 2000). Thus, synthesis-dependent HR from a DSB is a mechanism that can produce both repeat expansions and contractions.

Studies from the Freudenreich lab further investigated the role of the various DSB repair proteins and pathways in CAG repeat maintenance. CAG tracts of either 70 or 155 repeats contained on a yeast artificial chromosome (YAC) (Callahan *et al*., 2003) were evaluated for repeat fragility and instability (Sundararajan *et al*., 2010). Proteins needed for HR repair of DSBs (Rad51, Rad52, Mre11, Sae2, and Exo1) as well as those required for non-homologous end joining (NHEJ) repair (Mre11 and Dnl4) were important for preventing CAG fragility. Deletion of both Rad52 and Dnl4 (human Lig4) was additive, indicating that both pathways play important but separate roles in healing breaks within repetitive DNA.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Indeed, in the absence of these repair pathways, yeast cells containing long CAG tracts showed increased cell death compared with control cells with no repeat, underscoring the importance of these pathways in repair of DSBs that occur at expanded repeat tracts (Sundararajan & Freudenreich, 2011; Sundararajan *et al*., 2010).

DSB repair proteins were also required to protect against CAG/CTG instability in this yeast model. Consistent with the earlier study, expansions were increased 11-fold in the absence of Mre11, pointing to a key role for this protein in preventing repeat instability (Sundararajan *et al*., 2010). These expansions were suppressed in the absence of Rad52, and, therefore, occurred through aberrant recombination (Sundararajan *et al*., 2010). It was proposed that MRX plays an important role in stabilizing one-ended DSBs that occur at stalled forks, promoting correct strand alignment during HR-dependent fork restart (similar to Figure 2C). One-ended DSBs could form by fork breakage, or alternatively fork reversal is predicted to create a double-stranded DNA end to which MRX could bind. Deletion of the Mre11-associated Sae2 and Exo1 nucleases also increased both CAG fragility and expansions, although to a lesser extent than *mre11* , reinforcing a role for appropriate MRX-mediated processing of ends to promote efficient and correct repair of repeat-associated breaks (Sundararajan *et al*., 2010). Deletion of the gene encoding the Ctf18 protein, another factor speculated to have a role in fork stabilization, also increases CAG fragility and instability, with the expansions and contractions occurring during a Rad52-dependent process (Gellon *et al*., 2011). In summary, although HR is needed to repair DSBs within expanded repeats, it can also be a source of repeat instability. This could occur through misalignment (Figure 2C) and/or because of low fidelity of the repair process.

Despite the strong evidence for HR-induced repeat instability in yeast, in a DM1 mouse model, loss of the Rad54 gene product did not significantly suppress CTG instability, and the absence of the Rad52 gene decreased the size of expansions, but did not eliminate them (Savouret *et al*., 2003). These results have been interpreted to mean that HR has no effect on CAG/CTG instability in mammalian cells. So how can they be reconciled with the data from *E. coli* and yeast model systems that support a role for HR? First, consistent with the mouse data, deletion of HR proteins in wild-type yeast does not eliminate expansions. The effect of HR seems to be most important in repairing breaks, and it comes into play primarily when breaks are increased, such as in the *mre11* or *ctf18* backgrounds, at long repeats that break more frequently, or when gap repair by HR is compromised (such as during template switching in the absence of H4-K16ac). Furthermore, the knockouts used in the DM1 mouse studies, Rad52 and Rad54, do not eliminate HR in mammalian cells. Rad52 knockouts exhibit only a slightly reduced HR frequency and are not hyper-sensitive to DSB-inducing agents; Rad54 is not an essential HR protein, but serves to facilitate chromatin remodeling during HR (Suwaki *et al*., 2011; van Veelen *et al*., 2005). Despite this, the mean size of expansions was significantly decreased in Rad52$^{-/-}$ mice, supporting the idea that rare large expansions could occur during HR in these mice. In summary, data in multiple systems are in agreement that HR is not the major expansion mechanism operating in wild-type cells. Nonetheless, it is important in repairing breaks at long TNRs, and can produce expansions in some situations. Due to the technical difficulties of eliminating HR in mammalian cells, the role of HR in repeat instability in mammals remains to be fully assessed.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Recent results show that replication in the context of HR repair or HR-dependent fork restart proceeds with less fidelity and more mutations than normal replication, even without the complication of copying DNA repeats (Deem *et al*., 2011; Hicks *et al*., 2010; Iraqui *et al*., 2012). Recently, GAA repeats have been shown to induce mutagenesis up to 8 kb from the repeat site in yeast, which is enhanced in strains with polymerase defects and largely dependent on the Polζ TLS polymerase (Saini *et al*., 2013; Shah *et al*., 2012; Tang *et al*., 2013). The evidence indicates that DSBs are formed at the repeat, followed by strand invasion of the sister chromatid, and mutagenic synthesis by Polζ during HR repair (Saini *et al*., 2013). This repeat-induced mutagenesis (RIM) has also been observed for H-DNA- and Z-DNA-forming sequences introduced into mammalian cells (Wang & Vasquez, 2004, 2009; Wang *et al*., 2006). Thus, the *cis*-effects of structure-forming DNA extend even beyond the repeat itself, influencing genome mutation rates over a large region.

Of note, the absence of HR may also *create* repeat instability (Sundararajan *et al*., 2010). In this case, since HR is defective, the tract length changes must occur through another pathway, such as end-joining or single-strand annealing (SSA).

The role of single-strand annealing (SSA) and end-joining mechanisms (NHEJ and MMEJ) in repair at repeats

NHEJ does not use a homologous sequence as a template for repair, and instead directly ligates the broken ends together by the Dnl4/Lif1 ligase complex in yeast, Lig4/XRCC4 in mammals. End joining is an error-prone process, which can result in insertions or deletions of nucleotides or translocations (Chapman *et al*., 2012; Lieber, 2010). Although an error-prone pathway, NHEJ is favored over HR in G_1 or non-replicating cells, as copying from the homologous chromosome could result in loss of heterozygosity (Shrivastav *et al*., 2008). NHEJ does not require 5′ end resection, and in fact, occurs when end resection is blocked by binding of the Ku70–Ku80 (Ku) heterodimer to the broken ends (Lieber, 2010). If resection does occur to expose homology between the 3′ ends, those ends can anneal in a pathway called SSA, resulting in deletion of the resected sequence (Figure 2D). SSA is Rad51 independent and can be Rad52 independent in some circumstances (Ira & Haber, 2002; Mott & Symington, 2011; Prado & Aguilera, 1995).

Just as HR is required for error-free repair of CAG repeats in yeast, so too is intact NHEJ. In the absence of Dnl4 (Lig4), CAG repeat contraction frequency is significantly increased and remains so in *dnl4 rad52* cells, indicating that these contractions were not occurring through HR (Sundararajan *et al*., 2010). The absence of HR proteins Rad52, Rad51, and Rad54 also increased repeat contractions ~2.5-fold (Sundararajan *et al*., 2010). Because broken DNA ends are less protected by Ku in both *rad52* and *dnl4* backgrounds, end processing followed by SSA between repeats is an attractive mechanism to explain these contractions (Figure 2D, right); another possibility is MMEJ (see below). Indeed, in a yeast system where breaks are induced within a CAG repeat, contraction via SSA is a prominent outcome (Richard *et al*., 1999), and induction of a DSB by a CAG/CTG-specific TALEN induced 100% contractions in a highly specific manner (Richard *et al*., 2014). The same mechanism appears to be operating in human cells, as a CAG-targeted DSB by a zinc finger nuclease also induced frequent contractions (Mittelman *et al*., 2009). This makes nuclease-

directed cleavage of TNRs an attractive method for inducing contractions, which could potentially be used therapeutically.

Repeat expansions are also increased in the absence of HR proteins in some systems (Sundararajan *et al*., 2010) although not others (Miret *et al*., 1998; Savouret *et al*., 2003; Schweitzer & Livingston, 1999). In the yeast study by Sundararajan *et al*. (2010), more than half of the expansions in a *rad52* strain were eliminated in the *rad52 dnl4* background, indicating that an end-joining pathway may contribute to expansions in this system (Figure 2D, left). The observation of large expansions created during repair of DSBs by HR or end-joining repair is intriguing, as large expansions that occur during maternal transmission of the DM1 CTG and FXS CGG repeat appear to happen during oogenesis. A prominent stage of oogenesis is meiosis, where many breaks occur and are repaired. In yeast, breaks occur frequently at CAG repeats during meiosis and are repaired to give both expansions and contractions (Jankowski *et al*., 2000; Schweitzer *et al*., 2001).

A second pathway of end joining is microhomology-mediated end-joining (MMEJ), which uses short stretches of homology (5–25 bp) to align the broken DNA ends and promote ligation to resolve the lesion (McVey & Lee, 2008). This process is error-prone and favors contraction events, but can also result in templated insertions during Polβ or Polγ-dependent fill-in (Chan *et al*., 2010; McVey & Lee, 2008). Using human proteins in a reconstituted *in vitro* system, it was revealed that Polβ-mediated MMEJ promotes CAG expansions, and the frequency of these events is limited by the 9-1-1 (Rad9/Hus1/Rad1) DNA damage checkpoint complex (Crespan *et al*., 2012; Figure 2D, left).

Role of the DNA damage checkpoint in sensing damage at TNRs and regulating repair

Since expanded repeats can form unusual structures with regions of single stranded DNA, gaps, stalled forks, and DSBs, it is reasonable to think that they may activate the DNA damage checkpoint. The DNA damage checkpoint is the cellular response to damaged DNA. It consists of damage sensors, mediators, and transducers whose signal is mediated primarily through phosphorylation signaling cascades (Errico & Costanzo, 2012; Gobbini *et al*., 2013). One of the key inducers of the checkpoint is extensive single-stranded DNA. ssDNA can occur at large gaps, stalled forks (especially if the replisome is uncoupled from the replicative helicase), or DNA breaks that have been processed to expose ssDNA ends. Briefly, the RPA ssDNA-binding complex recruits the mammalian ATRIP/ATR complex (S. cerevisiae (sc) Ddc2/Mec1). ATR/Mec1 is a kinase that phosphorylates downstream effector kinases (Chk1 and scRad53 (scChk1 playing a lesser role)), and is thus a key transducer of the DNA damage checkpoint response. ssDNA gaps as small as 35 bp can weakly induce an ATR-dependent DNA damage checkpoint response; ssDNA greater than 200 nt induces a more robust response (MacDougall *et al*., 2007). In addition, ssDNA–dsDNA junctions bind the alternative PCNA clamp 9-1-1 (scRad17, Mec3, and Ddc1), loaded by clamp loader complex Rad17–RFC (scRad24–RFC), which contributes to ATR/Mec1 activation. At stalled forks, the mediator Claspin/scMrc1 facilitates ATR activation of the effector kinases. A second branch of the DNA damage checkpoint response flows through the ATM kinase and is responsive to DSBs. Unprocessed DSB ends activate the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

checkpoint primarily through MRN/MRX binding, activating the ATM/scTel1 kinase, which phosphorylates Chk2/scRad53, the main effector kinases of the DSB response cascade. MDC1 and 53BP1 are primary mediators of the DSB response in mammalian cells; Rad9 assumes this role in S. cerevisiae. The identity of checkpoint proteins important for maintaining repeats can provide insight into the prominent type of DNA damage they cause in a particular cell type or situation.

The role of the DNA damage checkpoint in cells with expanded CAG repeats was first investigated in yeast (Lahiri *et al*., 2004). In this study, it was found that CAG repeats of 85 and 135–155 units were highly sensitive to deletion of the genes encoding either the Mec1 and Ddc2 proteins (ATR/ATRIP) or the Rad9 protein; fragility and instability were increased in their absence as well as in cells with defective Rad53 signaling (Freudenreich & Lahiri, 2004; Lahiri *et al*., 2004). These data suggest that both stalled forks and DSBs exist in cells with expanded CAG tracts, and the checkpoint response is important in facilitating repair of these lesions. Of interest, $(CAG)_{85}$ tracts, but not the longer repeats, were also sensitive to the absence of a functional 9-1-1 complex (*rad17 rad24* mutant), supporting the model depicted in Figure 5 that there are less stalled forks and more gapped structures for this CAG repeat length. The most prominent instability phenotype in all these checkpoint mutants was an increase in contractions, ranging from 2 to 6-fold over wild-type cells, consistent with the increase in DSBs occurring. Related, shorter $(CAG)_{20}$ repeats that do not increase fragility did not show increased contractions in checkpoint mutants (Razidlo & Lahue, 2008). For $(CAG)_{85}$, expansions were modestly increased in some backgrounds, ~3-fold for *rad53*, *rad17*, and *rad24* mutants. Strengthening this result, in a more sensitive system where expansions could be selected for, a 3- to 6-fold increase was observed at a $(CAG)_{13}$ tract in checkpoint mutants (Razidlo & Lahue, 2008). It is interesting that a repeat this short would induce enough damage (presumably gaps) to require checkpoint proteins for appropriate repair. A recent study showed that the 9-1-1 complex might reduce CAG expansions more directly by inhibiting Polβ expansion-prone synthesis during MMEJ (Crespan *et al*., 2012).

In a FXS mouse model, the loss of one copy of the *Atr* gene was associated with an increased risk of expansion on maternal transmission (Entezam & Usdin, 2008). The paternal expansion frequency was not significantly affected in $Atr^{+/-}$ mice, perhaps because the expansion frequency was much higher to start with in males or because of sex differences in the levels of Atr in heterozygous animals. However, loss of one copy of the *Atr* gene resulted in the transmitted alleles being significantly larger in the offspring of both $Atr^{+/-}$ males and females relative to the offspring of wild-type animals. This would be consistent with the idea that ATR protects the genome against repeat expansion in animals of both sexes. Loss of one *Atm* allele also increased the expansion frequency in both males and females (Entezam & Usdin, 2009). Thus, both yeast and mouse models show that the ATR and ATM checkpoint proteins are important for preventing repeat instability, although the ratio of expansions to contractions was somewhat different in the various systems. Consistently, inhibition of these kinases leads to an increase in the incidence of chromosome fragility in both yeast and human cells (Freudenreich & Lahiri, 2004; Kumari & Usdin, 2009; Lahiri *et al*., 2004).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The Mrc1/Tof1/Csm3 (human Claspin/Timeless/Tipin) complex travels with the replisome and has an important role in coupling Polε to the MCM replicative helicase in both yeast and mammalian cells (Katou *et al*., 2003; Lou *et al*., 2008; Tourriere *et al*., 2005; Zegerman & Diffley, 2009). In addition, the Mrc1 protein has multiple amino acid motifs (SQ or TQ residues) that are phosphorylated in an ATM/Mec1-dependent manner, and facilitate CHK2/ Rad53 activation in response to stalled replication forks (Errico & Costanzo, 2012; Naylor *et al*., 2009; Voineagu *et al*., 2009a). By studying an Mrc1 mutant specifically deficient in checkpoint function (Mrc1AQ), it was determined that that Mrc1 checkpoint function plays a role in preventing $(CAG)_{13}$ repeat expansions in yeast, consistent with an earlier study using a *mrc1-1* allele and a $(CAG)_{85}$ tract (Freudenreich & Lahiri, 2004; Razidlo & Lahue, 2008). However, $(CAG)_{20}$ contractions are increased in *mrc1* cells independent of Mrc1 checkpoint function (Razidlo & Lahue, 2008). In human DM1 and Hela cells, knockdown of Claspin, Timeless, or Tipin all significantly increased expansions and contractions of an expanded CAG tract, either $(CAG)_{100}$ at an ectopic locus or $(CTG)_{400–1000}$ at the DM1 locus, confirming the importance of this complex in patient-derived cells (Liu & Leffak, 2012). The Mrc1/Claspin complex also has an important role independent of its checkpoint function in stabilization of forks stalled at CGG repeats. In yeast cells lacking Mrc1 and Tof1, stalling at an expanded $(CGG)_{40}$ repeat was increased, however, cells with the checkpoint deficient but replication proficient Mrc1AQ allele replicated the CGG tract normally (Voineagu *et al*., 2009a). The stall caused by CGG repeats may escape the replication checkpoint due to stable hairpin structures preventing exposure of ssDNA, resulting in fragility even in wild-type cells (Voineagu *et al*., 2009b). In summary, all results to date consistently support that the DNA damage checkpoint response is important for preventing triplet repeat expansions for all lengths, and additionally some checkpoint proteins have an additional role in preventing stalling, fragility, and contractions of longer repeats that interfere with fork progression.

Recently, more direct evidence has accumulated that the checkpoint response to DNA damage induced by expanded repeats may be a significant burden on cells, reducing their viability and lifespan. In yeast cells, $(CAG)_{70–155}$ repeat tracts can induce a length-dependent DNA checkpoint response that is increased in the absence of DNA repair pathways (Sundararajan & Freudenreich, 2011). In the absence of HR (*rad52*), CAG-dependent Rad53 phosphorylation is persistent, indicating that intact DSB repair is required for recovery from the checkpoint response. The fact that the checkpoint response was increased when DNA repair was deficient implies that it occurred as a response to DNA damage (rather than RNA or protein toxicity). Additionally, CAG tract-induced Rad53 phosphorylation is abolished in an *mre11* mutant, implicating this complex in sensing damage at the repeat to initiate a checkpoint response (Sundararajan & Freudenreich, 2011). Human neuronal cells are sensitive to DNA damage and induce apoptosis in response, so it would be important to test whether expanded repeat tracts impact cell health in humans. This is complicated by the fact that the RNA transcribed from the repeat can also be toxic, as can the translated protein, especially for the polyglutamine diseases. It has been shown that bidirectional transcription through an expanded CAG repeat in human cells induces a cell stress response, cell cycle arrest, and eventually massive cell death via apoptosis (Lin *et al*., 2010a,b). This response is signaled through RPA, ATR, CHK1, and p53, indicating that it is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

due to single-stranded DNA of some nature that occurs with convergent transcription. The aberrant structure that induces the response is speculated to be a ‘‘double bubble’’, with stalled RNAPII complexes on both strands; indeed the toxicity is worse in TCR mutants (Lin & Wilson, 2012). Together, these results point to a potential mechanism of cell death in the terminally differentiated cells that are affected in repeat diseases, and suggest this is an area that deserves future study.

Conclusions

As evident from this review, work in various model organisms and *in vitro* model systems has demonstrated that there are many ways in which repeats can expand and contract. Which pathway is used seems to depend on a variety of cis- and *trans*-acting factors, including the transcriptional activity of the locus containing the repeat, the chromatin structure of the region, the level of oxidative stress, and the level of expression of repair proteins in that particular cell type. However, many questions remain as to which of these pathways occur at the disease locus in humans and which of them are disease relevant. Because of the importance of locus-specific factors, tissue type, and developmental window in the choice between pathways, none of the model systems, although valuable, can be assumed to recapitulate completely the situation in patients. Nonetheless, knowledge of the pathways and proteins that can contribute to repeat instability in these systems can provide the basis for therapeutic ideas. In addition, through study of repeat expansion and contraction mechanisms, scientists have learned a great deal about DNA replication and DNA repair pathways and how they operate in the context of structure-forming sequences and in different cells and tissues. Insights into protein functions outside of their canonical pathways have also been made, and contributed to our understanding of DNA repair. The field has contributed many novel concepts, such as the possibility of structure-induced gratuitous repair, the importance of 5′ flap processing for preventing insertions in genomes, methods of fork progression through structural barriers and helicase unwinding preferences, the toxicity of bidirectional transcription, and insights into how chromatin structure and modifications prevent genome instability, to name a few. More surprises surely remain to be discovered, providing new insights into repeat expansion diseases and the sophisticated yet imperfect mechanisms of genome maintenance.

Acknowledgements

We thank Sergei Mirkin, Vincent Dion, Melissa Koch, Jane Kim, Bruce Hayward, and Simran Kaushal for helpful comments. Thanks to Guy-Franck Richard for sharing unpublished data.

Research in CHF’s laboratory is supported by the National Institutes of Health (Award P01GM105473), National Science Foundation (MCB1330743), and Tufts University. Research in KU’s laboratory is supported by the Intramural Program of the National Institute of Diabetes, Kidney and Digestive Diseases (DK057808).

References

Anand RP, Shah KA, Niu H, et al. Overcoming natural replication barriers: differential helicase requirements. Nucleic Acids Res. 2012; 40:1091–105. [PubMed: 21984413]

Balakumaran BS, Freudenreich CH, Zakian VA. CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in *Saccharomyces cerevisiae*. Hum Mol Genet. 2000; 9:93–100. [PubMed: 10587583]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Belotserkovskii BP, Liu R, Tornaletti S, et al. Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. Proc Natl Acad Sci USA. 2010; 107:12816–21. [PubMed: 20616059]

Belotserkovskii BP, Mirkin SM, Hanawalt PC. DNA sequences that interfere with transcription: implications for genome function and stability. Chem Rev. 2013; 113:8620–37. [PubMed: 23972098]

Bergoglio V, Boyer AS, Walsh E, et al. DNA synthesis by Pol eta promotes fragile site stability by preventing under-replicated DNA in mitosis. J Cell Biol. 2013; 201:395–408. [PubMed: 23609533]

Bhattacharyya S, Lahue RS. *Saccharomyces cerevisiae* Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. Mol Cell Biol. 2004; 24:7324–30. [PubMed: 15314145]

Bhattacharyya S, Lahue RS. Srs2 helicase of *Saccharomyces cerevisiae* selectively unwinds triplet repeat DNA. J Biol Chem. 2005; 280:33311–17. [PubMed: 16085654]

Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem. 2013; 5:182–6. [PubMed: 23422559]

Boiteux S, Jinks-Robertson S. DNA repair mechanisms and the bypass of DNA damage in *Saccharomyces cerevisiae*. Genetics. 2013; 193:1025–64. [PubMed: 23547164]

Bourn RL, De Biase I, Pinto RM, et al. Pms2 suppresses large expansions of the (GAA.TTC)n sequence in neuronal tissues. PLoS One. 2012; 7:e47085. [PubMed: 23071719]

Bowater RP, Jaworski A, Larson JE, et al. Transcription increases the deletion frequency of long CTG.CAG triplet repeats from plasmids in Escherichia coli. Nucleic Acids Res. 1997; 25:2861–8. [PubMed: 9207036]

Brock GJ, Anderson NH, Monckton DG. Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. Hum Mol Genet. 1999; 8:1061–7. [PubMed: 10332038]

Byun TS, Pacek M, Yee MC, et al. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. Genes Dev. 2005; 19:1040–52. [PubMed: 15833913]

Callahan JL, Andrews KJ, Zakian VA, Freudenreich CH. Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. Mol Cell Biol. 2003; 23:7849–60. [PubMed: 14560028]

Cannavo E, Marra G, Sabates-Bellver J, et al. Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. Cancer Res. 2005; 65:10759–66. [PubMed: 16322221]

Casella M, Lucarelli M, Simili M, et al. Spontaneous chromosome loss and colcemid resistance in lymphocytes from patients with myotonic dystrophy type 1. Cytogenet Genome Res. 2003; 100:224–9. [PubMed: 14526184]

Chan NL, Guo J, Zhang T, et al. Coordinated processing of 3′ slipped (CAG)n/(CTG)n hairpins by DNA polymerases beta and delta preferentially induces repeat expansions. J Biol Chem. 2013; 288:15015–22. [PubMed: 23585564]

Chan NL, Hou C, Zhang T, et al. The Werner syndrome protein promotes CAG/CTG repeat stability by resolving large (CAG)(n)/(CTG)(n) hairpins. J Biol Chem. 2012; 287:30151–6. [PubMed: 22787159]

Chan SH, Yu AM, McVey M. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. PLoS Genet. 2010; 6:e1001005. [PubMed: 20617203]

Chapman JR, Taylor MR, Boulton SJ. Playing the end game: DNA double-strand break repair pathway choice. Mol Cell. 2012; 47:497–510. [PubMed: 22920291]

Charbonneau N, Amunugama R, Schmutte C, et al. Evidence that hMLH3 functions primarily in meiosis and in hMSH2-hMSH3 mismatch repair. Cancer Biol Ther. 2009; 8:1411–20. [PubMed: 19483466]

Cherng N, Shishkin AA, Schlager LI, et al. Expansions, contractions, and fragility of the spinocerebellar ataxia type 10 pentanucleotide repeat in yeast. Proc Natl Acad Sci USA. 2011; 108:2843–8. [PubMed: 21282659]

Chung MY, Ranum LP, Duvick LA, et al. Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. Nat Genet. 1993; 5:254–8. [PubMed: 8275090]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Claassen DA, Lahue RS. Expansions of CAG.CTG repeats in immortalized human astrocytes. Hum Mol Genet. 2007; 16:3088–96. [PubMed: 17881653]

Clark RM, De Biase I, Malykhina AP, et al. The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. Hum Genet. 2007; 120:633–40. [PubMed: 17024371]

Cleary JD, Nichol K, Wang YH, Pearson CE. Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. Nat Genet. 2002; 31:37–46. [PubMed: 11967533]

Cleary JD, Pearson CE. Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. Trends Genet. 2005; 21:272–80. [PubMed: 15851063]

Cleary JD, Tome S, Lopez Castel A, et al. Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. Nat Struct Mol Biol. 2010; 17:1079–87. [PubMed: 20711191]

Crespan E, Czabany T, Maga G, Hubscher U. Microhomology-mediated DNA strand annealing and elongation by human DNA polymerases lambda and beta on normal and repetitive DNA sequences. Nucleic Acids Res. 2012; 40:5577–90. [PubMed: 22373917]

Daee DL, Mertz T, Lahue RS. Postreplication repair inhibits CAG.CTG repeat expansions in *Saccharomyces cerevisiae*. Mol Cell Biol. 2007; 27:102–10. [PubMed: 17060452]

De Temmerman N, Sermon K, Seneca S, et al. Intergenerational instability of the expanded CTG repeat in the DMPK gene: studies in human gametes and preimplantation embryos. Am J Hum Genet. 2004; 75:325–9. [PubMed: 15185171]

Dean NL, Tan SL, Ao A. Instability in the transmission of the myotonic dystrophy CTG repeat in human oocytes and preimplantation embryos. Fertil Steril. 2006; 86:98–105. [PubMed: 16716318]

Debacker K, Frizzell A, Gleeson O, et al. Histone deacetylase complexes promote trinucleotide repeat expansions. PLoS Biol. 2012; 10:e1001257. [PubMed: 22363205]

Deem A, Keszthelyi A, Blackgrove T, et al. Break-induced replication is highly inaccurate. PLoS Biol. 2011; 9:e1000594. [PubMed: 21347245]

DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011; 72:245–56. [PubMed: 21944778]

Dhar A, Lahue RS. Rapid unwinding of triplet repeat hairpins by Srs2 helicase of *Saccharomyces cerevisiae*. Nucleic Acids Res. 2008; 36:3366–73. [PubMed: 18440969]

Dion V. Tissue specificity in DNA repair: lessons from trinucleotide repeat instability. Trends Genet: TIG. 2014; 30:220–9. [PubMed: 24842550]

Dion V, Lin Y, Hubert L Jr, et al. Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. Hum Mol Genet. 2008; 17:1306–17. [PubMed: 18252747]

Dion V, Wilson JH. Instability and chromatin structure of expanded trinucleotide repeats. Trends Genet: TIG. 2009; 25:288–97. [PubMed: 19540013]

Dixon MJ, Lahue R. Examining the potential role of DNA polymerases eta and zeta in triplet repeat instability in yeast. DNA Repair. 2002; 1:763–70. [PubMed: 12509280]

Dragileva E, Hendricks A, Teed A, et al. Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. Neurobiol Dis. 2009; 33:37–47. [PubMed: 18930147]

Du J, Campau E, Soragni E, et al. Length-dependent CTG.CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells. Hum Mol Genet. 2013; 22:5276–87. [PubMed: 23933738]

Du J, Campau E, Soragni E, et al. Role of mismatch repair enzymes in GAA.TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells. J Biol Chem. 2012; 287:29861–72. [PubMed: 22798143]

Duquette ML, Handa P, Vincent JA, et al. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev. 2004; 18:1618–29. [PubMed: 15231739]

Dure L St, Landwehrmeyer GB, Golden J, et al. IT15 gene expression in fetal human brain. Brain Res. 1994; 659:33–41. [PubMed: 7820679]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Eichler EE, Holden JJ, Popovich BW, et al. Length of uninterrupted CGG repeats determines instability in the FMR1 gene. Nat Genet. 1994; 8:88–94. [PubMed: 7987398]

Ennis S, Murray A, Brightwell G, et al. Closely linked cis-acting modifier of expansion of the CGG repeat in high risk FMR1 haplotypes. Hum Mutat. 2007; 28:1216–24. [PubMed: 17674408]

Entezam A, Lokanga AR, Le W, et al. Potassium bromate, a potent DNA oxidizing agent, exacerbates germline repeat expansion in a fragile X premutation mouse model. Hum Mutat. 2010; 31:611–16. [PubMed: 20213777]

Entezam A, Usdin K. ATR protects the genome against CGG.CCG-repeat expansion in Fragile X premutation mice. Nucleic Acids Res. 2008; 36:1050–6. [PubMed: 18160412]

Entezam A, Usdin K. ATM and ATR protect the genome against two different types of tandem repeat instability in Fragile X premutation mice. Nucleic Acids Res. 2009; 37:6371–7. [PubMed: 19710035]

Erlich RL, Fry RC, Begley TJ, et al. Anc1, a protein associated with multiple transcription complexes, is involved in postreplication repair pathway in S. cerevisiae. PLoS One. 2008; 3:e3717. [PubMed: 19005567]

Errico A, Costanzo V. Mechanisms of replication fork protection: a safeguard for genome stability. Crit Rev Biochem Mol Biol. 2012; 47:222–35. [PubMed: 22324461]

Ezzatizadeh V, Pinto RM, Sandi C, et al. The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model. Neurobiol Dis. 2012; 46:165–71. [PubMed: 22289650]

Ezzatizadeh V, Sandi C, Sandi M, et al. MutLalpha heterodimers modify the molecular phenotype of Friedreich ataxia. PLoS One. 2014; 9:e100523. [PubMed: 24971578]

Foiry L, Dong L, Savouret C, et al. Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. Human Genet. 2006; 119:520–6. [PubMed: 16552576]

Fojtik P, Vorlickova M. The fragile X chromosome (GCC) repeat folds into a DNA tetraplex at neutral pH. Nucleic Acids Res. 2001; 29:4684–90. [PubMed: 11713318]

Follonier C, Oehler J, Herrador R, Lopes M. Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions. Nat Struct Mol Biol. 2013; 20:486–94. [PubMed: 23454978]

Fortune MT, Vassilopoulos C, Coolbaugh MI, et al. Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. Hum Mol Genet. 2000; 9:439–45. [PubMed: 10655554]

Fouche N, Ozgur S, Roy D, Griffith JD. Replication fork regression in repetitive DNAs. Nucleic Acids Res. 2006; 34:6044–50. [PubMed: 17071963]

Freudenreich CH, Kantrow SM, Zakian VA. Expansion and length-dependent fragility of CTG repeats in yeast. Science. 1998; 279:853–6. [PubMed: 9452383]

Freudenreich CH, 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. 2004; 3:1370–4. [PubMed: 15483399]

Freudenreich CH, Stavenhagen JB, Zakian VA. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. Mol Cell Biol. 1997; 17:2090–8. [PubMed: 9121457]

Frizzell A, Nguyen JH, Petalcorin MI, et al. RTEL1 inhibits trinucleotide repeat expansions and fragility. Cell Rep. 2014; 6:827–35. [PubMed: 24561255]

Fry M, Loeb LA. The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. Proc Natl Acad Sci USA. 1994; 91:4950–4. [PubMed: 8197163]

Gacy AM, Goellner G, Juranic N, et al. Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. Cell. 1995; 81:533–40. [PubMed: 7758107]

Gangavarapu V, Prakash S, Prakash L. Requirement of RAD52 group genes for postreplication repair of UV-damaged DNA in *Saccharomyces cerevisiae*. Mol Cell Biol. 2007; 27:7758–64. [PubMed: 17785441]

Gannon AM, Frizzell A, Healy E, Lahue RS. MutSbeta and histone deacetylase complexes promote expansions of trinucleotide repeats in human cells. Nucleic Acids Res. 2012; 40:10324–33. [PubMed: 22941650]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Gao R, Matsuura T, Coolbaugh M, et al. Instability of expanded CAG/CAA repeats in spinocerebellar ataxia type 17. Eur J Hum Genet. 2008; 16:215–22. [PubMed: 18043721]

Gatchel JR, Zoghbi HY. Diseases of unstable repeat expansion: mechanisms and common principles. Nat Rev Genet. 2005; 6:743–55. [PubMed: 16205714]

Gellon L, Razidlo DF, Gleeson O, et al. New functions of Ctf18-RFC in preserving genome stability outside its role in sister chromatid cohesion. PLoS Genet. 2011; 7:e1001298. [PubMed: 21347277]

Gerhardt J, Tomishima MJ, Zaninovic N, et al. The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells. Mol Cell. 2014a; 53:19–31. [PubMed: 24289922]

Gerhardt J, Zaninovic N, Zhan Q, et al. Cis-acting DNA sequence at a replication origin promotes repeat expansion to fragile X full mutation. J Cell Biol. 2014b; 206:599–607. [PubMed: 25179629]

Gobbini E, Cesena D, Galbiati A, et al. Interplays between ATM/Tel1 and ATR/Mec1 in sensing and signaling DNA double-strand breaks. DNA Repair (Amst). 2013; 12:791–9. [PubMed: 23953933]

Gomes-Pereira M, Fortune MT, Ingram L, et al. Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. Hum Mol Genet. 2004; 13:1815–25. [PubMed: 15198993]

Gomes-Pereira M, Hilley JD, Morales F, et al. Disease-associated CAG.CTG triplet repeats expand rapidly in non-dividing mouse cells, but cell cycle arrest is insufficient to drive expansion. Nucleic Acids Res. 2014; 42:7047–56. [PubMed: 24860168]

Gonitel R, Moffitt H, Sathasivam K, et al. DNA instability in postmitotic neurons. Proc Natl Acad Sci USA. 2008; 105:3467–72. [PubMed: 18299573]

Gorbunova V, Seluanov A, Mittelman D, Wilson JH. Genome-wide demethylation destabilizes CTG.CAG trinucleotide repeats in mammalian cells. Hum Mol Genet. 2004; 13:2979–89. [PubMed: 15459182]

Goula AV, Berquist BR, Wilson DM III, et al. Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. PLoS Genet. 2009; 5:e1000749. [PubMed: 19997493]

Goula AV, Festenstein R, Merienne K. Tissue-dependent regulation of RNAP II dynamics: the missing link between transcription and trinucleotide repeat instability in diseases? Transcription. 2013; 4:172–6. [PubMed: 23989661]

Goula AV, Pearson CE, Della Maria J, et al. The nucleotide sequence, DNA damage location, and protein stoichiometry influence the base excision repair outcome at CAG/CTG repeats. Biochemistry. 2012a; 51:3919–32. [PubMed: 22497302]

Goula AV, Stys A, Chan JP, et al. Transcription elongation and tissue-specific somatic CAG instability. PLoS Genet. 2012b; 8:e1003051. [PubMed: 23209427]

Grabczyk E, Mancuso M, Sammarco MC. A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP *in vitro*. Nucleic Acids Res. 2007; 35:5351–9. [PubMed: 17693431]

Grabczyk E, Usdin K. The GAATTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. Nucleic Acids Res. 2000; 28:2815–22. [PubMed: 10908340]

Grigg JC, Shumayrikh N, Sen D. G-quadruplex structures formed by expanded hexanucleotide repeat RNA and DNA from the neurodegenerative disease-linked c9orf72 gene efficiently sequester and activate heme. PLoS One. 2014; 9:e106449. [PubMed: 25207541]

Groh M, Lufino MM, Wade-Martins R, Gromak N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet. 2014; 10:e1004318. [PubMed: 24787137]

Halabi A, Ditch S, Wang J, Grabczyk E. DNA mismatch repair complex MutSbeta promotes GAA.TTC repeat expansion in human cells. J Biol Chem. 2012; 287:29958–67. [PubMed: 22787155]

Hanawalt PC, Spivak G. Transcription-coupled DNA repair: two decades of progress and surprises. Nat Rev Mol Cell Biol. 2008; 9:958–70. [PubMed: 19023283]

Author Manuscript

Hansen RS, Canfield TK, Fjeld AD, et al. A variable domain of delayed replication in FRAXA fragile X chromosomes: X inactivation-like spread of late replication. Proc Natl Acad Sci USA. 1997; 94:4587–92. [PubMed: 9114034]

Hansen RS, Canfield TK, Lamb MM, et al. Association of fragile X syndrome with delayed replication of the FMR1 gene. Cell. 1993; 73:1403–9. [PubMed: 8324827]

Hashimoto Y, Chaudhuri AR, Lopes M, Costanzo V. Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. Nat Struct Mol Biol. 2010; 17:1305–11. [PubMed: 20935632]

Hebert ML, Wells RD. Roles of double-strand breaks, nicks, and gaps in stimulating deletions of CTG.CAG repeats by intramolecular DNA repair. J Mol Biol. 2005; 353:961–79. [PubMed: 16213518]

Heidenfelder BL, Makhov AM, Topal MD. Hairpin formation in Friedreich's ataxia triplet repeat expansion. J Biol Chem. 2003; 278:2425–31. [PubMed: 12441336]

Heller RC, Marians KJ. Replication fork reactivation downstream of a blocked nascent leading strand. Nature. 2006; 439:557–62. [PubMed: 16452972]

Henricksen LA, Tom S, Liu Y, Bambara RA. Inhibition of flap endonuclease 1 by flap secondary structure and relevance to repeat sequence expansion. J Biol Chem. 2000; 275:16420–7. [PubMed: 10748145]

Hicks WM, Kim M, Haber JE. Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. Science. 2010; 329:82–5. [PubMed: 20595613]

Hohl M, Thorel F, Clarkson SG, Scharer OD. Structural determinants for substrate binding and catalysis by the structure-specific endonuclease XPG. J Biol Chem. 2003; 278:19500–8. [PubMed: 12644470]

Hou C, Zhang T, Tian L, et al. The role of XPG in processing (CAG)n/(CTG)n DNA hairpins. Cell Biosci. 2011; 1:11. [PubMed: 21711735]

House NC, Koch MR, Freudenreich CH. Chromatin modifications and DNA repair: beyond double-strand breaks. Front Genet. 2014a; 5 Article 296.

House NC, Yang JH, Walsh SC, et al. NuA4 initiates dynamic histone H4 acetylation to promote high-fidelity sister chromatid recombination at postreplication gaps. Mol Cell. 2014b; 55:818–28. [PubMed: 25132173]

Huber MD, Lee DC, Maizels N. G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. Nucleic Acids Res. 2002; 30:3954–61. [PubMed: 12235379]

Hubert L Jr, Lin Y, Dion V, Wilson JH. Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. Hum Mol Genet. 2011; 20:4822–30. [PubMed: 21926083]

Imbert G, Saudou F, Yvert G, et al. Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. Nat Genet. 1996; 14:285–91. [PubMed: 8896557]

Ira G, Haber JE. Characterization of RAD51-independent break-induced replication that acts preferentially with short homologous sequences. Mol Cell Biol. 2002; 22:6384–92. [PubMed: 12192038]

Iraqui I, Chekkal Y, Jmari N, et al. Recovery of arrested replication forks by homologous recombination is error-prone. PLoS Genet. 2012; 8:e1002976. [PubMed: 23093942]

Jalal SM, Lindor NM, Michels VV, et al. Absence of chromosome fragility at 19q13.3 in patients with myotonic dystrophy. Am J Med Genet. 1993; 46:441–3. [PubMed: 8357018]

Jankowski C, Nasar F, Nag DK. Meiotic instability of CAG repeat tracts occurs by double-strand break repair in yeast. Proc Natl Acad Sci USA. 2000; 97:2134–9. [PubMed: 10681451]

Jarem DA, Wilson NR, Delaney S. Structure-dependent DNA damage and repair in a trinucleotide repeat sequence. Biochemistry. 2009; 48:6655–63. [PubMed: 19527055]

Jarem DA, Wilson NR, Schermerhorn KM, Delaney S. Incidence and persistence of 8-oxo-7,8-dihydroguanine within a hairpin intermediate exacerbates a toxic oxidation cycle associated with trinucleotide repeat expansion. DNA Repair (Amst). 2011; 10:887–96. [PubMed: 21727036]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Javeri A, Lyons JG, Huang XX, Halliday GM. Downregulation of Cockayne syndrome B protein reduces human 8-oxoguanine DNA glycosylase-1 expression and repair of UV radiation-induced 8-oxo-7,8-dihydro-2′-deoxyguanine. Cancer Sci. 2011; 102:1651–8. [PubMed: 21668583]

Jeffreys AJ, Tamaki K, MacLeod A, et al. Complex gene conversion events in germline mutation at human minisatellites. Nat Genet. 1994; 6:136–45. [PubMed: 8162067]

Jung J, Bonini N. CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. Science. 2007; 315:1857–9. [PubMed: 17332375]

Kadyrov FA, Genschel J, Fang Y, et al. A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. Proc Natl Acad Sci USA. 2009; 106:8495–500. [PubMed: 19420220]

Kang S, Jaworski A, Ohshima K, Wells RD. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. Nat Genet. 1995; 10:213–18. [PubMed: 7663518]

Kantartzis A, Williams GM, Balakrishnan L, et al. Msh2–Msh3 interferes with Okazaki fragment processing to promote trinucleotide repeat expansions. Cell Rep. 2012; 2:216–22. [PubMed: 22938864]

Katou Y, Kanoh Y, Bando M, et al. S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. Nature. 2003; 424:1078–83. [PubMed: 12944972]

Kennedy L, Evans E, Chen CM, et al. Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. Hum Mol Genet. 2003; 12:3359–67. [PubMed: 14570710]

Kerrest A, Anand RP, Sundararajan R, et al. SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. Nat Struct Mol Biol. 2009; 16:159–67. [PubMed: 19136956]

Kettani A, Kumar RA, Patel DJ. Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. J Mol Biol. 1995; 254:638–56. [PubMed: 7500339]

Kim HM, Narayanan V, Mieczkowski PA, et al. Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. EMBO J. 2008; 27:2896–906. [PubMed: 18833189]

Kim JC, Mirkin SM. The balancing act of DNA repeat expansions. Curr Opin Genet Dev. 2013; 23:280–8. [PubMed: 23725800]

Kovalenko M, Dragileva E, Claire J St, et al. Msh2 acts in medium-spiny striatal neurons as an enhancer of CAG instability and mutant huntingtin phenotypes in huntington’s disease knock-in mice. PLoS One. 2012; 7:e44273. doi:10.1371/journal.pone.0044273. [PubMed: 22970194]

Kovtun IV, Johnson KO, McMurray CT. Cockayne syndrome B protein antagonizes OGG1 in modulating CAG repeat length *in vivo*. Aging. 2011; 3:509–14. [PubMed: 21566259]

Kovtun IV, Liu Y, Bjoras M, et al. OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. Nature. 2007; 447:447–52. [PubMed: 17450122]

Kovtun IV, McMurray CT. Trinucleotide expansion in haploid germ cells by gap repair. Nat Genet. 2001; 27:407–11. [PubMed: 11279522]

Kovtun IV, Thornhill AR, McMurray CT. Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations. Human Mol Genet. 2004; 13:3057–68. [PubMed: 15496421]

Krasilnikova MM, Kireeva ML, Petrovic V, et al. Effects of Friedreich’s ataxia (GAA)n*(TTC)n repeats on RNA synthesis and stability. Nucleic Acids Res. 2007; 35:1075–84. [PubMed: 17264130]

Krasilnikova MM, Mirkin SM. Replication stalling at Friedreich’s ataxia (GAA)n repeats *in vivo*. Mol Cell Biol. 2004; 24:2286–95. [PubMed: 14993268]

Ku S, Soragni E, Campau E, et al. Friedreich’s ataxia induced pluripotent stem cells model intergenerational GAATTC triplet repeat instability. Cell Stem Cell. 2010; 7:631–7. [PubMed: 21040903]

Kumari D, Usdin K. Chromatin remodeling in the noncoding repeat expansion diseases. J Biol Chem. 2009; 284:7413–17. [PubMed: 18957431]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

La Spada AR, Wilson EM, Lubahn DB, et al. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature. 1991; 352:77–9. [PubMed: 2062380]

Ladd PD, Smith LE, Rabaia NA, et al. An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Hum Mol Genet. 2007; 16:3174–87. [PubMed: 17921506]

Lahiri M, Gustafson TL, Majors ER, Freudenreich CH. Expanded CAG repeats activate the DNA damage checkpoint pathway. Mol Cell. 2004; 15:287–93. [PubMed: 15260979]

Lai Y, Beaver JM, Lorente K, et al. Base excision repair of chemotherapeutically-induced alkylated DNA damage predominantly causes contractions of expanded GAA repeats associated with Friedreich's ataxia. PLoS One. 2014; 9:e93464. [PubMed: 24691413]

Lai Y, Xu M, Zhang Z, Liu Y. Instability of CTG repeats is governed by the position of a DNA base lesion through base excision repair. PLoS One. 2013; 8:e56960. [PubMed: 23468897]

Lam EY, Beraldi D, Tannahill D, Balasubramanian S. G-quadruplex structures are stable and detectable in human genomic DNA. Nat Commun. 2013; 4 Article no. 1796.

Lang WH, Coats JE, Majka J, et al. Conformational trapping of mismatch recognition complex MSH2/MSH3 on repair-resistant DNA loops. Proc Natl Acad Sci USA. 2011; 108:E837–44. [PubMed: 21960445]

Lia AS, Seznec H, Hofmann-Radvanyi H, 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. 1998; 7:1285–91. [PubMed: 9668171]

Libby RT, Hagerman KA, Pineda VV, et al. CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. PLoS Genet. 2008; 4:e1000257. [PubMed: 19008940]

Libby RT, Monckton DG, Fu Y-H, et al. Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. Hum Mol Genet. 2003; 12:41–50. [PubMed: 12490531]

Lieber MR. NHEJ and its backup pathways in chromosomal translocations. Nat Struct Mol Biol. 2010; 17:393–5. [PubMed: 20368722]

Lin Y, Dent SY, Wilson JH, et al. R loops stimulate genetic instability of CTG.CAG repeats. Proc Natl Acad Sci USA. 2010a; 107:692–7. [PubMed: 20080737]

Lin Y, Dion V, Wilson JH. Transcription promotes contraction of CAG repeat tracts in human cells. Nat Struct Mol Biol. 2006; 13:179–80. [PubMed: 16388310]

Lin Y, Hubert L Jr, Wilson JH. Transcription destabilizes triplet repeats. Mol Carcinog. 2009; 48:350–61. [PubMed: 18973172]

Lin Y, Leng M, Wan M, Wilson JH. Convergent transcription through a long CAG tract destabilizes repeats and induces apoptosis. Mol Cell Biol. 2010b; 18:4435–51. [PubMed: 20647539]

Lin Y, Wilson JH. Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. Mol Cell Biol. 2007; 27:6209–17. [PubMed: 17591697]

Lin Y, Wilson JH. Nucleotide excision repair, mismatch repair, and R-loops modulate convergent transcription-induced cell death and repeat instability. PLoS One. 2012; 7:e46807. [PubMed: 23056461]

Liquori CL, Ikeda Y, Weatherspoon M, et al. Myotonic dystrophy type 2: human founder haplotype and evolutionary conservation of the repeat tract. Am J Hum Genet. 2003; 73:849–62. [PubMed: 14505273]

Liu G, Chen X, Bissler JJ, et al. Replication-dependent instability at (CTG) × (CAG) repeat hairpins in human cells. Nat Chem Biol. 2010; 6:652–9. [PubMed: 20676085]

Liu G, Leffak M. Instability of (CTG)n*(CAG)n trinucleotide repeats and DNA synthesis. Cell Biosci. 2012; 2:7. [PubMed: 22369689]

Liu Y, Bambara RA. Analysis of human flap endonuclease 1 mutants reveals a mechanism to prevent triplet repeat expansion. J Biol Chem. 2003; 278:13728–39. [PubMed: 12554738]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Liu Y, Beard WA, Shock DD, et al. DNA polymerase beta and flap endonuclease 1 enzymatic specificities sustain DNA synthesis for long patch base excision repair. J Biol Chem. 2005; 280:3665–74. [PubMed: 15561706]

Liu Y, Prasad R, Beard WA, et al. Coordination between Pol beta and FEN1 can modulate CAG repeat expansion. J Biol Chem. 2009; 284:28352–66. [PubMed: 19674974]

Liu Y, Wilson SH. DNA base excision repair: a mechanism of trinucleotide repeat expansion. Trends Biochem Sci. 2012; 37:162–72. [PubMed: 22285516]

Liu Y, Zhang H, Veeraraghavan J, et al. *Saccharomyces cerevisiae* flap endonuclease 1 uses flap equilibration to maintain triplet repeat stability. Mol Cell Biol. 2004; 24:4049–64. [PubMed: 15082797]

Lobachev KS, Gordenin DA, Resnick MA. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. Cell. 2002; 108:183–93. [PubMed: 11832209]

Loesch D, Hagerman R. Unstable mutations in the FMR1 gene and the phenotypes. Adv Exp Med Biol. 2012; 769:78–114. [PubMed: 23560306]

Lokanga RA, Entezam A, Kumari D, et al. Somatic expansion in mouse and human carriers of Fragile X premutation alleles. Hum Mutat. 2013; 34:157–66. [PubMed: 22887750]

Lokanga RA, Zhao XN, Entezam A, Usdin K. X inactivation plays a major role in the gender bias in somatic expansion in a mouse model of the Fragile X-related disorders: implications for the mechanism of repeat expansion. Hum Mol Genet. 2014a; 23:4985–94. [PubMed: 24858908]

Lokanga RA, Zhao X-N, Usdin K. The mismatch repair protein, MSH2, is rate-limiting for repeat expansion in a Fragile X premutation mouse model. Hum Mutat. 2014b; 35:129–36. [PubMed: 24130133]

Loomis EW, Sanz LA, Chedin F, Hagerman PJ. Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. PLoS Genet. 2014; 10:e1004294. [PubMed: 24743386]

Lopez Castel A, Cleary JD, Pearson CE. Repeat instability as the basis for human diseases and as a potential target for therapy. Nat Rev Mol Cell Biol. 2010; 11:165–70. [PubMed: 20177394]

Lopez Castel A, Nakamori M, Tome S, et al. Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. Hum Mol Genet. 2011; 20:1–15. [PubMed: 21044947]

Lou H, Komata M, Katou Y, et al. Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. Mol Cell. 2008; 32:106–17. [PubMed: 18851837]

MacDougall CA, Byun TS, Van C, et al. The structural determinants of checkpoint activation. Genes Dev. 2007; 21:898–903. [PubMed: 17437996]

Mangiarini L, Sathasivam K, Mahal A, et al. Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. Nat Genet. 1997; 15:197–200. [PubMed: 9020849]

Manley K, Shirley TL, Flaherty L, Messer A. Msh2 deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. Nat Genet. 1999; 23:471–3. [PubMed: 10581038]

Mariappan SV, Silks LA 3rd, Chen X, et al. Solution structures of the Huntington's disease DNA triplets, (CAG)n. J Biomol Struct Dyn. 1998; 15:723–44. [PubMed: 9514249]

Martins S, Coutinho P, Silveira I, et al. Cis-acting factors promoting the CAG intergenerational instability in Machado–Joseph disease. Am J Med Genet B Neuropsychiatr Genet: Official Publ Int Soc Psychiatric Genet. 2008; 147b:439–46.

Martins S, Pearson CE, Coutinho P, et al. Modifiers of (CAG)n instability in Machado–Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. Hum Genet. 2014; 133:1311–18. [PubMed: 25026993]

Mason AG, Tome S, Simard JP, et al. Expression levels of DNA replication and repair genes predict regional somatic repeat instability in the brain but are not altered by polyglutamine disease protein expression or age. Hum Mol Genet. 2014; 23:1606–18. [PubMed: 24191263]

Matsuura T, Yamagata T, Burgess DL, et al. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet. 2000; 26:191–4. [PubMed: 11017075]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Maurer DJ, O'Callaghan BL, Livingston DM. Orientation dependence of trinucleotide CAG repeat instability in *Saccharomyces cerevisiae*. Mol Cell Biol. 1996; 16:6617–22. [PubMed: 8943315]

McMurray CT. Hijacking of the mismatch repair system to cause CAG expansion and cell death in neurodegenerative disease. DNA Repair (Amst). 2008; 7:1121–34. [PubMed: 18472310]

McMurray CT. Mechanisms of trinucleotide repeat instability during human development. Nat Rev Genet. 2010; 11:786–99. [PubMed: 20953213]

McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. 2008; 24:529–38. [PubMed: 18809224]

Mimitou EP, Symington LS. DNA end resection – unraveling the tail. DNA Repair (Amst). 2011; 10:344–8. [PubMed: 21227759]

Minca EC, Kowalski D. Multiple Rad5 activities mediate sister chromatid recombination to bypass DNA damage at stalled replication forks. Mol Cell. 2010; 38:649–61. [PubMed: 20541998]

Miret JJ, Pessoa-Brandao L, Lahue RS. Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA. 1998; 95:12438–43. [PubMed: 9770504]

Mirkin EV, Mirkin SM. To switch or not to switch: at the origin of repeat expansion disease. Mol Cell. 2014; 53:1–3. [PubMed: 24411078]

Mirkin SM. DNA structures, repeat expansions and human hereditary disorders. Curr Opin Struct Biol. 2006; 16:351–8. [PubMed: 16713248]

Mirkin SM, Smirnova EV. Positioned to expand. Nat Genet. 2002; 31:5–6. [PubMed: 11984556]

Mitas M, Yu A, Dill J, Haworth IS. The trinucleotide repeat sequence d(CGG)15 forms a heat-stable hairpin containing Gsyn. Ganti base pairs. Biochemistry. 1995a; 34:12803–11. [PubMed: 7548035]

Mitas M, Yu A, Dill J, et al. Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)15. Nucleic Acids Res. 1995b; 23:1050–9. [PubMed: 7731793]

Mittelman D, Moye C, Morton J, et al. Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. Proc Natl Acad Sci USA. 2009; 106:9607–12. [PubMed: 19482946]

Modrich P. Mechanisms in eukaryotic mismatch repair. J Biol Chem. 2006; 281:30305–9. [PubMed: 16905530]

Moldovan GL, Dejsuphong D, Petalcorin MI, et al. Inhibition of homologous recombination by the PCNA-interacting protein PARI. Mol Cell. 2012; 45:75–86. [PubMed: 22153967]

Mollersen L, Rowe AD, Illuzzi JL, et al. Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice. Hum Mol Genet. 2012; 21:4939–47. [PubMed: 22914735]

Monckton DG, Caskey CT. Unstable triplet repeat diseases. Circulation. 1995; 91:513–20. [PubMed: 7805257]

Morales F, Couto JM, Higham CF, et al. Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. Hum Mol Genet. 2012; 21:3558–67. [PubMed: 22595968]

Morales F, Vasquez M, Cuenca P, et al. Parental age effects, but no evidence for an intrauterine effect in the transmission of myotonic dystrophy type 1. Eur J Hum Genet. 2014 [Epub ahead of print]. doi: 10.1038/ejhg.2014.138.

Moseley ML, Zu T, Ikeda Y, et al. Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nat Genet. 2006; 38:758–69. [PubMed: 16804541]

Mott C, Symington LS. RAD51-independent inverted-repeat recombination by a strand-annealing mechanism. DNA Repair (Amst). 2011; 10:408–15. [PubMed: 21317047]

Mozlin AM, Fung CW, Symington LS. Role of the *Saccharomyces cerevisiae* Rad51 paralogs in sister chromatid recombination. Genetics. 2008; 178:113–26. [PubMed: 18202362]

Muftuoglu M, de Souza-Pinto NC, Dogan A, et al. Cockayne syndrome group B protein stimulates repair of formamidopyrimidines by NEIL1 DNA glycosylase. J Biol Chem. 2009; 284:9270–9. [PubMed: 19179336]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Muftuoglu M, Sharma S, Thorslund T, et al. Cockayne syndrome group B protein has novel strand annealing and exchange activities. Nucleic Acids Res. 2006; 34:295–304. [PubMed: 16410611]

Mulvihill DJ, Nichol Edamura K, Hagerman KA, et al. 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. 2005; 280:4498–503. [PubMed: 15574425]

Murray A, Ennis S, Youings SA, et al. Stability and haplotype analysis of the FRAXE region. Eur J Hum Genet. 2000; 8:583–9. [PubMed: 10951520]

Nadel Y, Weisman-Shomer P, Fry M. The fragile X syndrome single strand d(CGG)n nucleotide repeats readily fold back to form unimolecular hairpin structures. J Biol Chem. 1995; 270:28970–7. [PubMed: 7499428]

Nag DK, Suri M, Stenson EK. Both CAG repeats and inverted DNA repeats stimulate spontaneous unequal sister-chromatid exchange in *Saccharomyces cerevisiae*. Nucleic Acids Res. 2004; 32:5677–84. [PubMed: 15494455]

Nakamori M, Pearson CE, Thornton CA. Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. Hum Mol Genet. 2011; 20:580–8. [PubMed: 21088112]

Napierala M, Parniewski P, Pluciennik A, Wells RD. Long CTG.CAG repeat sequences markedly stimulate intramolecular recombination. J Biol Chem. 2002; 277:34087–100. [PubMed: 12045198]

Naylor ML, Li JM, Osborn AJ, Elledge SJ. Mrc1 phosphorylation in response to DNA replication stress is required for Mec1 accumulation at the stalled fork. Proc Natl Acad Sci USA. 2009; 106:12765–70. [PubMed: 19515819]

Nestor CE, Monckton DG. Correlation of inter-locus polyglutamine toxicity with CAG*CTG triplet repeat expandability and flanking genomic DNA GC content. PLoS One. 2011; 6:e28260. [PubMed: 22163004]

Neville CE, Mahadevan MS, Barcelo JM, Korneluk RG. High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. Hum Mol Genet. 1994; 3:45–51. [PubMed: 7909252]

Nolin SL, Glicksman A, Ding X, et al. Fragile X analysis of 1112 prenatal samples from 1991 to 2010. Prenat Diagn. 2011; 31:925–31. [PubMed: 21717484]

Nolin SL, Sah S, Glicksman A, et al. Fragile X AGG analysis provides new risk predictions for 45-69 repeat alleles. Am J Med Genet A. 2013; 161A:771–8. [PubMed: 23444167]

Northam MR, Moore EA, Mertz TM, et al. DNA polymerases zeta and Rev1 mediate error-prone bypass of non-B DNA structures. Nucleic Acids Res. 2014; 42:290–306. [PubMed: 24049079]

Oberle I, Rousseau F, Heitz D, et al. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science. 1991; 252:1097–102. [PubMed: 2031184]

Owen BA, Lang WH, McMurray CT. The nucleotide binding dynamics of human MSH2–MSH3 are lesion dependent. Nat Struct Mol Biol. 2009; 16:550–7. [PubMed: 19377479]

Owen BA, Yang Z, Lai M, et al. (CAG)(n)-hairpin DNA binds to Msh2–Msh3 and changes properties of mismatch recognition. Nat Struct Mol Biol. 2005; 12:663–70. [PubMed: 16025128]

Panigrahi GB, Cleary JD, Pearson CE. *In vitro* (CTG)*(CAG) expansions and deletions by human cell extracts. J Biol Chem. 2002; 277:13926–34. [PubMed: 11832482]

Panigrahi GB, Lau R, Montgomery SE, et al. Slipped (CTG)*(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. Nat Struct Mol Biol. 2005; 12:654–62. [PubMed: 16025129]

Panigrahi GB, Slean MM, Simard JP, et al. Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired. Proc Natl Acad Sci USA. 2010; 107:12593–8. [PubMed: 20571119]

Pearson CE, Edamura KN, Cleary JD. Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet. 2005; 6:729–42. [PubMed: 16205713]

Pearson CE, Wang YH, Griffith JD, Sinden RR. Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG) n.(CAG) n repeats from the myotonic dystrophy locus. Nucleic Acids Res. 1998; 26:816–23. [PubMed: 9443975]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Pelletier R, Krasilnikova MM, Samadashwily GM, et al. Replication and expansion of trinucleotide repeats in yeast. Mol Cell Biol. 2003; 23:1349–57. [PubMed: 12556494]

Pinto RM, Dragileva E, Kirby A, et al. Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. PLoS Genet. 2013; 9:e1003930. [PubMed: 24204323]

Pluciennik A, Burdett V, Baitinger C, et al. Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLalpha endonuclease activation. Proc Natl Acad Sci USA. 2013; 110:12277–82. [PubMed: 23840062]

Pluciennik A, Iyer RR, Napierala M, et al. Long CTG.CAG repeats from myotonic dystrophy are preferred sites for intermolecular recombination. J Biol Chem. 2002; 277:34074–86. [PubMed: 12087090]

Potaman VN, Bissler JJ, Hashem VI, et al. Unpaired structures in SCA10 (ATTCT)n.(AGAAT)n repeats. J Mol Biol. 2003; 326:1095–111. [PubMed: 12589756]

Potaman VN, Oussatcheva EA, Lyubchenko YL, et al. Length-dependent structure formation in Friedreich ataxia (GAA)n*(TTC)n repeats at neutral pH. Nucleic Acids Res. 2004; 32:1224–31. [PubMed: 14978261]

Prado F, Aguilera A. Role of reciprocal exchange, one-ended invasion crossover and single-strand annealing on inverted and direct repeat recombination in yeast: different requirements for the RAD1, RAD10, and RAD52 genes. Genetics. 1995; 139:109–23. [PubMed: 7705617]

Ranjha L, Anand R, Cejka P. The *Saccharomyces cerevisiae* Mlh1-Mlh3 heterodimer is an endonuclease that preferentially binds to Holliday junctions. J Biol Chem. 2014; 289:5674–86. [PubMed: 24443562]

Razidlo DF, Lahue RS. Mrc1, Tof1 and Csm3 inhibit CAG.CTG repeat instability by at least two mechanisms. DNA Repair (Amst). 2008; 7:633–40. [PubMed: 18321795]

Reddy K, Schmidt MH, Geist JM, et al. Processing of double-R-loops in (CAG).(CTG) and C9orf72 (GGGGCC).(GGCCCC) repeats causes instability. Nucleic Acids Res. 2014; 42:10473–87. [PubMed: 25147206]

Reddy K, Tam M, Bowater RP, et al. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. Nucleic Acids Res. 2011; 39:1749–62. [PubMed: 21051337]

Renton AE, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011; 72:257–68. [PubMed: 21944779]

Richard GF, Dujon B, Haber JE. Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats. Mol Gen Genet. 1999; 261:871–82. [PubMed: 10394925]

Richard GF, Goellner GM, McMurray CT, Haber JE. Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11-RAD50-XRS2 complex. EMBO J. 2000; 19:2381–90. [PubMed: 10811629]

Richard GF, Viterbo D, Khanna V, et al. Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. PLoS One. 2014; 9:e95611. [PubMed: 24748175]

Richards RI, Holman K, Friend K, et al. Evidence of founder chromosomes in fragile X syndrome. Nat Genet. 1992; 1:257–60. [PubMed: 1302021]

Richards RI, Sutherland GR. Simple repeat DNA is not replicated simply. Nat Genet. 1994; 6:114–16. [PubMed: 8162063]

Roesner LM, Mielke C, Fahnrich S, et al. Stable expression of MutLgamma in human cells reveals no specific response to mismatched DNA, but distinct recruitment to damage sites. J Cell Biochem. 2013; 114:2405–14. [PubMed: 23696135]

Rolfsmeier ML, Dixon MJ, Pessoa-Brandao L, et al. Cis-elements governing trinucleotide repeat instability in *Saccharomyces cerevisiae*. Genetics. 2001; 157:1569–79. [PubMed: 11290713]

Saha T, Usdin K. Tetraplex formation by the progressive myoclonus epilepsy type-1 repeat: implications for instability in the repeat expansion diseases. FEBS Lett. 2001; 491:184–7. [PubMed: 11240124]

Saini N, Zhang Y, Nishida Y, et al. Fragile DNA motifs trigger mutagenesis at distant chromosomal loci in *Saccharomyces cerevisiae*. PLoS Genet. 2013; 9:e1003551. [PubMed: 23785298]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Sakamoto N, Larson JE, Iyer RR, et al. GGA*TCC-interrupted triplets in long GAA*TTC repeats inhibit the formation of triplex and sticky DNA structures, alleviate transcription inhibition, and reduce genetic instabilities. J Biol Chem. 2001; 276:27178–87. [PubMed: 11325966]

Salinas-Rios V, Belotserkovskii BP, Hanawalt PC. DNA slip-outs cause RNA polymerase II arrest *in vitro*: potential implications for genetic instability. Nucleic Acids Res. 2011; 39:7444–54. [PubMed: 21666257]

Saugar I, Ortiz-Bazan MA, Tercero JA. Tolerating DNA damage during eukaryotic chromosome replication. Exp Cell Res. 2014; 329:170–7. [PubMed: 25038291]

Savouret C, Brisson E, Essers J, et al. CTG repeat instability and size variation timing in DNA repair-deficient mice. EMBO J. 2003; 22:2264–73. [PubMed: 12727892]

Savouret C, Garcia-Cordier C, Megret J, et al. MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 Transgenic Mice. Mol Cell Biol. 2004; 24:629–37. [PubMed: 14701736]

Schweitzer JK, Livingston DM. Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. Hum Mol Genet. 1997; 6:349–55. [PubMed: 9147637]

Schweitzer JK, Livingston DM. The effect of DNA replication mutations on CAG tract stability in yeast. Genetics. 1999; 152:953–63. [PubMed: 10388815]

Schweitzer JK, Reinke SS, Livingston DM. Meiotic alterations in CAG repeat tracts. Genetics. 2001; 159:1861–5. [PubMed: 11779820]

Seriola A, Spits C, Simard JP, et al. Huntington's and myotonic dystrophy hESCs: down-regulated trinucleotide repeat instability and mismatch repair machinery expression upon differentiation. Hum Mol Genet. 2011; 20:176–85. [PubMed: 20935170]

Shah KA, McGinty RJ, Egorova VI, Mirkin SM. Coupling transcriptional state to large-scale repeat expansions in yeast. Cell Rep. 2014; 9:1594–602. [PubMed: 25464841]

Shah KA, Shishkin AA, Voineagu I, et al. Role of DNA polymerases in repeat-mediated genome instability. Cell Rep. 2012; 2:1088–95. [PubMed: 23142667]

Shen J, Loeb LA. Unwinding the molecular basis of the Werner syndrome. Mech Ageing Dev. 2001; 122:921–44. [PubMed: 11348659]

Shishkin AA, Voineagu I, Matera R, et al. Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol Cell. 2009; 35:82–92. [PubMed: 19595718]

Shrivastav M, De Haro LP, Nickoloff JA. Regulation of DNA double-strand break repair pathway choice. Cell Res. 2008; 18:134–47. [PubMed: 18157161]

Simard O, Gregoire MC, Arguin M, et al. Instability of trinucleotidic repeats during chromatin remodeling in spermatids. Hum Mutat. 2014; 35:1280–4. [PubMed: 25136821]

Sopher BL, Ladd PD, Pineda VV, et al. CTCF regulates ataxin-7 expression through promotion of a convergently transcribed, antisense noncoding RNA. Neuron. 2011; 70:1071–84. [PubMed: 21689595]

Spiro C, McMurray CT. Nuclease-deficient FEN-1 blocks Rad51/BRCA1-mediated repair and causes trinucleotide repeat instability. Mol Cell Biol. 2003; 23:6063–74. [PubMed: 12917330]

Spiro C, Pelletier R, Rolfsmeier ML, et al. Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. Mol Cell. 1999; 4:1079–85. [PubMed: 10635332]

Staresincic L, Fagbemi AF, Enzlin JH, et al. Coordination of dual incision and repair synthesis in human nucleotide excision repair. EMBO J. 2009; 28:1111–20. [PubMed: 19279666]

Subramanian PS, Nelson DL, Chinault AC. Large domains of apparent delayed replication timing associated with triplet repeat expansion at FRAXA and FRAXE. Am J Hum Genet. 1996; 59:407–16. [PubMed: 8755928]

Sun H, Karow JK, Hickson ID, Maizels N. The Bloom's syndrome helicase unwinds G4 DNA. J Biol Chem. 1998; 273:27587–92. [PubMed: 9765292]

Sundararajan R, Freudenreich CH. Expanded CAG/CTG repeat DNA induces a checkpoint response that impacts cell proliferation in *Saccharomyces cerevisiae*. PLoS Genet. 2011; 7:e1001339. [PubMed: 21437275]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Sundararajan R, Gellon L, Zunder RM, Freudenreich CH. Double-strand break repair pathways protect against CAG/CTG repeat expansions, contractions and repeat-mediated chromosomal fragility in *Saccharomyces cerevisiae*. Genetics. 2010; 184:65–77. [PubMed: 19901069]

Suwaki N, Klare K, Tarsounas M. RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis. Semin Cell Dev Biol. 2011; 22:898–905. [PubMed: 21821141]

Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. Annu Rev Genet. 2011; 45:247–71. [PubMed: 21910633]

Takiyama Y, Igarashi S, Rogaeva EA, et al. Evidence for inter-generational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. Hum Mol Genet. 1995; 4:1137–46. [PubMed: 8528200]

Tang W, Dominska M, Greenwell PW, et al. Friedreich's ataxia (GAA)n*(TTC)n repeats strongly stimulate mitotic crossovers in *Saccharomyces cerevisae*. PLoS Genet. 2011; 7:e1001270. [PubMed: 21249181]

Tang W, Dominska M, Gawel M, et al. Genomic deletions and point mutations induced in *Saccharomyces cerevisae* by the trinucleotide repeats (GAA.TTC) associated with Friedreich's ataxia. DNA Repair. 2013; 12:10–17. [PubMed: 23182423]

Tian L, Hou C, Tian K, et al. Mismatch recognition protein MutSbeta does not hijack (CAG)n hairpin repair *in vitro*. J Biol Chem. 2009; 284:20452–6. [PubMed: 19525234]

Tittel-Elmer M, Alabert C, Pasero P, Cobb JA. The MRX complex stabilizes the replisome independently of the S phase checkpoint during replication stress. EMBO J. 2009; 28:1142–56. [PubMed: 19279665]

Tome S, Holt I, Edelmann W, et al. MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice. PLoS Genet. 2009; 5:e1000482. doi:10.1371/journal.pgen. 1000482. [PubMed: 19436705]

Tome S, Manley K, Simard JP, et al. MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. PLoS Genet. 2013; 9:e1003280. [PubMed: 23468640]

Tourriere H, Versini G, Cordon-Preciado V, et al. Mrc1 and Tof1 promote replication fork progression and recovery independently of Rad53. Mol Cell. 2005; 19:699–706. [PubMed: 16137625]

Tuo J, Chen C, Zeng X, et al. Functional crosstalk between hOgg1 and the helicase domain of Cockayne syndrome group B protein. DNA Repair (Amst). 2002; 1:913–27. [PubMed: 12531019]

Usdin K, Woodford KJ. CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis *in vitro*. Nucleic Acids Res. 1995; 23:4202–9. [PubMed: 7479085]

van den Broek WJ, Nelen MR, van der Heijden GW, et al. Fen1 does not control somatic hypermutability of the (CTG)(n)*(CAG)(n) repeat in a knock-in mouse model for DM1. FEBS Lett. 2006; 580:5208–14. [PubMed: 16978612]

van Veelen LR, Essers J, van de Rakt MW, et al. Ionizing radiation-induced foci formation of mammalian Rad51 and Rad54 depends on the Rad51 paralogs, but not on Rad52. Mutat Res. 2005; 574:34–49. [PubMed: 15914205]

Vannier JB, Pavicic-Kaltenbrunner V, Petalcorin MI, et al. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. Cell. 2012; 149:795–806. [PubMed: 22579284]

Vannier JB, Sandhu S, Petalcorin MI, et al. RTEL1 is a replisome-associated helicase that promotes telomere and genome-wide replication. Science. 2013; 342:239–42. [PubMed: 24115439]

Vanoli F, Fumasoni M, Szakal B, et al. Replication and recombination factors contributing to recombination-dependent bypass of DNA lesions by template switch. PLoS Genet. 2010; 6:e1001205. [PubMed: 21085632]

Verkerk AJ, Pieretti M, Sutcliffe JS, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell. 1991; 65:905–14. [PubMed: 1710175]

Voineagu I, Freudenreich CH, Mirkin SM. Checkpoint responses to unusual structures formed by DNA repeats. Mol Carcinog. 2009a; 48:309–18. [PubMed: 19306277]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Voineagu I, Surka CF, Shishkin AA, et al. Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. Nat Struct Mol Biol. 2009b; 16:226–8. [PubMed: 19136957]

Wang G, Christensen LA, Vasquez KM. Z-DNA-forming sequences generate large-scale deletions in mammalian cells. Proc Natl Acad Sci USA. 2006; 103:2677–82. [PubMed: 16473937]

Wang G, Vasquez KM. Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. Proc Natl Acad Sci USA. 2004; 101:13448–53. [PubMed: 15342911]

Wang G, Vasquez KM. Models for chromosomal replication-independent non-B DNA structure-induced genetic instability. Mol Carcinog. 2009; 48:286–98. [PubMed: 19123200]

Warby SC, Montpetit A, Hayden AR, et al. CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. Am J Hum Genet. 2009; 84:351–66. [PubMed: 19249009]

Watanabe H, Tanaka F, Doyu M, 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. 2000; 107:452–7. [PubMed: 11140942]

Wenger SL, Giangreco CA, Tarleton J, Wessel HB. Inability to induce fragile sites at CTG repeats in congenital myotonic dystrophy. Am J Hum Genet. 1996; 66:60–3.

Wheeler VC, Lebel LA, Vrbanac V, et al. Mismatch repair gene Msh2 modifies the timing of early disease in hdh(Q111) striatum. Human Mol Genet. 2003; 12:273–81. [PubMed: 12554681]

Williams RS, Williams JS, Tainer JA. Mre11-Rad50-Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin template. Biochem Cell Biol. 2007; 85:509–20. [PubMed: 17713585]

Wohrle D, Salat U, Hameister H, et al. Demethylation, reactivation, and destabilization of human fragile X full-mutation alleles in mouse embryocarcinoma cells. Am J Hum Genet. 2001; 69:504–15. [PubMed: 11462172]

Xu M, Gabison J, Liu Y. Trinucleotide repeat deletion via a unique hairpin bypass by DNA polymerase beta and alternate flap cleavage by flap endonuclease 1. Nucleic Acids Res. 2013; 41:1684–97. [PubMed: 23258707]

Xu M, Lai Y, Torner J, et al. Base excision repair of oxidative DNA damage coupled with removal of a CAG repeat hairpin attenuates trinucleotide repeat expansion. Nucleic Acids Res. 2014; 42:3675–91. [PubMed: 24423876]

Yang J, Freudenreich CH. Haploin sufficiency of yeast FEN1 causes instability of expanded CAG/CTG tracts in a length-dependent manner. Gene. 2007; 393:110–15. [PubMed: 17383831]

Yang JH, Freudenreich CH. The Rtt109 histone acetyltransferase facilitates error-free replication to prevent CAG/CTG repeat contractions. DNA Repair (Amst). 2010; 9:414–20. [PubMed: 20083442]

Yoon SR, Dubeau L, de Young M, et al. Huntington disease expansion mutations in humans can occur before meiosis is completed. Proc Natl Acad Sci USA. 2003; 100:8834–8. [PubMed: 12857955]

Yrigollen CM, Durbin-Johnson B, Gane L, et al. AGG interruptions within the maternal FMR1 gene reduce the risk of offspring with fragile X syndrome. Genet Med. 2012; 14:729–36. [PubMed: 22498846]

Yrigollen CM, Martorell L, Durbin-Johnson B, et al. AGG interruptions and maternal age affect FMR1 CGG repeat allele stability during transmission. J Neurodev Disord. 2014; 6:24. [PubMed: 25110527]

Yu A, Barron MD, Romero RM, et al. At physiological pH, d(CCG)15 forms a hairpin containing protonated cytosines and a distorted helix. Biochemistry. 1997; 36:3687–99. [PubMed: 9132022]

Yu S, Pritchard M, Kremer E, et al. Fragile X genotype characterized by an unstable region of DNA. Science. 1991; 252:1179–81. [PubMed: 2031189]

Yudkin D, Hayward BE, Aladjem MI, et al. Chromosome fragility and the abnormal replication of the FMR1 locus in Fragile X syndrome. Hum Mol Genet. 2014; 23:2940–52. [PubMed: 24419320]

Zatz M, Passos-Bueno MR, Cerqueira A, 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. 1995; 4:401–6. [PubMed: 7795594]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Zegerman P, Diffley JF. DNA replication as a target of the DNA damage checkpoint. DNA Repair (Amst). 2009; 8:1077–88. [PubMed: 19505853]

Zhang H, Lawrence CW. The error-free component of the RAD6/RAD18 DNA damage tolerance pathway of budding yeast employs sister-strand recombination. Proc Natl Acad Sci USA. 2005; 102:15954–9. [PubMed: 16247017]

Zhang T, Huang J, Gu L, Li GM. *In vitro* repair of DNA hairpins containing various numbers of CAG/CTG trinucleotide repeats. DNA Repair (Amst). 2012a; 11:201–9. [PubMed: 22041023]

Zhang Y, Shishkin AA, Nishida Y, et al. Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. Mol Cell. 2012b; 48:254–65. [PubMed: 22959270]

Zhao X-N, Usdin K. Gender and cell-type specific effects of the transcription coupled repair protein, ERCC6/CSB, on repeat expansion in a mouse model of the Fragile X-related disorders. Hum Mutat. 2014; 35:341–9. [PubMed: 24352881]

Zlatanou A, Despras E, Braz-Petta T, et al. The hMsh2-hMsh6 complex acts in concert with monoubiquitinated PCNA and Pol eta in response to oxidative DNA damage in human cells. Mol Cell. 2011; 43:649–62. [PubMed: 21855803]

Zu T, Liu Y, Banez-Coronel M, et al. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc Natl Acad Sci USA. 2013; 110:E4968–77. [PubMed: 24248382]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

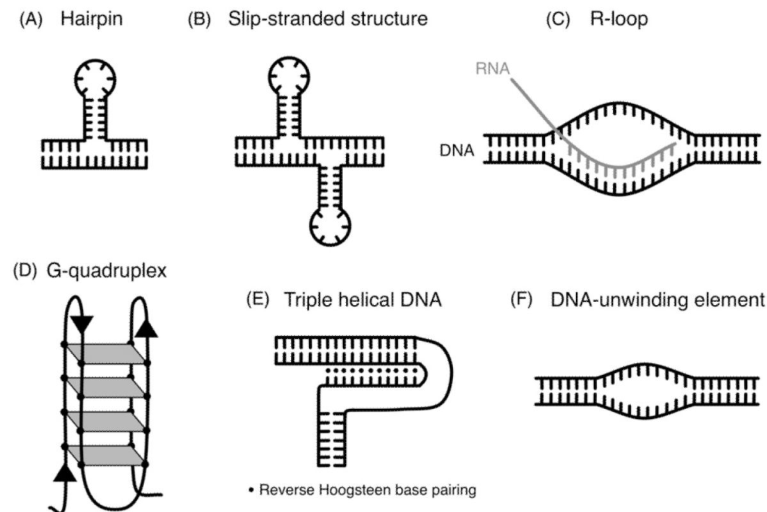

Figure 1.

Stable non-B form DNA structures formed by expandable repeats. (A) Hairpins can be formed by G–C-rich expanding triplet repeats with stability: CGG>CCG>CTG>CAG. (B) A slipped-strand structure occurs when hairpins form simultaneously on both strands. (C) RNA/DNA hybrids (R-loops) occur during transcription when the newly synthesized RNA transcript stably pairs with single-stranded DNA in the transcription bubble. Purine-rich repeats are especially prone to forming persistent R-loops. (D) G-quadruplex structures can be formed by CGG repeats and G_4C_2 repeats. G-quadruplexes can be parallel (not shown) or anti-parallel (shown). (E) Triple helical DNA structures form at GAA/TCC repeats and can be either the purine:purine:pyrimidine triplexes or pyrimidine:purine pyrimidine triplexes. (F) AT-rich ATTCT/AGAAT repeats are DNA-unwinding elements, melting the double helix to form a region of unpaired DNA.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

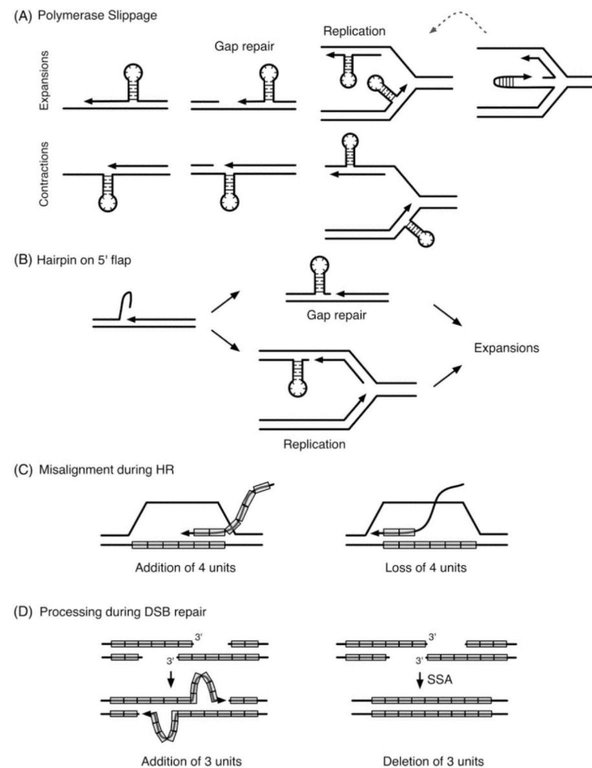

Figure 2.

General mechanisms of trinucleotide repeat instability. (A) Polymerase slippage can lead to formation of secondary structures on the extending nascent strand (expansion) or template strand (contraction). Slippage can also occur at replication barriers, and structures can form on the nascent stands or template strands at the stalled replication fork. Reversed forks can be processed into a hairpin on the leading strand, causing an expansion. (B) Hairpins or quadruplexes caused by unprocessed 5′ flaps can form during gap repair or on an Okazaki fragment, leading to expansions. (C) Misalignment of repeat units (gray blocks) during homologous recombination can lead to the addition of repeat units (expansion event) or loss of repeat units (contraction event). (D) Resection during double-strand break repair exposes repeat units that can misalign and anneal, leading to repeat unit deletion (right). Misalignment followed by slippage during gap filling could lead to an expansion (left).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

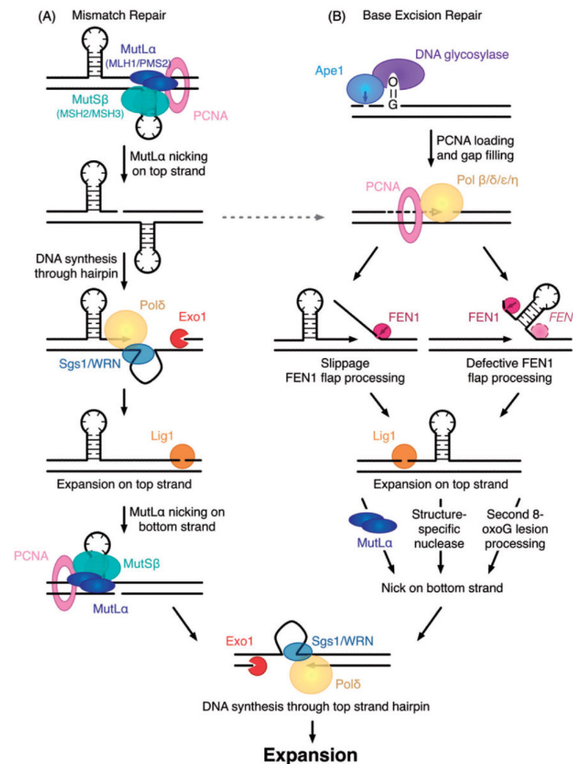

Figure 3.

Model for generation of repeat expansions during MMR and BER. (A) Mismatch repair (MMR) is initiated when MutSβ (a heterodimer of MSH2/MSH3) binds to and stabilizes the hairpin formed by the repetitive DNA element. PCNA loading (facilitated by RFC, not shown) also occurs at the extrusion. The interaction between MutSβ and PCNA activates the cryptic endonuclease activity of the MutLα complex (a heterodimer of MLH1/PMS2). MutLα can nick either strand; here, a nick on the top strand is shown. Exo1 exonucleolytic activity digests the end to form a single-stranded gap that extends ~150 nt past the hairpin. The hairpin is unwound by helicase activity, potentially Sgs1/WRN (see section ‘‘The role of helicases in resolving hairpins during repair and fork restart’’), and Polδ fills the gap. The nick is sealed by Lig1. These steps are repeated on the other strand to resolve the second hairpin; DNA synthesis through the hairpin at this stage leads to an expansion on both strands. Note that after nicking, repair could also proceed by strand displacement (indicated by dotted gray arrow to part B), leading to an expansion by slippage during replication or inefficient flap cleavage by Fen1. (B) BER is initiated to repair 8-oxoG lesions caused by oxidative damage. 8-oxoG is first recognized by a DNA glycosylase (Ogg1 or NEIL1). ApeI creates a nick, and strand slippage during fill in by Polβ, Polδ, Polε, or Polη can lead to an expansion (left pathway). Alternatively, hairpins can impede Fen1 cleavage (indicated by dotted Fen1), leading to inefficient or ‘‘alternative’’ flap cleavage (indicated by solid Fen1) and expansions (right pathway). A nick on the complementary (bottom) strand followed by DNA synthesis through the hairpin leads to an expansion on both strands. Alternatively, direct nicking of the hairpin, or oxidative damage within the hairpin followed by nicking, could start the BER cycle again, leading to further expansions (toxic oxidation cycle, not shown). (see colour version of this figure at www.informahealthcare.com/bmg).

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

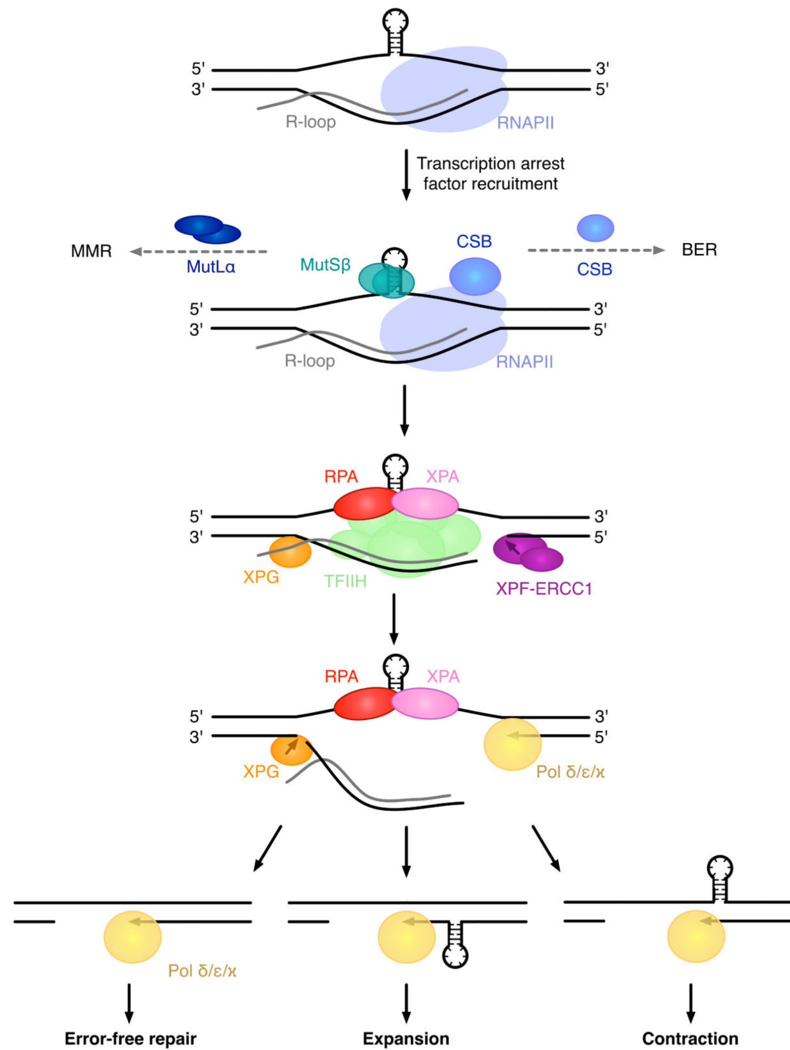

Figure 4.

Repeat instability through transcription-coupled repair. In this model, RNA polymerase stalls due to R loop formation and/or the formation of secondary structures on the non-template strand, which could be facilitated by an R-loop on the template strand. The folded structures could also facilitate the stall by sequestering the non-template DNA strand, thus decreasing its propensity to displace RNA (Belotserkovskii *et al.*, 2013). An alternate model (not shown) is that the RNAPII stalls at pre-formed template hairpins (e.g. see Lin *et al.*, 2009). Structures formed on the non-template strand during transcription could be stabilized by MutSβ binding, further increasing the strength of RNAPII stalling. Stalled transcription recruits the transcription arrest factors, including CSB and XPG, which initiate TCR. Alternatively, CSB may facilitate glycosylase activity to initiate BER, or the MutSβ complex could recruit MutLα to initiate MMR (left and right arrows). RNAPII is displaced and TFIIH is recruited; RPA and XPA stabilize the denatured bubble. The RNAPII-blocking lesion then undergoes dual incisions, the first carried out by XPF-ERCC1 that cleaves 5′ of the lesion. The second cleavage occurs downstream of the lesion and is carried out by XPG. The result would be a 25–30 nt gap that is filled by Polδ, Polε, and/or Polκ; repair

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

replication may begin before 3′ cleavage (Staresincic *et al*., 2009). If the polymerase fills the gap faithfully no tract length change occurs (left). If strand slippage occurs during DNA synthesis, this leads to an expansion (center). If the polymerase replicates over a template hairpin formed due to exposed ssDNA on the bottom strand or the hairpin is excised, this will result in a contraction. (see colour version of this figure at www.informahealthcare.com/bmg).

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

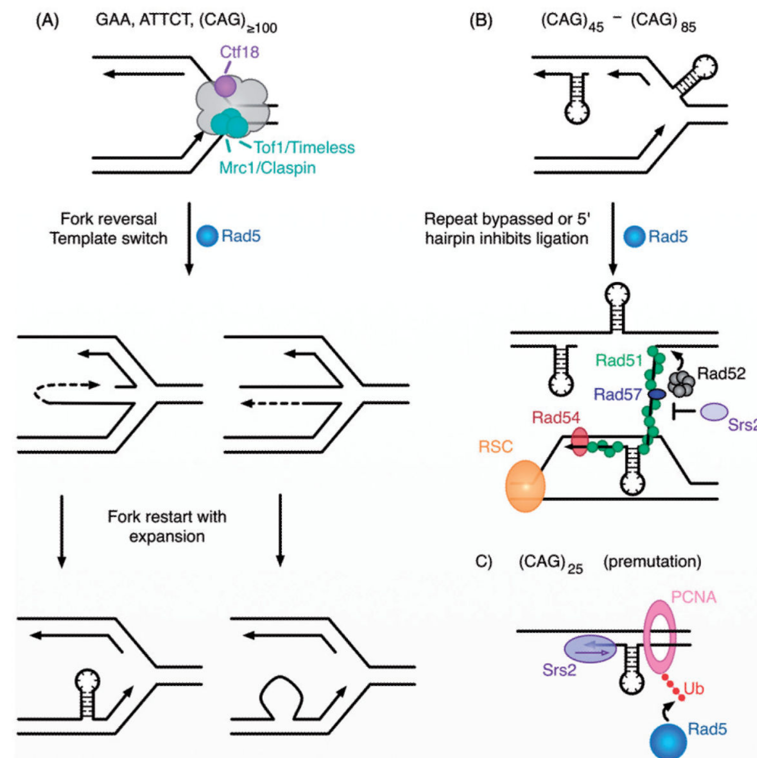

Figure 5.

Post-replicative repair at repetitive DNA elements. Repeat units indicated reflect experimental constructs, but the mechanisms described here likely work along a continuum based on repeat length. (A) Repeats that cause fork stalling or fork reversal can initiate Rad5-dependent template switching, leading to expansions. In *rad5* cells, these expansions are eliminated. In the absence of fork stabilization proteins (Mrc1, Tof1, and Ctf18) instability and fork breakage are both increased. (B) Hairpins at medium CAG repeat tracts are bypassed during replication and Rad5-dependent template switching is initiated at post-replication gaps. Rad5-, Rad51-, Rad52-, Rad57-, and Rad54-mediated expansions occur during sister chromatid recombination. Excessive recombination is inhibited by Srs2 anti-recombinase function. (C) At short repeat lengths, Rad5-dependent PRR and unwinding by the Srs2 helicase are sufficient to fill gaps and prevent expansions without sister chromatid recombination. (see colour version of this figure at www.informahealthcare.com/bmg).

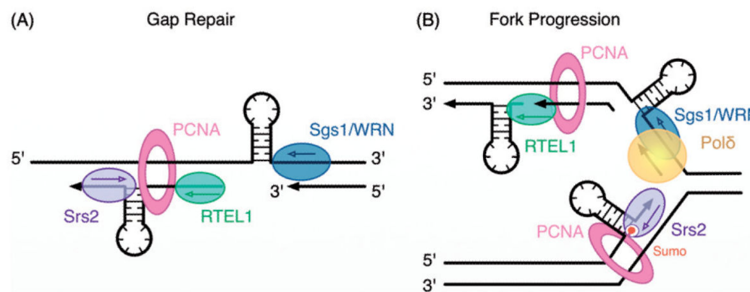

Figure 6.

Helicases resolve hairpins to promote repeat stability. (A) The helicases Srs2, RTEL1, Sgs1, and WRN unwind hairpins at gaps to prevent expansions of repeat DNA. Based on instability profiles in mutants, Sgs1/WRN are hypothesized to act on the template strand to prevent contractions, and Srs2/RTEL1 on the nascent strand to prevent expansions. (B) Helicases promote replication fork progression through hairpin unwinding. Sgs1/WRN and RTEL1 may also unwind G-quadruplex structures. The placement of Srs2 at the advancing fork is based on its ability to bind the sumoylated form of PCNA, found at the replication fork, and to lessen fork stalling at CGG repeats (Anand *et al*., 2012). (see colour version of this figure at www.informahealthcare.com/bmg).

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Characteristics of expandable repeats that cause disease.

Disease	Repeat location	Normal repeat length	Pathogenic repeat length
CAG expansion (hairpin, slip-stranded DNA, R-loop[*])			
Dentatorubral-pallidoluysian atrophy (DRPLA)	*ATN1* exon	7–34	29–88
Huntington disease (HD)	*HTT* exon	11–34	40–121
Spinal bulbar muscular atrophy (SBMA)	*AR* exon	9–36	38–62
Spinocerebellar ataxia 1 (SCA1)	*ATXN1* exon	6–39	40–82
Spinocerebellar ataxia 2 (SCA2)	*ATXN2* exon	15–36	32–200
Spinocerebellar ataxia 3 (SCA3)	*ATXN3* exon	13–36	61–84
Spinocerebellar ataxia 6 (SCA6)	*CACNA1A* exon	4–20	20–29
Spinocerebellar ataxia 7 (SCA7)	*ATXN7* exon	4–35	37–306
Spinocerebellar ataxia 12 (SCA12)	*PPP2R2B* 5′ UTR	7–45	55–78
Spinocerebellar ataxia 17 (SCA17)	*TBP* exon	25–42	47–63
CTG expansion (hairpin, slip-stranded DNA, R-loop)			
Myotonic dystrophy type 1 (DM1)	*DMPK* 3′ UTR	5–37	50– 1000
Huntington disease like 2 (HDL2)	*JPH3* 3′ UTR	7–28	66–78
Spinocerebellar ataxia 8 (SCA8)	*ATXN8* 3′ UTR	16–34	>74
CGG expansion (hairpin, slip-stranded DNA, G-quadruplex, R-loop)			
Fragile X syndrome (FXS)[†]	*FMR1* 5′ UTR	6–55[‡]	>200
Fragile X-associated primary ovarian insufficiency (FXPOI)[†]	*FMR1* 5′ UTR	6–55[‡]	55–200
Fragile X tremor/ataxia syndrome (FXTAS)[†]	*FMR1* 5′ UTR	6–55[‡]	55–200
FRAXE mental retardation syndrome (FRAXE MR)	*FMR2* 5′ UTR	4–36	200–900
GAA expansion (triplex DNA, R-loop)			
Friedrich ataxia (FRDA)	*FXN* intron	6–32	200–1700
CCTG expansion (hairpin, slip-stranded DNA)			
Myotonic dystrophy type 2 (DM2)	*ZNF9* intron	10–26	75–11 000
ATTCT expansion (DNA unwinding element)			
Spinocerebellar ataxia 10 (SCA1O)	*ATXN10* intron	10–20	500–4500
G_2C_2TG expansion (hairpin)			
Spinocerebellar ataxia 36 (SCA36)	*NOP56* intron	<8	1500–2500
G_4CC expansions (hairpin, G-quadruplex, R-loop)			
Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)	*C9orf72* intron	<23	250–1600
C_4GC_4GCG expansion (hairpin, G-quadruplex)			
Progressive myoclonic epilepsy 1 (EPM1)	*CSTB* promoter	2–3 12–17[¶]	30–80

[*] For all repeats, experimentally demonstrated structures are listed.

[†] Members of the Fragile X-related disorders.

[‡] 30 repeats: normal allele; 45–54 repeats: intermediate allele, uncertain pathogenicity.

[¶] Rare, highly unstable.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

The requirement of MMR proteins for repeat expansion in different mouse models of the TNRs.

Genes (complex)	Disease model	Germ line	Somatic	References
MSH2	DM1 (TG)	Yes (95%)	Yes	Savouret *et al.* (2003) and Tome *et al.* (2009)
(MutSα; β)	HD (TG)	Yes	Yes	Manley *et al.* (1999) and Kovtun & McMurray (2001)
	HD (KI)	Yes (♂), no (♀)	Yes	Wheeler *et al.* (2003)
	FXS (KI)	Yes	Yes	Lokanga *et al.* (2014b)
	FRDA (TG)	No	Yes	Ezzatizadeh *et al.* (2012) and Bourn *et al.* (2012)
MSH3	DM1 (TG)	Yes (~80%)	ND	Foiry *et al.* (2006)
(MutSβ)	DM1 (KI)	ND	Yes	van den Broek *et al.* (2006)
	HD (TG)	ND	Yes	Tome *et al.* (2013)
	HD (KI)	No	Yes	Kovalenko (2012) and Dragileva *et al.* (2009)
	FRDA (TG)	No	No	Bourn *et al.* (2012)
MSH6	DM1 (TG)	Yes (♀)*, no (♂)	ND	Foiry *et al.* (2006)
(MutSα)	DM1 (KI)	ND	Protects	van den Broek *et al.* (2006)
	HD (TG)	No	No	Kovtun *et al.* (2004) and Tome *et al.* (2013)
	HD (KI)	No	No	Kovalenko *et al.* (2012) and Dragileva *et al.* (2009)
	FRDA (TG)	Protects	Yes	Ezzatizadeh *et al.* (2012) and Bourn *et al.* (2012)
MLH1	HD (KI)	ND	Yes	Pinto *et al.* (2013)
(MutLα; β; γ)	FRDA (TG)	Yes	Yes	Ezzatizadeh *et al.* (2014)
MLH3 (MutLγ)	HD (KI)	ND	Yes	Pinto *et al.* (2013)
PMS2	DM1 (TG)	ND	Yes (50%)	Gomes-Pereira *et al.* (2004)
(MutLα)	FRDA (TG)	Protects	Protects†	Ezzatizadeh *et al.* (2012) and Bourn *et al.* (2012)

TG, transgenic; KI, knock-in; ND, not determined; %, fraction of expansions that are dependent on the gene in question.

*Effect may be indirect via loss of MSH3.

†In neural tissue but not non-neuronal tissue.