

Review

Modifiers of CAG/CTG Repeat Instability: Insights from Mammalian Models

Vanessa C. Wheeler^{a,b,1,*} and Vincent Dion^{c,1,*}

^a*Molecular Neurogenetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA*

^b*Department of Neurology, Harvard Medical School, Boston, MA, USA*

^c*UK Dementia Research Institute at Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff, UK*

Abstract. At fifteen different genomic locations, the expansion of a CAG/CTG repeat causes a neurodegenerative or neuromuscular disease, the most common being Huntington's disease and myotonic dystrophy type 1. These disorders are characterized by germline and somatic instability of the causative CAG/CTG repeat mutations. Repeat lengthening, or expansion, in the germline leads to an earlier age of onset or more severe symptoms in the next generation. In somatic cells, repeat expansion is thought to precipitate the rate of disease. The mechanisms underlying repeat instability are not well understood. Here we review the mammalian model systems that have been used to study CAG/CTG repeat instability, and the modifiers identified in these systems. Mouse models have demonstrated prominent roles for proteins in the mismatch repair pathway as critical drivers of CAG/CTG instability, which is also suggested by recent genome-wide association studies in humans. We draw attention to a network of connections between modifiers identified across several systems that might indicate pathway crosstalk in the context of repeat instability, and which could provide hypotheses for further validation or discovery. Overall, the data indicate that repeat dynamics might be modulated by altering the levels of DNA metabolic proteins, their regulation, their interaction with chromatin, or by direct perturbation of the repeat tract. Applying novel methodologies and technologies to this exciting area of research will be needed to gain deeper mechanistic insight that can be harnessed for therapies aimed at preventing repeat expansion or promoting repeat contraction.

Keywords: Huntington's disease, myotonic dystrophy 1, DNA repair, trinucleotide repeat instability, mismatch repair

INTRODUCTION

The first discoveries in the 1990s that inherited unstable trinucleotide repeats caused human neurological and neuromuscular diseases stimulated con-

siderable early interest in understanding the mechanisms underlying repeat instability (see [1] for a historical perspective). These so-called repeat expansion disorders exhibit instability of the expanded repeat tract upon transmission to the next generation as well as instability in somatic cells [2–30]. To study mechanisms of repeat instability, a wide variety of model systems in different organisms, from *E. coli* to mice, have been implemented. In particular, a large body of research on repeat instability has been conducted in *S. cerevisiae* [31–35] and has guided many of the subsequent mammalian

¹These authors contributed equally to this work.

*Correspondence to: Vanessa C. Wheeler, Center for Genomic Medicine, Massachusetts Hospital, Boston MAA 02115, USA. E-mail: wheeler@helix.mgh.harvard.edu and Vincent Dion, UK Dementia Research Institute at Cardiff University, Hadyn Ellis Building, Maindy Road, CF24 4HQ Cardiff, UK. E-mail: Dionv@cardiff.ac.uk.

studies. Collectively across all organisms, many genes and candidate pathways have been identified that can play a role in altering repeat dynamics in the various model systems. Much less well understood is the relevance of most of these genes and pathways to human disease, as most have not been studied in the context of relevant cell-types in animal models or have no direct validation in patients.

Here we review the modifiers and potential mechanisms underlying the instability of CAG/CTG repeats, gleaned primarily from mammalian cell-based and animal models, and put these observations into the context of recent human genetic data. We focus on the coding CAG repeat causing Huntington's disease (HD) (OMIM #143100) and the non-coding CTG repeat causing myotonic dystrophy type 1 (DM1) (OMIM# 160900). These are the most common of the thirteen CAG/CTG repeat expansion disorders [22], upon which the majority of the mouse and patient cell-based models are based.

The HD CAG repeat is located within exon 1 of the *HTT* gene. CAG repeat lengths of 6–35 are not clearly disease-associated, whereas 36–39 CAGs and 40+ CAGs are associated with incompletely or fully penetrant disease, respectively [3, 36]. A rarer juvenile-onset form of the disease is usually associated with >~60 CAGs [37]. Intergenerational changes in repeat length, associated in males with instability in sperm DNA, are seen in the disease range (36+ CAGs) as well as in the high normal range (27–35 CAGs), the latter giving rise to new mutations [26, 38]. Somatic instability has been studied in individuals with disease-associated repeat lengths. The instability in somatic cells is expansion-biased, occurs in both in the brain and in peripheral tissues, but shows tissue- and cell-type specificity; forebrain regions show relatively high levels of expansion, the cerebellum exhibits low levels of expansion, and neurons have greater expansions than glia [2, 29, 30]. Of the peripheral tissues, liver is relatively unstable [2]. Somatic expansion in postmortem cortex was found to be inversely correlated with the age of disease onset [4].

The DM1 CTG repeat is located within the 3' untranslated region (UTR) of the *DMPK* gene. Individuals are classified into five categories based on the age at onset: congenital, infantile, juvenile, adult, and late-onset [39, 40]. These are largely correlated with repeat size, but there is significant overlap between the categories [41]. Repeat lengths of 5–~37 CTGs occur in the general population, and it is thought that alleles between 16 and 30 repeats are the source of

rare *de novo* expansions [42]. Affected individuals can harbor from ~50 to several thousand repeats, with congenital DM1 often presenting with alleles >1,000 CTGs [43, 44]. Repeats of ~50–79 exhibit high levels of male germline expansion, whereas 80+ repeats tend to exhibit greater expansions in female transmissions [41]. Somatic instability can be detected in fetal tissues between 13 and 16 weeks of gestation, and in patients shows expansion-biased and tissue-dependent instability, with high levels seen in heart, skin, muscle and blood, as well as in brain tissues [41, 45].

Thus, patient data indicate that there may be differences in mechanisms of instability in HD and DM1, likely due to genomic location and repeat length. However, as revealed by studies in mice there are clearly shared underlying mechanisms. Thus, observations in DM1 models may well be relevant in HD where the CAG repeat can expand somatically to hundreds of repeats [4, 29]. The reverse is also true: the observations made in HD models are likely to be relevant to DM1, in particular for repeat lengths in the lower range. Overlapping mechanisms extend to other trinucleotide repeats, as discussed in [46]. Equally important however, are aspects of repeat instability and modifiers that distinguish the different disease-associated repeats that may yield important mechanistic insight.

As described in the following sections, many genes that function in DNA repair pathways have been found to modify CAG/CTG instability. In several cases (see: "*Non-Canonical Mechanisms Underlying Repeat Instability?*") it appears that the encoded proteins do not in fact "repair" DNA, but rather promote mutation. For the purpose of this review, we use the term "DNA repair" to describe the process or function to which genes or pathways are traditionally assigned, rather than the functional outcome of that process in the context of expanded repeats. More specifically, and as an example, we use the term "mismatch repair" (MMR) to mean the canonical post-replicative pathway and its components [47], acknowledging that the specific functions of the proteins in the pathway may be different in the context of expanded repeats.

THE LINK WITH HUMAN GENETIC STUDIES

An ongoing challenge is to understand which of the many modifiers identified in model systems is relevant in humans. Recent genome-wide associa-

tion studies (GWAS) for modifiers of onset in HD have identified genes in the MMR pathway (*MSH3*, *MLH1*, *PMS2*, *PMS1*) as modifiers of the rate of disease onset in patients [48–51]. *MSH3*, also modified a measure of HD progression [52]. Other DNA repair genes identified in the age-at-onset GWAS were *LIG1*, encoding a DNA ligase that can function in several DNA repair pathways, and *FAN1*, encoding FANCD2 and FANCI Associated Nuclease 1, with a role in repairing interstrand crosslinks. Several MMR pathway genes had been previously identified as critical drivers of somatic repeat expansion in mouse models (see below), thus providing strong support for somatic expansion as the underlying mechanism for disease modification in HD. Notably, the genes that have emerged in the human genetic studies to date do not obviously highlight a role for a general DNA damage response as a driver of disease onset, where key players include ATM and ATR for example [53]. Interestingly, two HD onset-modifying DNA repair genes, *FAN1* and *PMS2*, were also associated with age at onset in the spinocerebellar ataxias (SCAs), caused by CAG expansion mutations but having distinct neuropathological profiles [54], suggesting somatic expansion as common disease onset-modifying mechanism. MMR pathway genes *MSH3*, *MLH1* and *MLH3*, as well as *FAN1*, are also associated with repeat instability in HD or DM1 patient blood [50, 55–57].

Human genetic discovery efforts help to define the DNA repair factors and pathways that may play prominent roles in the process of somatic repeat expansion in cell-types that are relevant to the onset or progression of disease phenotypes, prioritizing targets for therapeutic intervention [58]. Larger sample sizes would be expected to reveal genetic modifiers that are rarer in the population and/or that have weaker effects. However, the absence or rarity of naturally occurring functional variation in a particular gene will preclude its detection in genetic association studies. Therefore, complementary investigation of instability in model systems allows the identification of modifier genes that may not be revealed in human genetic studies, but which may be part of the same pathway(s), better defining pathways and mechanisms and providing additional points for possible therapeutic intervention. An example is MMR pathway gene *MSH2*; this gene has not emerged as a human onset modifier, yet it is clearly required for somatic expansion in mouse models as well as in patient cell-based models (Table 1). This may indicate the lack of functional *MSH2* variation in the

human population at sufficiently high frequency to be detected in the GWAS studies to date.

NON-CANONICAL MECHANISMS UNDERLYING REPEAT INSTABILITY?

The underlying mechanisms by which DNA repair proteins modulate repeat instability remain to be clearly defined. However, it is conceivable that these proteins could act in non-canonical ways or in non-canonical pathways. DNA repair pathways are traditionally defined based on the repair of a specific lesion, such as a mismatch or a single-strand break. However, the substrate(s) or lesion(s) at repeat tracts may be highly unusual, and the classical definition of DNA repair pathways may break down. Therefore, it is plausible that DNA repair proteins have roles promoting or protecting against repeat instability that differ from those usually described during the repair of a specific lesion. Most notably, the same MMR proteins that normally act to suppress genomic instability by repairing post-replicative errors, promote CAG/CTG instability in non-dividing cells. This counterintuitive finding suggests that repeat tracts might engage components of the MMR machinery in unusual ways, and/or could implicate other pro-mutagenic processes in which MMR factors play roles [59, 60]. There is also considerable cross-talk between proteins in DNA repair pathways (see non-exhaustive examples in these reviews: [61–65]). Thus, DNA repair proteins may work together in unexpected ways to modulate repeat dynamics. For example, *FAN1* is not a component of the classically-defined MMR pathway, yet *Fan1* interacts genetically with *Mlh1* to control somatic CAG instability in mice [66].

Although it remains helpful to broadly classify modifiers of repeat instability by the pathways in which they were first described, as categorized below and illustrated in Fig. 1, we also present genetic modifiers identified in the various mammalian model systems in the context of known associations between them, in a manner that is agnostic to their classically defined functions (Fig. 2). This allows a better appreciation of the extent to which many of the modifiers (Table 1), regardless of the system(s) in which they were identified or their specific role in modifying repeat instability, are connected as part of a network. Importantly however, the vast majority of these genes have not been validated as modifiers of repeat instability in mouse models, and/or their

Table 1

Genetic modifiers of repeat instability found in mammalian systems. Excluded from this list: modifiers that act in the context of an exogenous perturbation or sensitizer (e.g., under replication stress, or upon double-strand break or nicks induced within the repeat tract)

Type of model/method to detect instability	CAG/CTG repeat substrate	Normal Gene function					References
		enhances contractions	suppresses contractions	enhances expansions	suppresses expansions	No effect found	
Integrated chromosomal selectable reporter. Only detects large contractions	(CAG) _{~95} including 19 nt 5' and 43 nt 3' flanking <i>DMPK</i> sequence in intron 2 of <i>HPRT</i> [67].	<i>BARD1, BRCA1, CSB, ERCC1, MSH2, MSH3, TFIIIS, XPA, XPG.</i>	<i>DNMT1, HSP90, MLH1, PARP1, PMS2, RAD51, RNASEH1, RNASEH2a, TOP1, TDPI, XRCC1.</i>			<i>APEX1, FEN1, MLH3, MSH6, OGG1, XPC.</i>	[67–73]
Shuttle vectors using human cells and reporter in <i>S. cerevisiae</i> . Only detects expansions.	(CTG) ₂₂ 5' of <i>S. cerevisiae</i> <i>CAN1</i> ATG or of <i>URA3</i> ATG			<i>HDAC3, HDAC5, MSH3, MSH2, PSMB3, PSMC5.</i>	<i>CBP, EP300, HDAC9, HLTF, RAD18, RTEL1.</i>	<i>CTIP, DSS1, FBH1, HDAC1, MRE11, MSH6, SHPRH.</i>	[74–78]
Ectopic loci in human cell lines: direct detection by PCR	(CTG) _{~800} in <i>DMPK</i> 3' UTR [79]; (CTG) _{~102} or (CAG) _{~102} adjacent to <i>Myc</i> replication origin [80, 81]; <i>HTT</i> exon1 with (CAG) _{~118} [82]	<i>MSH2, MSH3, TFIIIS.</i>	<i>CLSPN, FEN1[§], SETX, TIMELESS, TIPIN, *TOP1 + TDPI.</i>	<i>MSH2, MSH3, TFIIIS.</i>	<i>FAN1, MSH6.</i>	<i>MLH1, PMS2, *TOP1, *TDPI.</i>	[79–82]
Endogenous loci in patient-derived cells: direct detection by PCR	<i>DMPK</i> (CTG) _{~400–1000} [80]; <i>DMPK</i> (CTG) _{~773} [83]; <i>HTT</i> (CAG) _{~109} [82]; (CAG) _{~72} [84]		<i>CLSPN, TIMELESS, TIPIN.</i>	<i>MSH2.</i>	<i>FAN1.</i>		[80, 82–84]
Knock-in and transgenic mice: direct detection by PCR	<i>Htt</i> CAG knock-in mice: <i>Htt</i> ^{Q111} :(CAG) _{~120} <i>Htt</i> ^{Q150} :(CAG) _{~150} <i>Htt</i> ^{Q50} and <i>Htt</i> ^{neoQ50} :(CAG) ₄₈		<i>Msh2[⊖], Msh6[⊖].</i>	<i>Hdac2, Hdac3, Msh2, Msh3, Mlh1, Mlh3, Ogg1.</i>	<i>Fan1.</i>	<i>Xpc, Aif, Dat.</i>	[66, 85–91]
	R6/1 <i>HTT</i> exon 1 transgenic mice: (CAG) _{~120}	<i>Csb[#].</i>	<i>Fen1[⊖].</i>	<i>Msh3, Ogg1, Neil1, Fen1[⊖].</i>	<i>Csb[#].</i>	<i>Msh6, Aag1, Nht1.</i>	[92–101]
	<i>Dmpk</i> CTG knock-in mice: (CTG) _{~84}			<i>Msh3.</i>	<i>Msh6.</i>	<i>Fen1.</i>	[102, 103]
	<i>DMPK</i> transgenic mice: DM300-328: (CTG) _{>300} Dmt-D: (CTG) _{~160}		<i>Msh2, Msh3, Msh6[†] Lig1[†].</i>	<i>Msh2, Msh3, Msh6^{&}, Pms2, Lig1[†], Rad52^{&}.</i>		<i>Msh6, Rad54, DNAPKcs.</i>	[104–107]
	<i>Atxn1</i> CAG knock-in mice: (CAG) _{~145}	<i>Dnmt1^{&}.</i>		<i>Xpa.</i>		<i>Dnmt1^{&}.</i>	[70, 108]

[⊖]No effect of *MSH3* knockout on contractions in a (CTG)₃₃-*URA3* shuttle vector assay that detects contractions only [78]. ^{*}Effect only seen with simultaneous knockdown of both *TOP1* and *TDPI*. [#]In male germline transmissions only; in somatic cells, effect only seen with simultaneous knockout of both *Csb* and *Ogg1*. [§]Depends on the orientation of the repeat tract with respect to the origin of replication. [&]only for intergenerational transmissions. [⊖]Only in male germline transmissions. [†]Only during maternal germline transmissions. *Htt*^{Q111}, *Htt*^{Q50}, *Htt*^{neoQ50} and R6/1 mice have a canonical human (CAG)_nCAACAG repeat structure; *Htt*^{Q150} mice do not have a penultimate CAACAG. Note that the mouse knock-in models are also referred to using the original gene nomenclature as *Hdh*^{Q111}, *Hdh*^{Q50}, *Hdh*^{neoQ50}, and *Hdh*^{Q150}.

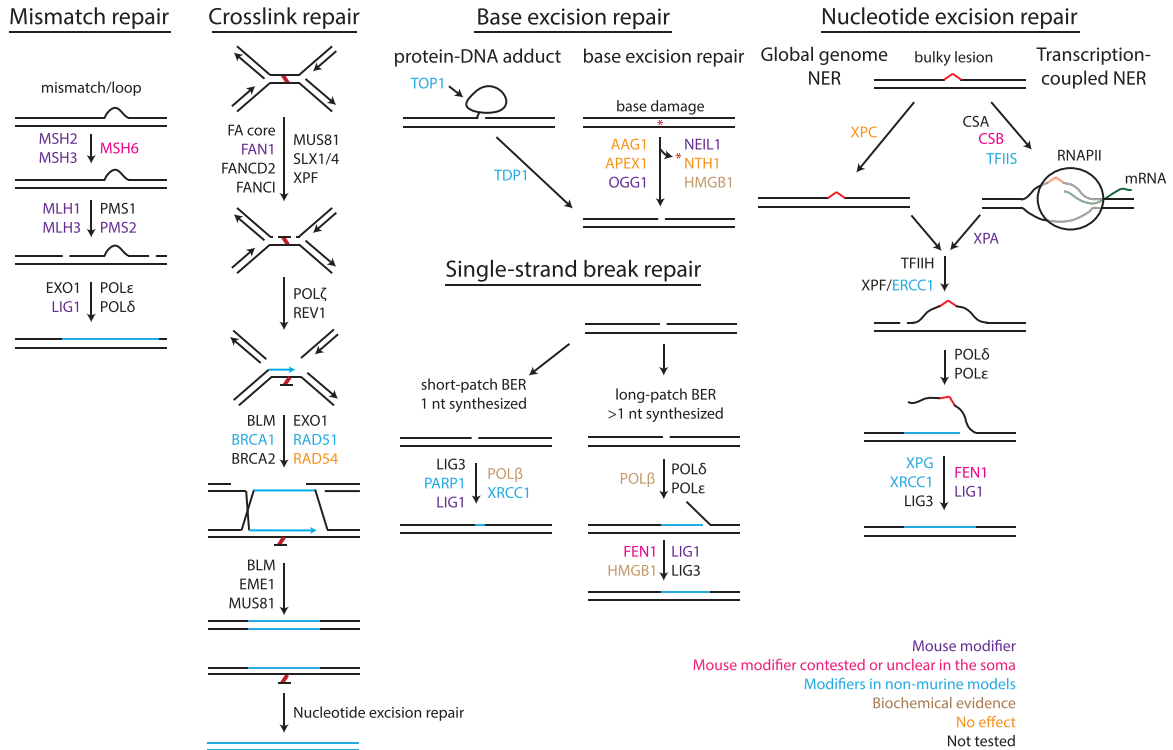


Fig. 1. Classically defined DNA repair pathways implicated in repeat instability. Outline of four different repair pathways, indicating a non-exhaustive list of proteins within those pathways, either based on *in vivo* evidence or from reconstituted systems. Modifiers of repeat instability are color-coded as follow: purple: modifier of repeat instability in murine models; pink: modifier with contested or unclear somatic instability data; blue: modifiers in non-murine models; brown: modifier in biochemical assays; orange: no effect when tested in a mammalian system; black: not tested for an effect on repeat instability. See Table 1 for direction of effect of modifiers in mammalian systems. Figure update from [109] with permission.

relevance to patients is unknown (see color coding in Fig. 2). The connections in this network may therefore provide testable hypotheses for future discovery of genes and pathways that might be relevant to repeat instability in disease.

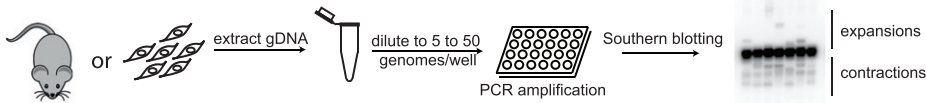
MAMMALIAN MODELS USED TO STUDY CAG INSTABILITY

The principal types of cell or animal-based mammalian model systems and assays that have been developed to investigate CAG/CTG instability, are summarized in Table 1 and in Fig. 3. Knock-in and transgenic mice have been generated with unstable repeats, typically containing more than 100 units [110]. Repeat instability is analyzed in these models by PCR, using either small pool-PCR with Southern blot detection, or using fragment analysis on a DNA sequencer [13, 87]. These models exhibit repeat length-and-time-dependent, expansion-biased

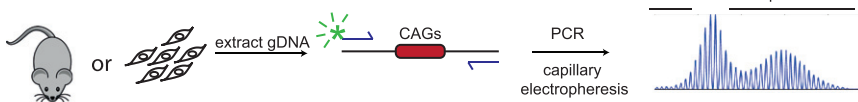
somatic instability that is tissue or cell type-specific [87, 89, 102, 111–117]. In the HD models, high levels of expansion are seen in the striatum, in particular in the medium-spiny striatal neurons (MSNs), and in the liver, attributable to hepatocytes [87, 89, 102, 116, 117]. The cerebellum exhibits relative stability [87, 116]. Overall, there is very good correlation with tissue-specific instability patterns observed in both adult and juvenile onset HD [2]. DM1 mouse models can exhibit relatively high instability in kidney, liver, skeletal muscle and brain, including the striatum, with relatively low instability in the spleen, heart and cerebellum [112, 115, 118]. There appears to be more variation in the tissue-specific instability patterns between different DM1 lines than apparent across HD mouse models, and in general, extensive comparisons with humans have been limited by the lack of availability of human tissues. Both HD and DM1 models exhibit intergenerational repeat length changes that recapitulate some, but not all of the features of intergenerational instability in patients.

Direct detection of repeat instability

Small-pool PCR

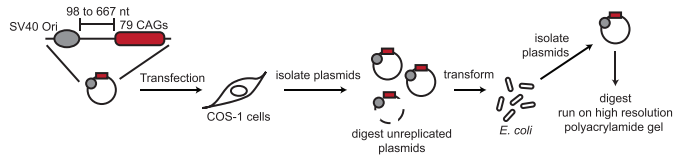


Fragment analysis

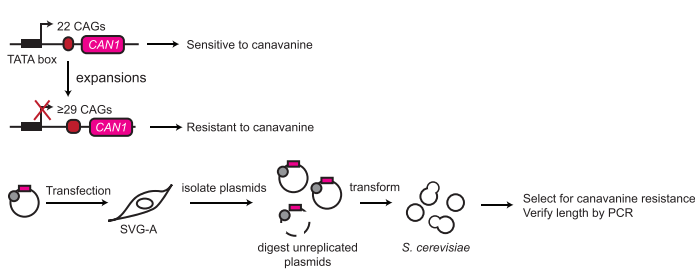


Shuttle vectors

Direct detection

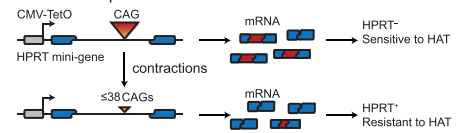


Selectable reporter



Integrated chromosomal reporters

Selectable reporter



Fluorescent reporter

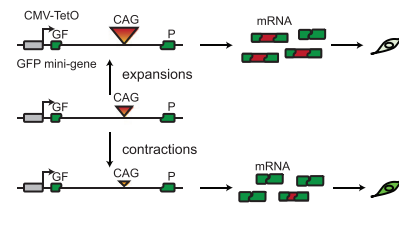


Fig. 3. Common assays for measuring repeat instability. Direct detection of repeat length changes is used to analyze instability in mouse models, patient-derived cells and cell-based models harboring repeat tracts at ectopic loci. Small-pool PCR is the current gold standard as it reduces bias against amplification of longer alleles, it is quantifiable, and can detect both expansions and contractions. It is prone to carry-over contamination and is time-consuming. Recently published methods address those concerns [144, 145]. Fragment analysis, e.g., using GeneMapper [111] is quicker and easier than the small-pool PCR, is not as sensitive to rare events but is also quantitative [87]. Methods for studying repeat instability are discussed in detail in [146]. A shuttle vector assay in which repeat length changes are detected directly has been used for mapping cis-acting elements, including origin of replication distance and direction as well as DNA methylation [129, 147]. It has not been used in conjunction with genetic perturbation. Shuttle vector assays using selectable reporters were first developed as a unidirectional contraction assay [148, 149] and later as an expansion-only assay [75, 128]. Two versions of the latter exist, one using the *S. cerevisiae* *CAN1* gene, the other the *URA3* gene. The *CAN1* reporter assay has been extensively used for genetic studies (see Table 1). Integrated chromosomal reporters have the advantage of being ultra-sensitive, but limited in the types of events they can detect. The first ones relied on APRT and HPRT function. An expansion assay was developed but was found to have impractically low frequencies of expansions [150]. The expression can be controlled with a doxycycline-inducible promoter [67]. HAT: medium containing hypoxanthine aminopterin and thymidine. An integrated chromosomal reporter based on GFP fluorescence can detect both expansion and contractions, but has not yet been used to uncover genetic modifiers [130, 133].

to achieve frequencies of intergenerational changes and/or large jumps in repeat size that are seen in patients [112, 120, 121]. Thus, parallels in many of the features of repeat instability in patients and in mice indicate that the mice models are likely to afford insight into mechanisms of instability that are relevant to disease in patients. Other advantages of mouse models are that they allow the dissection of

modifiers that might act in a tissue/cell type-specific manner, including distinguishing mechanisms that act in the germline and in the soma. They also permit the analysis of naturally occurring strain-specific variation that is associated with different levels of instability [90, 94, 117, 121, 122]. Moreover, comparisons between models that differ in repeat length and genomic contexts can provide information on the role

of *cis* elements that might contribute to repeat instability [111, 115, 119, 121, 123–125]. While yielding rich *in vivo* insights into potential mechanisms of repeat instability, the throughput in mice is low due to the need for extensive breeding to test genetic modifiers, and the time (several weeks to months) for somatic repeat expansions to accumulate. This is especially challenging in models with shorter inherited repeat lengths, e.g., within the adult-onset range of CAG lengths for HD (~40–50); although these models would better model the typical time-course of somatic expansion seen in most HD patients, the relatively slow rate of somatic expansion and the limited extent of expansion within the lifespan of a mouse [66] may make some modifier studies impractical. Finally, the ability to query a particular genetic modifier will depend on germ cell or mouse viability when the gene is knocked out or mutated.

At the other end of the spectrum are systems in cultured human and primate cells, either utilizing plasmid-based mammalian-*E. coli*/*S. cerevisiae* shuttle vectors, or chromosomally-integrated reporters (Fig. 3). In the shuttle vector systems, CAG/CTG-containing plasmids are first introduced into mammalian cells and, following the experimental perturbation, are transformed into *E. coli* or *S. cerevisiae* for read-outs of instability. In one system, which has been used extensively to test genetic modifiers repeat length changes can be determined in *S. cerevisiae* based upon a CAG length-dependent resistance to 5-fluoroorotic acid (5-FOA) [128] or to canavanine [75]. In another system, repeat length changes are measured by digesting plasmids isolated from single *E. coli* colonies and run on high-resolution polyacrylamide gels [129]. Shuttle vector systems have the major advantage of requiring only a few days. Integrated chromosomal reporters have been developed based on a CAG length-dependent sensitive hypoxanthine-guanine phosphoribosyltransferase (HPRT) or adenine phosphoribosyltransferase (APRT) activity, which can be selected in the appropriate culture media, or based on CAG length-dependent GFP fluorescence. These chromosomal reporter systems require a longer time for read-outs of instability compared to shuttle vectors but offer a higher throughput [67–73, 130–135]. Notably, the largest small molecule screen conducted so far was performed using a selectable chromosomal reporter system and included 880 compounds [72]. The use of selectable markers, in the context of either the shuttle vector or integrated reporter assays allows for the sensitive detection of low frequency instability

events, and the ease of combining both of these systems with knockdown, and more recently knockouts, has greatly facilitated the dissection of DNA repair pathways that are involved in CAG/CTG instability in cultured mammalian cells [69, 72, 74–76]. The disadvantages of these systems are that they work in immortalized and rapidly dividing cells, they probe the instability of repeats outside of their endogenous genomic loci, they detect rare events, and are only sensitive to specific types of repeat length changes that can be selected for in the respective assays. For example, the chromosomal reporter systems based on a (CAG)₉₅ repeat in an intron of the *HPRT* gene or the *APRT* gene are only sensitive to large repeat contractions that bring the repeat size below a threshold of ~38 repeats, which is required to restore HPRT or APRT function [132]. Conversely, the human astrocyte cell line (SVG-A)/*S. cerevisiae* shuttle vector system is only sensitive to expansion of a (CTG)₂₂ tract to 29 or more repeats or more, which alters transcription initiation and blocks expression of the *CAN1* or *URA3* reporter [75, 128]. Therefore, these systems are unlikely to recapitulate fully what occurs in disease-relevant cell types. Use of a green fluorescent protein (GFP) reporter mitigates one of these issues by enabling the simultaneous read-out of both expansions and contractions while still providing an assay system that is compatible with high throughput screening [130, 133, 134].

Other mammalian cell-based systems are those in which repeats are detected directly by PCR, allowing monitoring of both expansions and contractions that occur at relatively high frequencies. These include cell lines stably transfected with plasmids harboring expanded repeats at ectopic loci [79–82], those derived from mouse models [87, 136] or from patient-derived cells including lymphoblastoid cell lines and fibroblasts [80, 137–139], embryonic stem cells [140, 141], and induced pluripotent stem cells (iPSCs) [82–84]. Pluripotent cells have the major advantage of having the potential to differentiate into many different cell types. In practice, differentiation protocols can often yield a variable fraction of the intended cell type. Modeling repeat instability in patient cells has the advantage of providing direct insight into human-relevant modifiers of instability of repeats in their appropriate genomic contexts. However, these—and all patient-derived cell models—are slow, typically requiring long-term culturing of weeks to months to observe measurable instability. Dividing cells may also introduce selection or clonal artefacts that need to be considered [136, 142]. Differentiating stem cells

into specific cell types found to exhibit high levels of instability in mice and patients, e.g., MSNs, may increase the rate of repeat expansion in cultured cells. However, improved differentiation protocols are needed to increase the purity of the desired cell type. Human organoids have not yet been used for the study of repeat instability, but have the potential to model complex tissue environments in the context of human mutations [143].

Despite the disadvantages inherent to all these systems and approaches, each has yielded important insights into modifiers of CAG/CTG instability, contributing to our understanding of potential underlying mechanisms. Many factors, particularly MMR genes, modulate instability across multiple systems, indicating the value of initial high-throughput systems as screening tools to generate candidates for testing in mice or human cells (Table 1).

MISMATCH REPAIR FACTORS

The MMR pathway is best characterized for its role in correcting DNA mismatches generated during DNA replication but has additional functions in DNA recombination and in DNA damage signaling (reviewed in [47, 151, 152]). In mammals, MSH2-MSH6 dimers (MutS α) primarily recognize base-base mismatches, whereas MSH2-MSH3 dimers (MutS β) primarily recognize insertion-deletion loops. During the process of MMR, DNA recognition is followed by binding of MLH1-PMS2 (MutL α) that mediates the recruitment of downstream effector proteins to excise and repair the lesion. Two additional MutL dimers have been described: MLH1-MLH3 (MutL γ) plays a role in meiosis, whereas the role of MLH1-PMS1 (MutL β) is yet unclear. Mutations in *MSH2*, *MSH6*, *MLH1* and *PMS2* underlie the cancer prone Lynch syndrome in which biallelic mutations result in elevated instability of microsatellite repeat tracts [153]. This observation prompted an early interest in the role of MMR factors in the instability of disease-associated trinucleotide repeats.

A role for MMR pathway genes in CAG instability in an animal model of disease was first demonstrated by Manley et al., in which genetic knockout of *Msh2* eliminated somatic *HTT*CAG expansion in R6/1 exon 1 HD transgenic mice [97]. Constitutive, or striatal MSN-specific, *Msh2* knockout in *Htt*^{Q111} HD knock-in mice abrogated CAG expansion in the striatum [86, 88, 89], demonstrating that MSNs harbored the most highly expanded alleles and that a process dependent

on a MMR gene driving CAG expansion was active in post-mitotic neurons. Knockout of *Msh2* also suppressed male gametic expansion [98] and inherited repeat length changes in paternal transmissions [86, 88]. The absence of *Msh2* promoted repeat contraction in paternal transmissions [86, 88], indicating a distinct role for MSH2 in protecting against CAG contractions. Similarly, in a DM1 transgenic mouse model harboring a long (>300 CTG) repeat tract (DM300-328), *Msh2* knockout suppressed expansions and enhanced contractions in both the soma and the germline [106].

Genetic knockout of *Msh3* in multiple knock-in and transgenic models implicates MutS β as the major driver of somatic CAG/CTG expansion [88, 93, 103, 105]. In the germline, *Msh3* knockout had a moderate impact relative to that of the *Msh2* knockout in *Htt*^{Q111} mice, but strongly suppressed germline expansions in DM300-328 mice [88, 105]. *Msh3* knockout also promoted contractions in DM300-328 mice, indicating a contraction-suppressor role for MutS β [105]. In a cell-free SV40 replication assay MutS β also protected against contractions [154], with the suggestion that a contraction-suppressor role may be relevant to dividing cells *in vivo*. The contrasts between the *Htt*^{Q111} and DM300-328 models suggest potential differences in MMR-related mechanisms in the soma and germline that might depend on the disease gene context and/or CAG/CTG repeat length. Notably, in both HD and DM1 models, heterozygous *Msh3* knockout was sufficient to reduce CAG/CTG expansion, indicating that MSH3 levels are rate limiting in the expansion mechanism [88, 105]. Consistent with this, *Msh3* expression levels correlated with CAG expansion in mouse strains harboring naturally occurring *Msh3* variants [94] and in a human SVG-A astrocyte cell-based model [78]. The extent to which MSH3 levels contribute to CAG expansion that differs between tissues or between cell types [30, 83, 87, 140, 155, 156] is not well understood and warrants further investigation.

The role of *Msh6* is much less clear, with variable effects in mice depending on the model. In *Htt*^{Q111} mice, *Msh6* knockout had no impact on striatal expansions, whereas heterozygous knockout promoted contractions in the male germline [88]. In R6/1 mice, *Msh6* knockout slightly reduced expansion in a tissue-specific manner [93]. In DM1 models, *Msh6* knockout had either no impact [105] or promoted expansion [118] in somatic cells, and suppressed expansions and promoted contractions in the female germline [105]. Some of these effects

may be attributable to altered levels of the MutS β complex as a consequence of reduced MSH6 protein. However, this remains to be tested, and direct roles of MutS α are also possible. Overall the data indicate that although *Msh6* is not a key driver of somatic CAG/CTG expansion, it may modulate various aspects of repeat instability with the extent of its involvement perhaps depending in part on the relative levels of MSH6 and MSH3 in different cell types [157]. Cell-based systems, including reporter-based assays that exclusively detect either expansion or contraction events, have reinforced CAG/CTG instability-promoting roles of MSH2 and MSH3 observed in mice, whilst showing minimal impact of MSH6 [67, 69, 76, 79, 83].

In addition to *Msh2* and *Msh3*, both *Mlh1* and *Mlh3* are absolutely required for somatic CAG expansion in *Htt*^{Q111} mice [90], implicating MutL γ in this process. *Pms2* knockout partially reduced expansions in DM1 transgenic mice, also implicating MutL α [104]. Naturally occurring strain-specific *Mlh1* variation was associated with CAG expansion, an effect that could in part be explained by altered MLH1 expression [90]. Both MutL α and MutL γ possess endonuclease activities that, in reconstituted systems, can result in the retention of CAG or CTG loop-outs, representing expansion events [158, 159]. MutL α was also capable of eliminating such loop-outs in this system, representing potential contractions [158]. MutL γ endonuclease activity has been implicated in the expansion of GAA repeats [160]. Note that in cell-based systems, modifier roles of MutL genes were not found [79] or were inconsistent with observations in mice [69]. Importantly, the critical requirement for MutL genes for somatic expansion in mouse models supports the idea that active DNA repair is required to drive CAG/CTG expansion, rather than subversion of a normal MMR process (see below). This is also supported by data in mice indicating that the ATPase activity of MSH2, which is essential for the recruitment of MutL proteins, is required for repeat instability [95]. Similarly, an ATPase mutant in MSH3 behaved like a knockout in a human cell system for repeat expansion [78]. Downstream MMR factors (exonucleases, polymerases, ligases) that are involved in repeat expansion have not been delineated. The identification of *LIG1* as a human onset modifier [50], and the knowledge from reconstituted systems that this DNA ligase can function in MMR [161], suggests that it may be part of the MMR mechanism that drives somatic expansion. In mice, *Lig1* heterozygosity leads to more expansions and fewer

contractions exclusively in the female germline of DM1 transgenic mice [107], and *LIG1* expression levels were found to correlate with repeat instability in a cell-free replication assay [162]. Further studies on the role of *LIG1* in somatic expansion are of interest.

To gain insight into the mechanism by which MMR proteins promote expansion, several biochemical studies have also been conducted to understand the nature of the repeat substrate that can be bound by MutS β and the consequences of repeat binding. Both short (1–3 repeat) loop-outs and longer CAG repeat-containing hairpin structures can bind MutS β [93, 163–166]. Some, but not all, of these studies indicated that CAG hairpins can alter properties of MutS β binding or activity [93, 164, 166]. These observations have provided some controversy as to whether MutS β binding to repeats inhibits the normal process of MMR. In cell-free assays that measure the repair of repeat-containing substrates, the repair of short loop-outs was found to be dependent on MutS β and MutL α [163, 167] and hairpin repair was stimulated by MutS β [164]. Although these assays do not provide a direct readout of instability *per se*, these findings support the idea that MutS β binding to repeat structures stimulates, rather than inhibits, a repair process that ultimately results in expansion, consistent with genetic data in mice. Although CAG/CTG slipped-strand structures have been identified in patient tissues [168], and enrichment of MMR proteins can be observed close to CAG/CTG repeat tracts using chromatin immunoprecipitation [76, 79, 83, 96, 169], the nature of the substrate(s) bound by MutS β *in vivo* remains unknown. Further studies are needed to tie biochemical observations to repeat instability outcomes.

FAN1: A PROTEIN INVOLVED IN INTERSTRAND CROSS LINK REPAIR

FAN1 is required for the repair of interstrand DNA crosslinks (ICLs), possessing both 5'→3' exonuclease activity and a structure-specific endonuclease [170–174]. A potential role in CAG repeat instability was first indicated by human GWAS that identified *FAN1* as a modifier of HD onset [49]. Subsequently, it was shown that *FAN1* knockout in a human U2OS cell line model containing *HTT* exon 1 with 118 CAGs, or *FAN1* knockdown in HD patient iPSCs with ~109 CAGs or ~70 CAGs enhanced *HTT* CAG expansion [82, 84]. Knockout of *Fan1* also enhanced somatic CAG expansion in the striatum and other tissues of

Htt^{Q111} mice, in a manner that is dependent on *Mlh1* [66]. Significantly, *Fan1* knockout also promoted expansion of a 48-CAG repeat tract in HD knock-in mice, indicating that FAN1 normally acts to suppress the expansion of CAG repeats that fall into both the adult-onset and juvenile-onset inherited length ranges of the human disease [66]. Notably, FAN1 physically interacts with MLH1, MLH3, PMS1 and PMS2 [170, 175]. The mechanism by which FAN1 normally suppresses CAG expansion, whether physical interaction(s) between FAN1 and MMR proteins are required in the process of CAG repeat instability and whether additional ICL proteins play a role in repeat instability are currently unknown.

BASE EXCISION REPAIR AND SINGLE STRAND BREAK REPAIR

Base excision repair (BER) is initiated by the detection and excision of damaged bases in the DNA [176]. This is followed by the formation of an apurinic/aprimidinic (AP) site that is cleaved by APEX1, which leaves a single base pair gap. DNA polymerase beta (POL β) fills in the missing base pair for a short-patch BER followed by a ligation event or coordinates a long-patch synthesis that involves strand displacement and FEN1 activity in addition to DNA ligases (I and III). Single-strand break repair (SSBR) is often referred to as a sub pathway of BER [177–180] as it uses some of the same proteins, e.g., XRCC1, PARP1, and LIG1 (Fig. 1).

In the context of repeat instability, a chemical screen identified molecules targeting several enzymes in the SSBR pathway including topoisomerase 1, TDP1, and POL β , to stimulate repeat contractions in a selectable assay [72]. Knockdown of *TDP*, *TOP1* as well as other SSBR genes, including *XRCC1* and *PARP1*, confirmed that SSBR protects against contractions [72]. It is unknown whether these factors influence expansions as well. In a different system with 800 CTG repeats inserted at an ectopic site, individual knockdown of *TDP1* or *TOP1*, had no effect on repeat instability, though their simultaneous knockdown dramatically increased repeat contraction [79]. SSBR proteins have not been investigated beyond these two systems, but downstream factors in the SSBR pathway are shared with other pathways and several are involved in repeat instability (Fig. 1), notably XRCC1 and LIG1.

Extensive studies have been carried out to understand the role played by the repair of oxidized bases in repeat instability. Knockout of *Ogg1*, encoding a gly-

cosylase that excises oxidized guanines from DNA, reduced somatic expansion in ~70% of R6/1 mice [100], as well as in *Htt*^{Q150} knock-in mice [85]. Kovtun et al. proposed a model [100] whereby 8-oxoguanine (8-oxoG) excision from within the repeat tract by OGG1 triggers error-prone repair, involving FEN1 and long-patch BER. This process then leads to expansion either by DNA polymerase slippage or because the secondary structure formed within the displaced flap prevents FEN1 action and is ligated in rather than digested [181]. It has been further suggested that the stoichiometry of FEN1 and POL β and coordination between these proteins is important in determining the efficiency of gap-filling synthesis that would be expected to promote expansion [182–184]. Recent studies indicate that interaction between POL β and MutS β [185, 186] also promotes gap-filling synthesis, suggesting cross-talk between BER and MMR pathways [185]. However, a role for FEN1 that is predicted by this model, and previously shown in *S. cerevisiae* to protect against CAG/CTG instability, including expansions [187–189], has been more difficult to test directly in mice due to the lethality of the homozygous knockouts. Heterozygous *Fen1* knockout had no impact on somatic CAG expansion in R6/1 mice [99] or in DM1 knock-in mice [103]. There is some evidence that *Fen1* promotes expansions and suppresses contractions in the male germline [99] and suppresses contractions during DNA replication in a human cell-based model [81] but neither *OGG1* nor *FEN1* knockdown modified contraction frequency in a transcription-dependent chromosomal reporter assay [67, 68]. The role of POL β in CAG/CTG expansion in mice is unknown, though it was found to modify *FMR1* CGG instability in a mouse model of Fragile X-related disorders [190].

The idea that DNA repair triggered by oxidized bases leads to repeat instability is also supported by the observation that knockout of *Neill1*, encoding another glycosylase with a preference for oxidized pyrimidines and minimal activity towards 8-oxoG, reduced somatic expansion in R6/1 mice [92]. Oxidized bases or AP sites have also been implicated in CAG/CTG repeat instability in cell-free systems [191–196], supporting a role for oxidative damage triggering repair within a repeat tract. Interestingly, guanines are more susceptible to oxidation when in the loop of a hairpin, yet are inefficiently repaired [191, 196], further supported by the CAG length-dependent accumulation of oxidative lesions within the CAG repeat tract in HD knock-in mice [183].

These data appear to present somewhat of a paradox inasmuch as they imply the lack of repair at repeat tracts, rather than the active repair of oxidized lesions. However, a model has been proposed in which oxidative lesions that arise transiently in susceptible hairpin structures may be inefficiently repaired and thus incorporated into the repeat [196]. It has also been suggested that oxidized bases themselves may be incorporated [192].

A key tenet of the models involving recognition of oxidized bases is a “feed forward” toxic oxidation cycle whereby the accrual of oxidative lesions with aging and disease pathogenesis progressively increases the susceptibility of the repeat tract to error-prone repair [100], thus accelerating the rate of expansion. As the brain is highly susceptible to oxidative stress, and as there is evidence for the further accumulation of 8-oxoG in brains of R6/1 and R6/2 HD mice [100, 192] as well as in HD post-mortem caudate [197], this model provides a potential link between repeat instability and cumulative damage as part of ageing and neurodegeneration. However, additional observations do not support a clear link between DNA oxidation and CAG instability; for example, oxidative lesions within the *HTT* CAG repeat tract itself were not age-dependent, nor did they correlate with tissue-specific instability in HD mice [183]. In support of this, the extent to which disease pathogenesis itself, including the dysregulation of DNA repair processes that occurs in HD [198–200], contributes to somatic CAG expansion is unclear. Rather, there is evidence to support the idea that the primary driver(s) of somatic CAG expansion are independent of disease manifestation. Indeed, mouse models for clinically distinct diseases show similar tissue-specific patterns of instability [87, 109, 115, 201] and genetic and bioinformatic analyses in *Htt*^{Q111} mice indicated that the susceptibility of a tissue to CAG expansion was independent of the ongoing pathogenic process [87]. Taken together, therefore, data provide support that the repair of oxidative DNA lesions within the repeat tract can play a role in CAG instability but that they may not be necessary to drive expansion.

NUCLEOTIDE EXCISION REPAIR FACTORS AND TRANSCRIPTION-BASED MECHANISMS

Nucleotide excision repair (NER) removes bulky DNA lesions, including thymine dimers and 6-4 photoproducts [202]. The initial steps of NER differ

depending on how the lesion is detected. In cycling cells, the global genome repair branch of NER (GG-NER) detects lesions and leads to the recruitment of TFIIH. By contrast, lesions in transcribed regions will activate the transcription-coupled branch of NER (TC-NER) by causing the RNA polymerase to stall, CSB and CSA to be recruited, followed by TFIIH. The pathways merge with XPA binding to the lesion, followed by helicases and nucleases that remove 24–32 nucleotides around the lesion, leaving a DNA gap, which is filled by a DNA polymerase and ligated.

TC-NER was first found to be involved in repeat instability in immortalized human HT1080 cells that harbored a selectable reporter system to monitor exclusively large contractions [67, 68]. Knockdown of *CSA*, *CSB*, *XPA*, *ERCC1*, and *XPG* all reduced the frequency of large contractions in this system. In a *Drosophila* model of CAG/CTG intergenerational instability, a null allele in the *XPG* ortholog *mus201* suppressed expansions and contractions [203]. It is worth pointing out that CAG instability in *Drosophila* models has not been shown to be dependent on MMR genes, raising the possibility that some of the mechanisms of instability in *Drosophila* may not be shared with mammalian systems. Subsequently, using a mouse model for spinocerebellar ataxia type 1 (SCA1), knockout of *Xpa* was shown to suppress somatic CAG expansion [108]. Surprisingly, *Xpa* knockout had a tissue-specific effect whereby, of the tissues analyzed, only the neuronal tissues had stabilized repeats [108]. In contrast to *Xpa*, knocking out *Xpc*, which is specifically involved in GG-NER, did not impact either somatic or intergenerational instability in *Htt*^{Q111} mice [88]. This is consistent with the lack of impact of *XPC* knockdown in a selectable contraction-based reporter assay [67]. CSB was also implicated in repeat instability in R6/1 mice, but only had an effect on intergenerational instability when OGG1 was also simultaneously knocked out [101], again suggesting possible cross-talk between DNA repair pathways.

A role of TC-NER is in line with observations that transcription enhances CAG/CTG instability [67, 203, 204]. It is also notable that many NER proteins are part of the core transcription machinery [205]. In transgenic R6 mouse models with distinct sites of transgene integration it was suggested that transcription is necessary for repeat instability [111]. However, transcription does not necessarily account for differences in instability of a particular repeat in different chromatin environments [121, 124]. Further, steady state mRNA level does not correlate with

tissue-specific instability across tissues [87, 110, 125], indicating a potentially more complex interaction between transcriptional state and instability. It was suggested that transcription elongation may better determine somatic expansion than transcription initiation in HD mouse models [182]. In support of this, transcription elongator factor TFIIS (*TCEA2*) promoted instability in cell-based assays [68, 79], and downregulation of *BRCA1/BARD1*, a ubiquitin E3 ligase that regulates transcriptional elongation, suppressed instability [68]. Instability is also enhanced by bi-directional transcription through repeat tracts in cell-based models [71, 204]. As disease-associated repeats are transcribed in both sense and antisense directions (reviewed in [206]), this also provides a plausible mechanism by which transcription could modulate instability at endogenous repeat loci in patient cells or mouse models, though this remains to be tested.

Transcription through the repeat tract can also lead to R-loops, stable RNA-DNA hybrids that form by annealing of the newly synthesized RNA transcript to the DNA template. They are favored by G-rich RNA and by secondary structure-forming sequences in the displaced single stranded DNA, such as CAG/CTG repeats [32, 207–209]. R-loops are thought to be removed by either NER or dissolved by the action of RNase H enzymes. Indeed, the latter have been implicated in large contractions in human cells, in which knockdown of *RNaseH1* or *RNaseH2A* stimulated instability [71]. Moreover, in a cell-free system, human cell extracts from HeLa and SH-SY5Y cells could process R-loops, promoting repeat instability, an effect that was partially suppressed by treatment with RNase H [207, 208]. In HT 1080 cells harboring 800 CTG repeats, knockdown of *SETX* (Senataxin), a putative RNA/DNA helicase with a function in resolving R-loops, stimulated repeat contractions [79]. Interestingly, R-loop mediated CAG/CTG repeat instability in *S. cerevisiae* was found to be dependent both on cytosine deamination and BER, and on MutL γ nuclease activity [210]. These studies suggest that there may be multiple pathways by which R-loops at CAG/CTG tracts can be processed.

Although not thoroughly tested, transcription-based mechanisms are attractive in the context of post-mitotic neurons as transcription-associated repair of DNA lesions remains active in the absence of replication [211]. Potential intersection with R-loop biology, TC-NER, and other DNA repair pathways is an interesting area of further investigation.

ROLES OF ADDITIONAL DNA REPAIR OR REPLICATION FACTORS

In contrast to the involvement of MMR, BER, and NER-related mechanisms in various aspects of CAG/CTG instability, there has been little support for a prominent role of double strand break (DSB) repair mechanisms in repeat instability in mammalian cells. Although *RAD51*, involved in homologous recombination (HR), protected against large contractions [73], there was no impact of knocking out HR genes *Rad54* and *Rad52* on CTG somatic expansion in a DM1 mouse model [106]. *Rad52*, but not *Rad54* knockout did increase the size of intergenerational expansions in this model, potentially implicating single-strand annealing (SSA) mechanism of DSB repair, in which *RAD52* is also involved. Non-homologous end-joining (NHEJ) is the most commonly used DSB repair pathway in mammals. This has not been studied in depth in the context of repeat instability; however, knockout of NHEJ gene *DNAPKcs* had no impact on somatic or intergenerational instability in DM1 transgenic mice [106]. *CTIP* and *MRE11*, genes involved in the sensing and signaling of DSB repair, did not modify CAG expansion in a human selectable assay [76]. *BRCA1*, which enhanced contractions in a selectable assay [68], has diverse roles that include transcription, coordinating repair of DBSs, HR, as well as ICL repair [212]. Given the multifunctional nature of *BRCA1*, it is plausible that in the unusual context of expanded repeats, this protein might also intersect with other DNA repair mechanisms.

Several modifier genes have been identified that influence CAG/CTG instability in the context of DNA replication. Replication-based mechanisms may be more important in rapidly dividing cells, including the germline and during early embryonic development. In human cells, the *RTEL1* helicase blocked repeat expansions [74]. It is proposed to act via its hairpin unwinding activity, together with *HLTF*, a nucleosome remodeling factor and *RAD18*, an E3 ubiquitin ligase, both of which suppress expansions [74]. *RTEL1* appears to be a functional homolog of the *Srs2* helicase, originally found in *S. cerevisiae* to inhibit repeat expansion [213], and for which there are no sequence homologs in metazoans. It was proposed that *RTEL1/HLTF/RAD18* acts to prevent repeat expansions during post-replication repair mechanisms that allow replication forks to progress through lesions on damaged templates [74]. In a HeLa cell-based model containing an ectopic integrated

cassette that includes the *c-myc* replication origin and CAG/CTG repeats, as well as in DM1 fibroblasts, knockdown of one of *CLSPN*, *TIMELESS*, or *TIPIN*, which all play a role in replication stress and fork stabilization, resulted in a substantial increase in repeat contractions [80]. Whether any of the factors that appear to sense or resolve CAG/CTG-mediated replication stress play a role in modulating somatic or intergenerational repeat instability in mouse models is unknown.

CIS ELEMENTS, CHROMATIN, AND POST TRANSLATIONAL MODIFICATIONS IN REPEAT INSTABILITY

Some instability modifiers provide insight into ways in which DNA metabolic processes that influence repeat instability might be regulated. In addition to factors that act *in trans*, there is evidence that the instability of CAG/CTG repeats can be modulated *in cis* by DNA elements other than repeat length, repeat sequence and purity (reviewed in [206, 214]). For example, the expansion propensity of different disease-associated CAG/CTG repeats correlates with the GC-content of the flanking DNA [123, 215, 216]. Interestingly, there is an association between ancestral *HTT* haplotype and the mean length of the unexpanded CAG repeat, leading to the speculation that certain *HTT* haplotypes may be predisposed to CAG instability due to *cis*-acting elements [217]. In mouse models, the genomic contexts of transgenes or knock-in alleles can modulate instability [111, 115, 119, 121, 123–125]; notably, repeats flanked by genomic DNA are more unstable than those within cDNA transgenes [123, 124]. Insights into the precise nature of the *cis*-factors and the mechanisms by which they modulate repeat instability are currently very limited. Chromatin context regulates DNA repair, transcription and replication [218–220], and chromatin marks, including trimethylation of histone H3 at Lys36 (H3K36me3), and chromatin-remodeling factors have recently been shown to regulate MMR (reviewed in [221]), indicating possible routes for modification of CAG instability *in cis*. Replication fork dynamics, which can influence instability, are also influenced by *cis*-factors [222].

Highly expanded trinucleotide repeats tend to be associated with epigenetic marks of heterochromatin such as DNA methylation and methylation of histone H3 at Lys9 (H3K9) (reviewed in [206]). Greater levels of DNA methylation at the *DMPK*

CTG repeat locus were also found in tissues exhibiting greater somatic expansion [223]. Interestingly, disease-associated repeat loci were shown to localize to topology-associated domains (TAD) and subTAD chromatin boundaries [224]. CGG repeat expansion at the *FMR1* locus disrupted this boundary [224]. However, in a different study, 4C sequencing of the *HTT* and *DMPK* loci did not identify any changes in chromatin structure in the presence of expanded repeats [225]. The extent to which local epigenetic changes associated with expanded repeats might themselves contribute to repeat instability is unknown. In model systems global demethylation destabilized repeats [147, 226], and *Dnmt1* deficiency promoted germline CAG instability in a mouse model of SCA1 [70]. In these cases, instability modification may be due to local DNA methylation changes, to the altered expression of other genes that modulate instability *in trans*, or both. More direct evidence for a *cis*-modifier effect comes from SCA7 mouse models in which mutation of a CTCF binding site adjacent to the *Atxn7* CAG repeat promoted instability in specific tissues, notably in the kidney [124]. A similarly high level of expansion in a kidney was observed in a mouse with a wild type, but methylated, CTCF binding site. These results suggested a model in which CTCF protects against repeat expansion in a methylation-dependent manner [124].

Other proteins with chromatin-modifying activities also influence repeat instability. In a selectable assay for CAG expansions in cultured astrocytes, knockdown of histone deacetylase genes *HDAC3* or *HDAC5* suppressed expansions, whereas knockdown of histone acetyltransferase (HAT) gene *CREBBP*, encoding CREB-binding protein (CBP) or *EP300*, encoding the related P300 protein, promoted expansions [75, 76]. Loss of CBP also promoted CAG expansions in the *Drosophila* germline [203]. These genetic data are supported by the suppression of instability upon treatment with the Class I/II HDAC inhibitor trichostatin A (TSA) [203]. Genetic knockout of either *Hdac2* or *Hdac3* in MSNs of *Htt*^{Q111} knock-in mice moderately suppressed striatal expansions [91]. The impact of *Hdac3* knockout in this model is consistent with the expansion-suppressing effect of a selective HDAC3 inhibitor [227]. Relationships between HDACs, HATs and repeat instability appear to be complex, however. For example, knockdown of *HDAC9* promoted, rather than suppressed, CAG expansion in the cultured astrocyte model [76], and HDAC inhibitors promoted contractions in a selectable human cell-based assay [91]. In *S. cerevisiae*,

loss of function of different HDACs either enhanced or suppressed CAG instability [75, 228], with CAG stability being dependent on activities of both HDACs and HATs that control levels of H4 acetylation [228]. As in the case of DNA methylation, HDACs or HATs may modify repeat instability via epigenetic changes *in cis* and/or by altering the expression or activity of *trans*-acting factors. A *trans*-mediated effect was proposed to account for the impact of CBP loss in *Drosophila* [203]. In human astrocytes, epistasis experiments provided evidence that *HDAC3* and *MSH2* act in a common pathway to promote expansion [76]. Further molecular experiments indicated that *HDAC3* knockdown did not alter the binding of MutS β at the repeat tract, or the expression level of *MSH2* or *MSH3*, together suggesting an alternative hypothesis that *HDAC3* modulates instability by regulating *MSH2* or *MSH3* acetylation. Indeed, activities of both *MSH2* and *MLH1* can be altered by acetylation [229–231], indicating that this is a plausible mechanism for instability modification by HDACs and HATs. Recent studies show that *MSH3* peptides can be deacetylated by *HDAC3* and that the nuclear localization of *MSH3* is regulated by a selective *HDAC3* inhibitor [232]. Roles of histone modification *in cis* contributing to instability have not been demonstrated directly in mammalian systems; however, in *S. cerevisiae* a protective effect of chromatin remodeler *Isw1* against CAG expansion during transcription could be attributed at least in part to a direct effect at the repeat tract as nucleosome occupancy over the CAG tract was altered in *isw1* mutants [233].

CELLULAR STRESSORS IN REPEAT INSTABILITY

Cellular stresses can also induce changes in repeat instability. DNA damaging agents could work by inducing damage to the repeat tract directly (see section below), indirectly causing cellular stress, or both. For instance, Gorbunova et al. [132] treated cells with hydroxyurea, which depletes dNTP pools and slows DNA replication, ionizing radiation, which causes DNA breaks, or aphidicolin, an inhibitor of DNA polymerase α . In this early example, these treatments induced large contractions, with unknown effects on expansions. Other studies in cell-based models have found effects on CAG/CTG instability of other DNA damaging compounds that include DNA intercalating agents, nucleoside analogues, and hydrogen peroxide [72, 136, 234]. These experiments can often

be difficult to interpret due to effects of DNA damaging compounds on the cell cycle and potential clonal selection that could artifactually alter the distribution of CAG-containing alleles in a population without modifying repeat instability. Other types of cellular stresses that alter repeat instability include cold and heat shock [135], and pharmacological and genetic suppression of the proteasome [68, 77] or of HSP90 [73]. Cellular stressors could act indirectly by altering levels or activities of DNA repair proteins that control repeat instability and may provide a means to stimulate repeat instability to facilitate analyses in mammalian cell-based models where the natural time-course of somatic expansion is slow.

DIRECT PERTURBATION OF THE REPEAT TRACT

Alternative routes to modifying repeat instability involve direct perturbation of the repeat tract. Of the small molecule studies conducted thus far [72, 75, 132, 136, 203, 227, 235, 236], only one compound has been shown to engage directly the repeat tract at the DNA level. Indeed, Nakamori et al. [236], found that naphthyridine-azaquinolone (NA) can bind to CAG loops, stabilize them, and induce contractions and/or prevent expansions in human cell culture as well as in the R6/2 HD mouse model. Overall, there has been no small molecule that is non-toxic and efficiently prevents expansions or induces contractions at doses that are physiologically relevant. It will therefore be exciting to see whether NA can be turned into a viable therapeutic. Antisense oligonucleotides (ASOs) are being used in pre-clinical and clinical settings to reduce the toxic huntingtin protein or the toxic *DMPK* mRNA [237–239]. Interestingly, an ASO against the repeat tract itself was seen to reduce repeat instability through an unknown mechanism, though direct binding to the repeat is one possibility [238].

Directly damaging the repeat tract, by inducing a DSB within or very near it has been used to induce instability in yeast and mammalian cells [81, 130, 134, 240–245]. These studies have shown that homologous recombination and the DNA damage response are both involved in generating repeat instability. It is not clear that the same modifiers that affect repeat instability in the absence of a direct assault on the repeat tract, will be involved in these more artificial systems. Indeed, as pointed out above, several factors involved in DSB repair do not seem to influence repeat instability in mammalian models [106]. Never-

theless, directly damaging repeats may be harnessed for gene editing and may therefore have therapeutic relevance.

The first generation of customizable gene editing tools were Zinc Finger Nucleases (ZFNs). ZFNs consist of two FokI nuclease domains that must come together to induce a DSB at the target sequence [246]. Both halves are fused to customizable DNA binding domains, each targeting a DNA triplet. In the context of expanded repeats, they were first used in a chromosomal reporter assay that measures large contractions [245]. They led to a 14-fold increase in the number of contractions and induced rearrangements near the repeat tract. Another group, using expanded repeats inserted in the genome of HeLa cells, also found that they efficiently induced contractions [81]. Interestingly, in both studies, one of the two halves of the ZFN was shown to induce instability on its own. The first study suggested that one of the studies could not distinguish effectively between the CAG and CTG strands and thus cut both [245]. The authors of the second study rather postulated that the ZFN could cut out the secondary structures formed by the repeat tract, which led to the instability [81].

Transcription activator-like effectors (TALEs) were the basis for the second generation of programmable nucleases. They have the advantage that each DNA binding domain binds a specific base pair, making their design easier, although repetitive [247, 248]. Fusing the FokI nuclease to them as with the ZFNs allows for precise induction of a DSB in the human genome [249]. TALENs have been used against expanded CAG/CTG repeats in yeast to great effect [240, 241], but they are yet to be tried in mammalian systems.

CRISPR-Cas9 as a gene editing system was described in 2012 and has since revolutionized the way we manipulate genomes [250]. There are two major advantages of the CRISPR systems over other programmable nucleases. One is efficiency, the other is the ease of design. Indeed, Cas9, a nuclease, is guided to a sequence of choice using a single guide RNA (sgRNA) that contains a target sequence of choice [250, 251]. The most widely used Cas9, from *Staphylococcus pyogenes*, requires a protospacer adjacent motif (PAM) composed of three nucleotides, NGG, where N is any nucleotide, required abutting the target sequence [252]. Thus, any sequence in the genome that abuts a usable PAM can be targeted and cut. Cas9, directed to the repeat tract itself, has been used to induce DSBs within the repeat tract in the context of a chromosomal reporter [134]. This led to

a modest amount of expansions and contractions in roughly equal proportion, perhaps due to the suboptimal CAG and CTG PAMs. Interestingly, inducing a DSB near, but not within, the repeat tract led to a large increase in repeat instability, including expansions, contractions, and rearrangements, in immortalized mouse myoblasts from a transgenic DM1 mouse model containing 500 CTG repeats [253] as well as in yeast [254]. Together, studies inducing DSBs within the repeat tract suggest that they cause both expansions and contractions as well as larger rearrangements near the repeat tract.

The use of the CRISPR system is ideal to test whether different types of DNA damage induced within the repeat tract could lead to changes in repeat instability. This was tried by Cinesi et al. using a D10A mutant of the Cas9 enzyme [134]. This mutation turns Cas9 into a nickase, which cuts only one of the two strands of DNA, i.e., it introduces nicks. In this case, the same sgRNA was used, targeting the CAG/CTG repeat itself, enabling the comparison between DSBs and nicks. Surprisingly, the nickase induced predominantly contractions of the repeat tract, with up to a third of the cells having a contraction after 12 days of treatment. There were no rearrangements seen beyond those observed in the untreated population. Moreover, the effect of the nickase was repeat length-dependent with mildly expanded or short alleles not being targeted. In line with this, the number of off-target mutations at other endogenous, non-pathogenic, CAG/CTG repeat loci was below the detection level of the assay used [134]. The nickase remains to be tested in other systems, however.

CONCLUSIONS AND PERSPECTIVES

Through a variety of approaches and model systems considerable progress has been made in delineating factors that can modify the instability of CAG/CTG repeats. These include DNA repair proteins that act in pathways to enhance or suppress instability and reagents that interact directly with the repeat tract. Emerging data hint at possible mechanisms by which DNA repair pathways might be regulated, for example through interaction with *cis*-elements or by altering levels of DNA repair proteins or their activity, e.g., via post-translational modification. This is an interesting area of future investigation. Means by which repeat instability could be modulated are summarized in Fig. 4, indicating plausible routes for therapeutic intervention.

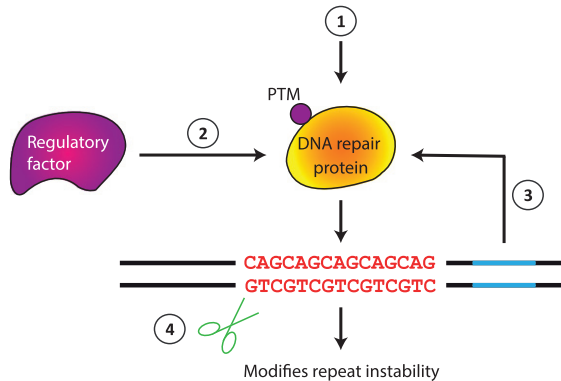


Fig. 4. Potential mechanisms for instability modifiers. Genetic or pharmacological manipulation can directly alter the level of activity of a DNA repair protein (1). Indirect effects are also possible via a regulatory factor that controls DNA repair gene expression, stability or activity (2), e.g., altering a posttranslational modification (PTM), or via interaction with a cis-element (3). The repeat can also be modified by agents that directly interact with the repeat (4).

The mechanism(s) underlying repeat instability, however, remain poorly understood, but they are important to delineate in order to refine therapeutic approaches. For example, if DNA repair proteins act in non-canonical ways to modify repeat instability, this may provide opportunities for interventions that specifically target the repeat instability mechanism, whilst minimizing the impact on DNA repair processes more globally. Relatively few of the modifiers uncovered in simpler cell-based assays have been tested in mouse models, and their significance in human disease is unclear. Although a process dependent on MMR genes is clearly critical for somatic CAG/CTG expansion in mouse models, potential intersections with factors in other DNA repair pathways such as BER or NER are not well understood. We highlight several questions and approaches to be addressed in future work.

First, how do mechanisms of instability differ between cell types, for example between the germline and soma, or between the brain and periphery? A number of studies indicate that mechanisms may not be the same [57, 70, 88, 92, 100, 101, 105, 106, 108, 255], and although many proteins function in repeat instability across cell types, factors may be specifically or preferentially involved, in a cell type-dependent manner. Several factors have been proposed to contribute to the tissue/cell type-specificity patterns of somatic expansion (reviewed in [109]), however the molecular underpinnings are unknown. Understanding the nature of these intrinsic cell type-specific instability drivers will also provide clues to

underlying mechanisms. Differences in mechanism between post-mitotic and dividing cells are also likely, and in the germline where instability can arise at multiple stages during spermatogenesis in humans [6], processes in both dividing and non-dividing cells may be relevant. Second, although much has been learned about factors involved in somatic expansion, we know surprisingly little about repeat contraction. This is a significant question to address as promoting repeat contraction would clearly be an important therapeutic goal. Third, the majority of the genetic studies in model systems involve the knockdown or knockout of specific genes. Mechanistic insight will be enhanced by further studying the impact of functional variation, for example to test the roles of enzymatic activities of DNA repair proteins [78, 82, 95]. Of particular significance are the functional variants discovered through human GWAS that will inform on ways in which instability can be modulated that, by definition, would be predicted to have an impact in patients. Fourth, what are the substrate(s) recognized by the DNA repair machinery *in vivo*? For example, is the substrate for MMR proteins simply an insertion loop that is recognized by the MutS β complex in the same way this complex would recognize any other insertion loop, or does MutS β interact with a CAG/CTG substrate in a distinct manner? Finally, there is currently a lack of connection between biochemical assays that are conducted in cell free systems and instability that occurs within cells. Developing systems to bridge this gap that would enable structure/function studies would be of value.

Novel and emerging technologies, including CRI SPR-based systems, single cell analyses and next-generation sequencing will enable the development of additional platforms for modifier screening/testing and a more comprehensive knowledge of the pathways and mechanisms that control repeat dynamics. Further, as current methods provide a steady state snapshot of repeat instability, innovative methodologies, including computational modeling, will be needed to capture the dynamic nature of the unstable repeats [256, 257]. Ultimately, a greater understanding of repeat instability will provide additional therapeutic avenues in HD, DM1, and other repeat expansion disorders.

ACKNOWLEDGMENTS

We thank Emma L. Randall and Alysha Taylor for help with Fig. 2. Oscar Rodríguez-Lima provided

the small-pool PCR figure and Ricardo Mouro Pinto provided the GeneMapper trace, both found in Fig. 3. Work in V.D.'s lab is supported by the Academic of Medical Sciences (AMSPR1\1014) and by the UK Dementia Research Institute, which receives its funding from DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK. V.C.W is supported by grants from the National Institutes of Health, the CHDI Foundation, Pfizer Inc. and the CCXDP, with previous support from the Huntington's Society of Canada.

CONFLICT OF INTEREST

V.C.W. is a scientific advisory board member of Triplet Therapeutics, a company developing new therapeutic approaches to address triplet repeat disorders such Huntington's disease and Myotonic Dystrophy. Her financial interests in Triplet Therapeutics were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. She is a scientific advisory board member of LoQus23 Therapeutics and has provided paid consulting services to Alnylam and Acadia Pharmaceuticals. V.D. is a named inventor on patent application WO2017178590A1 involving the use of the Cas9 nickase for gene editing of expanded CAG/CTG repeat disorders.

REFERENCES

- [1] Monckton DG. Somatic expansion of the CAG repeat in Huntington disease: An historical perspective. *J Huntingtons Dis.* 2020;doi: 10.3233/JHD-200429
- [2] Mouro Pinto R, Arning L, Giordano JV, Razghandi P, Andrew MA, Gillis T, et al. Patterns of CAG repeat instability in the central nervous system and periphery in Huntington's disease and in spinocerebellar ataxia type 1. *Hum Mol Genet.* 2020;29(15):2551-67.
- [3] Huntington's disease collaborative research group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell.* 1993;72(6):971-83.
- [4] Swami M, Hendricks AE, Gillis T, Massood T, Mysore J, Myers RH, et al. Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum Mol Genet.* 2009;18(16):3039-47.
- [5] Geraerts FCA, Snell RG, Faull RLM, Williams L, Jacobsen JC, Reid SJ. Comparison of Huntington's disease CAG repeat length stability in human motor cortex and cingulate gyrus. *J Huntingtons Dis.* 2016;5(3):297-301.
- [6] Yoon SR, Dubeau L, De Young M, Wexler NS, Arnheim N. Huntington disease expansion mutations in humans can occur before meiosis is completed. *Proc Natl Acad Sci U S A.* 2003;100(15):8834-8.
- [7] Chong SS, McCall AE, Cota J, Subramony SH, Orr HT, Hughes MR, et al. Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet.* 1995;10(3):344-50.
- [8] Watanabe M, Satoh A, Kanemoto M, Ohkoshi N, Shoji S. De novo expansion of a CAG repeat in a Japanese patient with sporadic Huntington's disease. *J Neurol Sci.* 2000;178(2):159-62.
- [9] Cancel G, Gourfinkel-An I, Stevanin G, Didierjean O, Abbas N, Hirsch E, et al. Somatic mosaicism of the CAG repeat expansion in spinocerebellar ataxia type 3/Machado-Joseph disease. *Hum Mutat.* 1998;11(1):23-7.
- [10] Martorell L, Gámez J, Cayuela ML, Gould FK, McAbney JP, Ashizawa T, et al. Germline mutational dynamics in myotonic dystrophy type 1 males: Allele length and age effects. *Neurology.* 2004;62(2):269-74.
- [11] Martorell L, Johnson K, Boucher CA, Baiget M. Somatic instability of the myotonic dystrophy (CTG)(n) repeat during human fetal development. *Hum Mol Genet.* 1997;6(6):877-80.
- [12] Martorell L, Monckton DG, Gamez J, Baiget M. Complex patterns of male germline instability and somatic mosaicism in myotonic dystrophy type 1. *Eur J Hum Genet.* 2000;8(6):423-30.
- [13] Monckton DG, Wong LJC, Ashizawa T, Caskey CT. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: Small pool PCR analyses. *Hum Mol Genet.* 1995;4(1):1-8.
- [14] Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, et al. Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat Genet.* 1993;4(4):387-92.
- [15] Harley HG, Rundle SA, Reardon W, Myring J, Crow S, Harper PS, et al. Unstable DNA sequence in myotonic dystrophy. *Lancet.* 1992;339(8802):1125-8.
- [16] Wong LJC, Ashizawa T, Monckton DG, Caskey CT, Richards CS. Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am J Hum Genet.* 1995;56(1):114-22.
- [17] Brunner HG, Bruggenwirth HT, Nillesen W, Jansen G, Hamel BCJ, Hoppe RLE, et al. Influence of sex of the transmitting parent as well as of parental allele size on the CTG expansion in myotonic dystrophy (DM). *Am J Hum Genet.* 1993;53(5):1016-23.
- [18] Ashizawa T, Dubel JR, Dunne PW, Fu YH, Pizzuti A, Caskey CT, et al. Anticipation in myotonic dystrophy: II. complex relationships between clinical findings and structure of the gct repeat. *Neurology.* 1992;42(10):1877-83.
- [19] Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, et al. Myotonic dystrophy mutation: An unstable CTG repeat in the 3' untranslated region of the gene. *Science.* 1992;255(5049):1253-5.
- [20] Aslanidis C, Jansen G, Amemiya C, Shutler G, Mahadevan M, Tsilfidis C, et al. Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature.* 1992;355(6360):548-51.
- [21] Redman JB, Fenwick RG, Fu YH, Pizzuti A, Caskey CT. Relationship between parental trinucleotide GCT repeat length and severity of myotonic dystrophy in offspring. *JAMA J Am Med Assoc.* 1993;269(15):1960-5.
- [22] Schmidt MHM, Pearson CE. Disease-associated repeat instability and mismatch repair. *DNA Repair (Amst).* 2016;38:117-26.

- [23] Leeftang EP, Zhang L, Tavaré S, Hubert R, Srinidhi J, MacDonald ME, et al. Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: Quantification of the mutation frequency spectrum. *Hum Mol Genet.* 1995;4(9):1519-26.
- [24] Telenius H, Kremer B, Goldberg YP, Theilmann J, Andrew SE, Zeisler J, et al. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat Genet.* 1994;6(4):409-14.
- [25] Zühlke C, Rless O, Bockel B, Lange H, Thies U. Mitotic stability and meiotic variability of the (CAG)_n repeat in the huntington disease gene. *Hum Mol Genet.* 1993;2(12):2063-7.
- [26] Wheeler VC, Persichetti F, McNeil SM, Mysore JS, Mysore SS, MacDonald ME, et al. Factors associated with HD CAG repeat instability in Huntington disease. *J Med Genet.* 2007;44(11):695-701.
- [27] Kennedy L, Evans E, Chen CM, Craven L, Detloff PJ, Ennis M, et al. Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum Mol Genet.* 2003;12(24):3359-67.
- [28] Veitch NJ, Ennis M, McAbney JP, Shelbourne PF, Monckton DG. Inherited CAG-CTG allele length is a major modifier of somatic mutation length variability in Huntington disease. *DNA Repair (Amst).* 2007;6(6):789-96.
- [29] Shelbourne PF, Keller-McGandy C, Bi WL, Yoon SR, Dubeau L, Veitch NJ, et al. Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Hum Mol Genet.* 2007;16(10):1133-42.
- [30] Gonitell R, Moffitt H, Sathasivam K, Woodman B, Detloff PJ, Faull RLM, et al. DNA instability in postmitotic neurons. *Proc Natl Acad Sci U S A.* 2008;105(9):3467-72.
- [31] Pearson CE, Nichol Edamura K, Cleary JD. Repeat instability: Mechanisms of dynamic mutations. *Nat Rev Genet.* 2005;6(10):729-42.
- [32] Khristich AN, Mirkin SM. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J Biol Chem.* 2020;295(13):4134-70.
- [33] Su XA, Dion V, Gasser SM, Freudenreich CH. Regulation of recombination at yeast nuclear pores controls repair and triplet repeat stability. *Genes Dev.* 2015;29(10):1006-17.
- [34] Castel AL, 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(3):165-70.
- [35] Usdin K, House NCM, Freudenreich CH. Repeat instability during DNA repair: Insights from model systems. *Crit Rev Biochem Mol Biol.* 2015;50(2):142-67.
- [36] Bates GP, Dorsey R, Gusella JF, Hayden MR, Kay C, Leavitt BR, et al. Huntington disease. *Nat Rev Dis Prim.* 2015;1(1):15005.
- [37] Nance MA, Myers RH. Juvenile onset Huntington's disease—clinical and research perspectives. *Ment Retard Dev Disabil Res Rev.* 2001;7(3):153-7.
- [38] Semaka A, Kay C, Doty C, Collins JA, Bijlsma EK, Richards F, et al. CAG size-specific risk estimates for intermediate allele repeat instability in Huntington disease. *J Med Genet.* 2013;50(10):696-703.
- [39] Dogan C, De Antonio M, Hamroun D, Varet H, Fabbro M, Rougier F, et al. Gender as a modifying factor influencing myotonic dystrophy type 1 phenotype severity and mortality: A nationwide multiple databases cross-sectional observational study. *PLoS One.* 2016;11(2):e0148264.
- [40] De Antonio M, Dogan C, Hamroun D, Mati M, Zerrouki S, Eymard B, et al. Unravelling the myotonic dystrophy type 1 clinical spectrum: A systematic registry-based study with implications for disease classification. *Rev Neurol (Paris).* 2016;172(10):572-80.
- [41] Tomé S, Gourdon G. DM1 phenotype variability and triplet repeat instability: Challenges in the development of new therapies. *Int J Mol Sci.* 2020;21(2):457.
- [42] Imbert G, Kretz C, Johnson K, Mandel JL. Origin of the expansion mutation in myotonic dystrophy. *Nat Genet.* 1993;4(1):72-6.
- [43] Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, et al. Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell.* 1992;68(4):799-808.
- [44] Barbé L, Lanni S, López-Castel A, Franck S, Spits C, Keymolen K, et al. CpG methylation, a parent-of-origin effect for maternal-biased transmission of congenital myotonic dystrophy. *Am J Hum Genet.* 2017;100(3):488-505.
- [45] Jinnai K, Mitani M, Futamura N, Kawamoto K, Funakawa I, Itoh K. Somatic instability of CTG repeats in the cerebellum of myotonic dystrophy type 1. *Muscle Nerve.* 2013;48(1):105-8.
- [46] Zhao X, Kumari D, Miller C, Kim G, Hayward B, Vitalo AG, et al. Modifiers of somatic repeat instability in mouse models of Friedreich ataxia and the Fragile X-related disorders: Implications for the mechanism of somatic expansion in Huntington disease. *J Huntingtons Dis.* 2020;doi: 10.3233/JHD-200423
- [47] Iyer R, Pluciennik A. DNA mismatch repair and its role in Huntington's disease. *J Huntingtons Dis.* 2020;doi: 10.3233/JHD-200438
- [48] Lee J-M, Chao MJ, Harold D, Abu Elneel K, Gillis T, Holmans P, et al. A modifier of Huntington's disease onset at the MLH1 locus. *Hum Mol Genet.* 2017;26(19):3859-67.
- [49] Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell.* 2015;162(3):516-26.
- [50] Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. CAG repeat not polyglutamine length determines timing of Huntington's disease onset. *Cell.* 2019;178(4):887-900.e14.
- [51] Hong E, MacDonald M, Wheeler V, Jones L, Holmans P, Orth M, et al. Huntington's disease pathogenesis: Two sequential components. *J Huntingtons Dis.* 2020;doi: 10.3233/JHD-200427
- [52] Moss DJH, Pardinás AF, Langbehn D, Lo K, Leavitt BR, Roos R, et al. Identification of genetic variants associated with Huntington's disease progression: A genome-wide association study. *Lancet Neurol.* 9AD;16(9):701-11.
- [53] Kantidze OL, Velichko AK, Luzhin AV, Petrova NV, Razin SV. Synthetically lethal interactions of ATM, ATR, and DNA-PKcs. *Trends Cancer.* 2018;4(11):755-68.
- [54] Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann Neurol.* 2016;79(6):983-90.
- [55] Morales F, Vásquez M, Santamaría C, Cuenca P, Corrales E, Monckton DG. A polymorphism in the MSH3 mismatch repair gene is associated with the levels of somatic instability of the expanded CTG repeat in the

- blood DNA of myotonic dystrophy type 1 patients. *DNA Repair (Amst)*. 2016;40:57-66.
- [56] Flower M, Lomeikaite V, Ciosi M, Cumming S, Morales F, Lo K, et al. MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain*. 2019;142(7):1876-86.
- [57] Ciosi M, Maxwell A, Cumming SA, Hensman Moss DJ, Alshammari AM, Flower MD, et al. A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine*. 2019;48:568-80.
- [58] Benn C, Gibson K, Reynolds D. Drugging DNA damage repair pathways for trinucleotide repeat expansion diseases. *J Huntingtons Dis*. 2020;doi: 10.3233/JHD-200421
- [59] Slean MM, Panigrahi GB, Ranum LP, Pearson CE. Mutagenic roles of DNA "repair" proteins in antibody diversity and disease-associated trinucleotide repeat instability. *DNA Repair (Amst)*. 2008;7(7):1135-54.
- [60] Peña-Díaz J, Bregenhorn S, Ghodgaonkar M, Follonier C, Artola-Borán M, Castor D, et al. Noncanonical mismatch repair as a source of genomic instability in human cells. *Mol Cell*. 2012;47(5):669-80.
- [61] Kovtun IV, McMurray CT. Crosstalk of DNA glycosylases with pathways other than base excision repair. *DNA Repair (Amst)*. 2007;6(4):517-29.
- [62] Mouw KW, D'Andrea AD. Crosstalk between the nucleotide excision repair and Fanconi anemia/BRCA pathways. *DNA Repair (Amst)*. 2014;19:130-4.
- [63] Limpose KL, Corbett AH, Doetsch PW. BERing the burden of damage: Pathway crosstalk and posttranslational modification of base excision repair proteins regulate DNA damage management. *DNA Repair (Amst)*. 2017;56:51-64.
- [64] Melis JPM, Van Steeg H, Luijten M. Oxidative DNA damage and nucleotide excision repair. *Antioxidants Redox Signal*. 2013;18(18):2409-19.
- [65] Risinger MA, Groden J. Crosslinks and crosstalk: Human cancer syndromes and DNA repair defects. *Cancer Cell*. 2004;6(6):539-45.
- [66] Loupe JM, Pinto RM, Kim K-H, Gillis T, Mysore JS, Andrew MA, et al. Promotion of somatic CAG repeat expansion by Fan1 knock-out in Huntington's disease knock-in mice is blocked by Mlh1 knock-out. *Hum Mol Genet*. 2020;doi: 10.1093/hmg/ddaa196
- [67] Lin Y, Dion V, Wilson JH. Transcription promotes contraction of CAG repeat tracts in human cells. *Nat Struct Mol Biol*. 2006;13(2):179-80.
- [68] 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(17):6209-17.
- [69] Lin Y, Wilson JH. Diverse effects of individual mismatch repair components on transcription-induced CAG repeat instability in human cells. *DNA Repair (Amst)*. 2009;8(8):878-85.
- [70] Dion V, Lin Y, Hubert L, Waterland RA, Wilson JH. Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Hum Mol Genet*. 2008;17(9):1306-17.
- [71] Lin Y, Dent SYR, Wilson JH, Wells RD, Napierala M. R loops stimulate genetic instability of CTG-CAG repeats. *Proc Natl Acad Sci U S A*. 2010;107(2):692-7.
- [72] Hubert L, Lin Y, Dion V, Wilson JH. Topoisomerase I and single-strand break repair modulate transcription-induced CAG repeat contraction in human cells. *Mol Cell Biol*. 2011;31(15):3105-12.
- [73] Mittelman D, Sykoudis K, Hersh M, Lin Y, Wilson JH. Hsp90 modulates CAG repeat instability in human cells. *Cell Stress Chaperones*. 2010;15(5):753-9.
- [74] Frizzell A, Nguyen JHG, Petalcorin MIR, Turner KD, Boulton SJ, Freudenreich CH, et al. RTEL1 inhibits trinucleotide repeat expansions and fragility. *Cell Rep*. 2014;6(5):827-35.
- [75] Debacker K, Frizzell A, Gleeson O, Kirkham-McCarthy L, Mertz T, Lahue RS. Histone deacetylase complexes promote trinucleotide repeat expansions. *PLoS Biol*. 2012;10(2):e1001257.
- [76] Gannon AMM, Frizzell A, Healy E, Lahue RS. MutS β and histone deacetylase complexes promote expansions of trinucleotide repeats in human cells. *Nucleic Acids Res*. 2012;40(20):10324-33.
- [77] Concannon C, Lahue RS. The 26S proteasome drives trinucleotide repeat expansions. *Nucleic Acids Res*. 2013;41(12):6098-108.
- [78] Keogh N, Chan KY, Li GM, Lahue RS. MutS β abundance and Msh3 ATP hydrolysis activity are important drivers of CTG-CAG repeat expansions. *Nucleic Acids Res*. 2017;45(17):10068-78.
- [79] Nakatani R, Nakamori M, Fujimura H, Mochizuki H, Takahashi MP. Large expansion of CTG-CAG repeats is exacerbated by MutS β in human cells. *Sci Rep*. 2015; 5:11020.
- [80] Liu G, Chen X, Gao Y, Lewis T, Barthelemy J, Leffak M. Altered replication in human cells promotes DMPK (CTG) $_n$ · (CAG) $_n$ repeat instability. *Mol Cell Biol*. 2012;32(9):1618-32.
- [81] Liu G, Chen X, Bissler JJ, Sinden RR, Leffak M. Replication-dependent instability at (CTG) $_n$ -(CAG) $_n$ repeat hairpins in human cells. *Nat Chem Biol*. 2010;6(9):652-9.
- [82] Goold R, Flower M, Moss DH, Medway C, Wood-Kaczmar A, Andre R, et al. FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. *Hum Mol Genet*. 2019;28(4):650-61.
- [83] Du J, Campau E, Soragni E, Jespersen C, Gottesfeld JM. Length-dependent CTG-CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells. *Hum Mol Genet*. 2013;22(25):5276-87.
- [84] Kim KH, Hong EP, Shin JW, Chao MJ, Loupe J, Gillis T, et al. Genetic and functional analyses point to FAN1 as the source of multiple Huntington disease modifier effects. *Am J Hum Genet*. 2020;107(1):96-110.
- [85] Budworth H, Harris FR, Williams P, Lee DY, Holt A, Pahnke J, et al. Suppression of somatic expansion delays the onset of pathophysiology in a mouse model of Huntington's disease. *PLoS Genet*. 2015;11(8):e1005267.
- [86] Wheeler VC, Lebel LA, Vrbanac V, Teed A, te Riele HT, MacDonald ME. Mismatch repair gene Msh2 modifies the timing of early disease in HdhQ111 striatum. *Hum Mol Genet*. 2003;12(3):273-81.
- [87] Lee J-M, Pinto RM, Gillis T, St Claire JC, Wheeler VC. Quantification of age-dependent somatic CAG repeat instability in Hdh CAG knock-in mice reveals different expansion dynamics in striatum and liver. *PLoS One*. 2011;6(8):e23647.
- [88] Dragileva E, Hendricks A, Teed A, Gillis T, Lopez ET, Friedberg EC, et al. Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol Dis*. 2009;33(1):37-47.

- [89] Kovalenko M, Dragileva E, St Claire J, Gillis T, Guide JR, New J, 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(9):e44273.
- [90] Pinto RM, Dragileva E, Kirby A, Lloret A, Lopez E, St. Claire J, 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(10):e1003930.
- [91] Kovalenko M, Erdin S, Andrew MA, St Claire J, Shaughnessy M, Hubert L, et al. Histone deacetylase knockouts modify transcription, CAG instability and nuclear pathology in Huntington disease mice. *Elife*. 2020;9:e55911.
- [92] Møllersen L, Rowe AD, Illuzzi JL, Hildrestrand GA, Gerhold KJ, Tveterås L, et al. Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice. *Hum Mol Genet*. 2012;21(22):4939-47.
- [93] Owen BAL, Yang Z, Lai M, Gajek M, Badger JD, Hayes JJ, et al. (CAG)_n-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat Struct Mol Biol*. 2005;12(8):663-70.
- [94] Tomé S, Manley K, Simard JP, Clark GW, Slean MM, Swami M, et al. MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. *PLoS Genet*. 2013;9(2):e1003280.
- [95] Tomé S, Holt I, Edelmann W, Morris GE, Munnich A, Pearson CE, et al. MSH2 ATPase domain mutation affects CTG•CAG repeat instability in transgenic mice. *PLoS Genet*. 2009;5(5):e1000482.
- [96] Kovtun IV, Thornhill AR, McMurray CT. Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations. *Hum Mol Genet*. 2004;13(24):3057-68.
- [97] 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(4):471-3.
- [98] Kovtun IV, McMurray CT. Trinucleotide expansion in haploid germ cells by gap repair. *Nat Genet*. 2001;27(4):407-11.
- [99] Spiro C, McMurray CT. Nuclease-deficient FEN-1 blocks Rad51/BRCA1-mediated repair and causes trinucleotide repeat instability. *Mol Cell Biol*. 2003;23(17):6063-74.
- [100] Kovtun IV, Liu Y, Bjoras M, Klungland A, Wilson SH, McMurray CT. OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature*. 2007;447(7143):447-52.
- [101] Kovtun IV, Johnson KO, McMurray CT. Cockayne syndrome B protein antagonizes OGG1 in modulating CAG repeat length *in vivo*. *Aging (Albany NY)*. 2011;3(5):509-14.
- [102] van den Broek WJAA, Wansink DG, Wieringa B. Somatic CTG•CAG repeat instability in a mouse model for myotonic dystrophy type 1 is associated with changes in cell nuclearity and DNA ploidy. *BMC Mol Biol*. 2007;8:61.
- [103] van den Broek WJAA, Nelen MR, van der Heijden GW, Wansink DG, Wieringa B. Fen1 does not control somatic hypermutability of the (CTG)_n·(CAG)_n repeat in a knock-in mouse model for DM1. *FEBS Lett*. 2006;580(22):5208-14.
- [104] Gomes-Pereira M, Fortune MT, Ingram L, McAbney JP, Monckton DG. 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(16):1815-25.
- [105] Foiry L, Dong L, Savouret C, Hubert L, te Riele H, Junien C, et al. Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Hum Genet*. 2006;119(5):520-6.
- [106] Savouret C, Brisson E, Essers J, Kanaar R, Pastink A, Te Riele H, et al. CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J*. 2003;22(9):2264-73.
- [107] Tomé S, Panigrahi GB, Castel AL, Foiry L, Melton DW, Gourdon G, et al. Maternal germline-specific effect of DNA ligase I on CTG/CAG instability. *Hum Mol Genet*. 2011;20(11):2131-43.
- [108] Hubert L, 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(24):4822-30.
- [109] Dion V. Tissue specificity in DNA repair: Lessons from trinucleotide repeat instability. *Trends Genet*. 2014;30(6):220-9.
- [110] Gomes-Pereira M, Foiry L, Gourdon G. Transgenic mouse models of unstable trinucleotide repeats: Toward an understanding of disease-associated repeat size mutation. *Genet Instab Neurol Dis Second Ed*. 2006;563-83.
- [111] Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, et al. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell*. 1996;87(3):493-506.
- [112] Seznec H, Lia-Baldini AS, Duros C, Fouquet C, Lacroix C, Hofmann-Radvanyi H, et al. Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum Mol Genet*. 2000;9(8):1185-94.
- [113] Wheeler VC, Auerbach W, White JK, Srinidhi J, Auerbach A, Ryan A, et al. Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Hum Mol Genet*. 1999;8(1):115-22.
- [114] Watase K, Weeber EJ, Xu B, Antalfy B, Yuva-Paylor L, Hashimoto K, et al. A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration. *Neuron*. 2002;34(6):905-19.
- [115] Fortune MT, Vassilopoulos C, Coolbaugh MI, Siciliano MJ, Monckton DG. Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum Mol Genet*. 2000;9(3):439-45.
- [116] Larson E, Fyfe I, Morton AJ, Monckton DG. Age-, tissue- and length-dependent bidirectional somatic CAG•CTG repeat instability in an allelic series of R6/2 Huntington disease mice. *Neurobiol Dis*. 2015;76:98-111.
- [117] Ament SA, Pearl JR, Grindeland A, Claire JS, Earls JC, Kovalenko M, et al. High resolution time-course mapping of early transcriptomic, molecular and cellular phenotypes in Huntington's disease CAG knock-in mice across multiple genetic backgrounds. *Hum Mol Genet*. 2017;26(5):913-22.
- [118] Van Den Broek WJAA, Nelen MR, Wansink DG, Coerwinkel MM, Te Riele H, Groenen PJTA, et al. Somatic expansion behaviour of the (CTG)_n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3

- and Msh6 mismatch-repair proteins. *Hum Mol Genet.* 2002;11(2):191-8.
- [119] Monckton DG, Coolbaugh MI, Ashizawa KT, Siciliano MJ, Caskey CT. Hypermutable myotonic dystrophy CTG repeats in transgenic mice. *Nat Genet.* 1997;15(2):193-6.
- [120] Gomes-Pereira M, Foiry L, Nicole A, Huguet A, Junien C, Munnich A, et al. CTG trinucleotide repeat "big jumps": Large expansions, small mice. *PLoS Genet.* 2007;3(4):0488-91.
- [121] Neto JL, Lee JM, Afridi A, Gillis T, Guide JR, Dempsey S, et al. Genetic contributors to intergenerational CAG repeat instability in Huntington's disease knock-in mice. *Genetics.* 2017;205(2):503-16.
- [122] Lloret A, Dragileva E, Teed A, Espinola J, Fossale E, Gillis T, et al. Genetic background modifies nuclear mutant huntingtin accumulation and HD CAG repeat instability in Huntington's disease knock-in mice. *Hum Mol Genet.* 2006;15(12):2015-24.
- [123] Gourdon G, Dessen P, Lia AS, Junien C, Hofmann-Radvanyi H. Intriguing association between disease associated unstable trinucleotide repeat and CPG island. *Ann Genet.* 1997;40(2):73-7.
- [124] Libby RT, Hagerman KA, Pineda VV, Lau R, Cho DH, Baccam SL, et al. CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: A novel basis for mutational hot spot determination. *PLoS Genet.* 2008;4(11):e1000257.
- [125] Lia AS, Seznec H, Hofmann-Radvanyi H, Radvanyi F, Duros C, Saquet C, 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(8):1285-91.
- [126] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-13.
- [127] Miller CJ, Kim G-Y, Zhao X, Usdin K. All three mammalian MutL complexes are required for repeat expansion in a mouse cell model of the Fragile X-related disorders. *PLoS Genet.* 2020;16(6):e1008902.
- [128] Claassen DA, Lahue RS. Expansions of CAG•CTG repeats in immortalized human astrocytes. *Hum Mol Genet.* 2007;16(24):3088-96.
- [129] 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(1):37-46.
- [130] Santillan BA, Moye C, Mittelman D, Wilson JH. GFP-based fluorescence assay for CAG repeat instability in cultured human cells. *PLoS One.* 2014;9(11):e113952.
- [131] 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(10):e46807.
- [132] Gorbunova V, Seluanov A, Dion V, Sandor Z, Meservy JL, Wilson JH. Selectable system for monitoring the instability of CTG/CAG triplet repeats in mammalian cells. *Mol Cell Biol.* 2003;23(13):4485-93.
- [133] Cinesi C, Yang B, Dion V. GFP reporters to monitor instability and expression of expanded CAG/CTG repeats. *Methods Mol Biol.* 2020;2056:255-68.
- [134] Cinesi C, Aeschbach L, Yang B, Dion V. Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nat Commun.* 2016;7:13272.
- [135] Chatterjee N, Lin Y, Santillan BA, Yotnda P, Wilson JH. Environmental stress induces trinucleotide repeat mutagenesis in human cells. *Proc Natl Acad Sci U S A.* 2015;112(12):3764-9.
- [136] Gomes-Pereira M, Monckton DG. Chemically induced increases and decreases in the rate of expansion of a CAG-CTG triplet repeat. *Nucleic Acids Res.* 2004;32(9):2865-72.
- [137] Yang Z, Lau R, Marcadier JL, Chitayat D, Pearson CE. Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells. *Am J Hum Genet.* 2003;73(5):1092-105.
- [138] Cannella M, Maglione V, Martino T, Ragona G, Frati L, Li G-M, et al. DNA instability in replicating Huntington's disease lymphoblasts. *BMC Med Genet.* 2009;10:11.
- [139] Hashem VI, Pytlos MJ, Klysik EA, Tsuji K, Khajavi M, Ashizawa T, et al. Chemotherapeutic deletion of CTG repeats in lymphoblast cells from DM1 patients. *Nucleic Acids Res.* 2004;32(21):6334-46.
- [140] Seriola A, Spits C, Simard JP, Hilven P, Haentjens P, Pearson CE, 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(1):176-85.
- [141] Jacquet L, Neueder A, Földes G, Karagiannis P, Hobbs C, Jolinon N, et al. Three Huntington's disease specific mutation-carrying human embryonic stem cell lines have stable number of CAG repeats upon *in vitro* differentiation into cardiomyocytes. *PLoS One.* 2015;10(5):e0126860.
- [142] Khajavi M, Tari AM, Patel NB, Tsuji K, Siwak DR, Meistrich ML, et al. "Mitotic drive" of expanded CTG repeats in myotonic dystrophy type 1 (DM1). *Hum Mol Genet.* 2001;10(8):855-63.
- [143] Conforti P, Besusso D, Bocchi VD, Faedo A, Cesana E, Rossetti G, et al. Faulty neuronal determination and cell polarization are reverted by modulating HD early phenotypes. *Proc Natl Acad Sci U S A.* 2018;115(4):E762-71.
- [144] Aeschbach L, Dion V. Minimizing carry-over PCR contamination in expanded CAG/CTG repeat instability applications. *Sci Rep.* 2017;7(1):18026.
- [145] Dandelot E, Gourdon G. The flash-small-pool PCR: How to transform blotting and numerous hybridization steps into a simple denatured PCR. *Biotechniques.* 2018;64(6):262-5.
- [146] Ciosi M, Cumming S, Chatzi A, Larson E, Tottey W, Lomeikaite V, et al. Approaches to sequence the HTT CAG repeat expansion and quantify repeat length variation. *J Huntingtons Dis.* 2020;doi: 10.3233/JHD-200433
- [147] Nichol K, Pearson CE. CpG methylation modifies the genetic stability of cloned repeat sequences. *Genome Res.* 2002;12(8):1246-56.
- [148] Farrell BT, Lahue RS. CAG-CTG repeat instability in cultured human astrocytes. *Nucleic Acids Res.* 2006;34(16):4495-505.
- [149] Pelletier R, Farrell BT, Miret JJ, Lahue RS. Mechanistic features of CAG•CTG repeat contractions in cultured cells revealed by a novel genetic assay. *Nucleic Acids Res.* 2005;33(17):5667-76.
- [150] Lin Y, Dion V, Wilson JH. A novel selectable system for detecting expansion of CAG-CTG repeats in mammalian cells. *Mutat Res.* 2005;572(1-2):123-31.

- [151] Li GM. Mechanisms and functions of DNA mismatch repair. *Cell Res.* 2008;18(1):85-98.
- [152] Guarné A. The functions of MutL in mismatch repair: The power of multitasking. *Prog Mol Biol Transl Sci.* 2012;110:41-70.
- [153] Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer.* 2015;15(3):181-94.
- [154] Slean MM, Panigrahi GB, Castel AL, Pearson AB, Tomkinson AE, Pearson CE. Absence of MutSB leads to the formation of slipped-DNA for CTG/CAG contractions at primate replication forks. *DNA Repair (Amst).* 2016;42:107-18.
- [155] Mason AG, Tomé S, Simard JP, Libby RT, Bammler TK, Beyer RP, 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(6):1606-18.
- [156] Hick A, Wattenhofer-Donzè M, Chintawar S, Tropel P, Simard JP, Vaucamps N, et al. Neurons and cardiomyocytes derived from induced pluripotent stem cells as a model for mitochondrial defects in Friedreich's ataxia. *Dis Model Mech.* 2013;6(3):608-21.
- [157] Tomé S, Simard JP, Slean MM, Holt I, Morris GE, Wojciechowicz K, et al. Tissue-specific mismatch repair protein expression: MSH3 is higher than MSH6 in multiple mouse tissues. *DNA Repair (Amst).* 2013;12(1):46-52.
- [158] Pluciennik A, Burdett V, Baitinger C, Iyer RR, Shi K, Modrich P. Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLa endonuclease activation. *Proc Natl Acad Sci U S A.* 2013;110(30):12277-82.
- [159] Kadyrova LY, Gujar V, Burdett V, Modrich PL, Kadyrov FA. Human MutL γ , the MLH1-MLH3 heterodimer, is an endonuclease that promotes DNA expansion. *Proc Natl Acad Sci U S A.* 2020;117(7):3535-42.
- [160] Halabi A, Fuselier KTB, Grabczyk E. GAA•TTC repeat expansion in human cells is mediated by mismatch repair complex MutL and depends upon the endonuclease domain in MLH3 isoform one. *Nucleic Acids Res.* 2018;46(8):4022-32.
- [161] Zhang Y, Yuan F, Presnell SR, Tian K, Gao Y, Tomkinson AE, et al. Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell.* 2005;122(5):693-705.
- [162] Castel AL, Tomkinson AE, Pearson CE. CTG/CAG repeat instability is modulated by the levels of human DNA ligase I and its interaction with proliferating cell nuclear antigen. A distinction between replication and slipped-DNA repair. *J Biol Chem.* 2009;284(39):26631-45.
- [163] Panigrahi GB, Slean MM, Simard JP, Gileadi O, Pearson CE. Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSB, but clustered slip-outs are poorly repaired. *Proc Natl Acad Sci U S A.* 2010;107(28):12593-8.
- [164] Tian L, Hou C, Tian K, Holcomb NC, Gu L, Li G-M. Mismatch recognition protein MutSbeta does not hijack (CAG) n hairpin repair *in vitro*. *J Biol Chem.* 2009;284(31):20452-6.
- [165] 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).* 2012;11(2):201-9.
- [166] Lang WH, Coats JE, Majka J, Hura GL, Lin Y, Rasnik I, et al. Conformational trapping of mismatch recognition complex MSH2/MSH3 on repair-resistant DNA loops. *Proc Natl Acad Sci U S A.* 2011;108(42):E837-44.
- [167] Panigrahi GB, Slean MM, Simard JP, Pearson CE. Human mismatch repair protein hMutL α is required to repair short slipped-DNAs of trinucleotide repeats. *J Biol Chem.* 2012;287(50):41844-50.
- [168] Axford MM, Wang YH, Nakamori M, Zannis-Hadjopoulos M, Thornton CA, Pearson CE. Detection of slipped-DNAs at the trinucleotide repeats of the myotonic dystrophy type I disease locus in patient tissues. *PLoS Genet.* 2013;9(12):e1003866.
- [169] Viterbo D, Michoud G, Mosbach V, Dujon B, Richard GF. Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. *DNA Repair (Amst).* 2016;42:94-106.
- [170] Smogorzewska A, Desetty R, Saito TT, Schlabach M, Lach FP, Sowa ME, et al. A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol Cell.* 2010;39(1):36-47.
- [171] MacKay C, Déclais AC, Lundin C, Agostinho A, Deans AJ, MacArtney TJ, et al. Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2. *Cell.* 2010;142(1):65-76.
- [172] Kratz K, Schöpf B, Kaden S, Sandoel A, Eberhard R, Lademann C, et al. Deficiency of FANCD2-associated nuclease KIAA1018/fan1 sensitizes cells to interstrand crosslinking agents. *Cell.* 2010;142(1):77-88.
- [173] Wang R, Persky NS, Yoo B, Ouerfelli O, Smogorzewska A, Elledge SJ, et al. Mechanism of DNA interstrand cross-link processing by repair nuclease FAN1. *Science.* 2014;346(6213):1127-30.
- [174] Deshmukh A, Porro A, Mohiuddin M, Lanni S, Panigrahi G, Caron M, et al. FAN1, a DNA repair nuclease, as a modifier of repeat expansion disorders. *J Huntingtons Dis.* 2020;doi: 10.3233/JHD-200448.
- [175] Cannavo E, Gerrits B, Marra G, Schlapbach R, Jiricny J. Characterization of the interactome of the human MutL homologues MLH1, PMS1, and PMS2. *J Biol Chem.* 2007;282(5):2976-86.
- [176] Wallace SS. Base excision repair: A critical player in many games. *DNA Repair (Amst).* 2014;19:14-26.
- [177] Fortini P, Dogliotti E. Base damage and single-strand break repair: Mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair (Amst).* 2007;6(4):398-409.
- [178] Almeida KH, Sobol RW. A unified view of base excision repair: Lesion-dependent protein complexes regulated by post-translational modification. *DNA Repair (Amst).* 2007;6(6):695-711.
- [179] Hegde ML, Izumi T, Mitra S. Oxidized base damage and single-strand break repair in mammalian genomes: Role of disordered regions and posttranslational modifications in early enzymes. *Prog Mol Biol Transl Sci.* 2012;110:123-53.
- [180] Hegde ML, Mantha AK, Hazra TK, Bhakat KK, Mitra S, Szczesny B. Oxidative genome damage and its repair: Implications in aging and neurodegenerative diseases. *Mech Ageing Dev.* 2012;133(4):157-68.
- [181] Spiro C, Pelletier R, Rolfmeier ML, Dixon MJ, Lahue RS, Gupta G, et al. Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol Cell.* 1999;4(6):1079-85.
- [182] Goula A-V, Stys A, Chan JPK, Trotter Y, Festenstein R, Merienne K. Transcription elongation and

- tissue-specific somatic CAG instability. *PLoS Genet.* 2012;8(11):e1003051.
- [183] Goula A-V, Berquist BR, Wilson DM, Wheeler VC, Trotter Y, Merienne K. 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(12):e1000749.
- [184] Liu Y, Prasad R, Beard WA, Hou EW, Horton JK, McMurray CT, et al. Coordination between polymerase β and FEN1 can modulate CAG repeat expansion. *J Biol Chem.* 2009;284(41):28352-66.
- [185] Lai Y, Budworth H, Beaver JM, Chan NLS, Zhang Z, McMurray CT, et al. Crosstalk between MSH2-MSH3 and $\text{pol}\beta$ promotes trinucleotide repeat expansion during base excision repair. *Nat Commun.* 2016;7:12465.
- [186] Guo J, Gu L, Leffak M, Li GM. MutS β promotes trinucleotide repeat expansion by recruiting DNA polymerase β to nascent (CAG) n or (CTG) n hairpins for error-prone DNA synthesis. *Cell Res.* 2016;26(7):775-86.
- [187] Freudenreich CH, Kantrow SM, Zakian VA. Expansion and length-dependent fragility of CTG repeats in yeast. *Science.* 1998;279(5352):853-6.
- [188] Schweitzer JK, Livingston DM. Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. *Hum Mol Genet.* 1998;7(1):69-74.
- [189] Yang J, Freudenreich CH. Haploinsufficiency of yeast FEN1 causes instability of expanded CAG/CTG tracts in a length-dependent manner. *Gene.* 2007;393(1-2):110-5.
- [190] Lokanga RA, Senejani AG, Sweasy JB, Usdin K. Heterozygosity for a hypomorphic $\text{Pol}\beta$ mutation reduces the expansion frequency in a mouse model of the fragile X-related disorders. *PLoS Genet.* 2015;11(4):e1005181.
- [191] Jarem DA, Wilson NR, Delaney S. Structure-dependent DNA damage and repair in a trinucleotide repeat sequence. *Biochemistry.* 2009;48(28):6655-63.
- [192] Cilli P, Ventura I, Minoprio A, Meccia E, Martire A, Wilson SH, et al. Oxidized dNTPs and the OGG1 and MUTYH DNA glycosylases combine to induce CAG/CTG repeat instability. *Nucleic Acids Res.* 2016;44(11):5190-203.
- [193] Beaver JM, Lai Y, Rolle SJ, Weng L, Greenberg MM, Liu Y. An oxidized abasic lesion inhibits base excision repair leading to DNA strand breaks in a trinucleotide repeat tract. *PLoS One.* 2018;13(2):e0192148.
- [194] 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(2):e56960.
- [195] Derevyanko AG, Endutkin AV, Ishchenko AA, Saparbaev MK, Zharkov DO. Initiation of 8-oxoguanine base excision repair within trinucleotide tandem repeats. *Biochemistry (Mosc)* 2012;77(3):270-9.
- [196] 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(8):887-96.
- [197] Browne SE, Bowling AC, MacGarvey U, Baik MJ, Berger SC, Muqit MMK, et al. Oxidative damage and metabolic dysfunction in huntington's disease: Selective vulnerability of the basal ganglia. *Ann Neurol.* 1997;41(5):646-53.
- [198] Maiuri T, Bowie LE, Truant R. DNA Repair signaling of huntingtin: The next link between late-onset neurodegenerative disease and oxidative DNA damage. *DNA Cell Biol.* 2019;38(1):1-6.
- [199] Gao R, Chakraborty A, Geater C, Pradhan S, Gordon KL, Snowden J, et al. Mutant huntingtin impairs PNKP and ATXN3, disrupting DNA repair and transcription. *Elife.* 2019;8:e42988.
- [200] Langfelder P, Cantle JP, Chatzopoulou D, Wang N, Gao F, Al-Ramahi I, et al. Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. *Nat Neurosci.* 2016;19(4):623-33.
- [201] Watase K, Venken KJT, Sun Y, Orr HT, Zoghbi HY. Regional differences of somatic CAG repeat instability do not account for selective neuronal vulnerability in a knock-in mouse model of SCA1. *Hum Mol Genet.* 2003;12(21):2789-95.
- [202] Goosen N. Nucleotide excision repair in eukaryotes. *Encycl Biol Chem Second Ed.* 2013;5(10):341-4.
- [203] Jung J, Bonini N. CREB-binding protein modulates repeat instability in a *Drosophila* model for polyQ disease. *Science.* 2007;315(5820):1857-9.
- [204] Nakamori M, Pearson CE, Thornton CA. Bidirectional transcription stimulates expansion and contraction of expanded (CTG) \cdot (CAG) repeats. *Hum Mol Genet.* 2011;20(3):580-8.
- [205] Fong YW, Cattoglio C, Tjian R. The intertwined roles of transcription and repair proteins. *Mol Cell.* 2013;52(3):291-302.
- [206] Dion V, Wilson JH. Instability and chromatin structure of expanded trinucleotide repeats. *Trends Genet.* 2009;25(7):288-97.
- [207] Reddy K, Tam M, Bowater RP, Barber M, Tomlinson M, Nichol Edamura K, et al. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res.* 2011;39(5):1749-62.
- [208] Reddy K, Schmidt MHM, Geist JM, Thakkar NP, Panigrahi GB, Wang YH, et al. Processing of double-R-loops in (CAG) \cdot (CTG) and C9orf72 (GGGGCC) \cdot (GGCCCC) repeats causes instability. *Nucleic Acids Res.* 2014;42(16):10473-87.
- [209] García-Muse T, Aguilera A. R Loops: From physiological to pathological roles. *Cell.* 2019;179(3):604-18.
- [210] Su XA, Freudenreich CH. Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A.* 2017;114(40):E8392-401.
- [211] Nospikel T. DNA repair in differentiated cells: Some new answers to old questions. *Neuroscience.* 2007;145(4):1213-21.
- [212] Tarsounas M, Sung P. The antitumorigenic roles of BRCA1-BARD1 in DNA repair and replication. *Nat Rev Mol Cell Biol.* 2020;21(5):284-99.
- [213] Bhattacharyya S, Lahue RS. *Saccharomyces cerevisiae* Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. *Mol Cell Biol.* 2004;24(17):7324-30.
- [214] Cleary JD, Pearson CE. The contribution of cis-elements to disease-associated repeat instability: Clinical and experimental evidence. *Cytogenet Genome Res.* 2003;100(1-4):25-55.
- [215] Brock GJR, 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(6):1061-7.
- [216] Nestor CE, Monckton DG. Correlation of inter-locus polyglutamine toxicity with CAG \cdot CTG triplet repeat expandability and flanking genomic DNA GC content. *PLoS One.* 2011;6(12):e28260.

- [217] Warby SC, Visscher H, Collins JA, Doty CN, Carter C, Butland SL, et al. HTT haplotypes contribute to differences in Huntington disease prevalence between Europe and East Asia. *Eur J Hum Genet.* 2011;19(5):561-6.
- [218] Hauer MH, Gasser SM. Chromatin and nucleosome dynamics in DNA damage and repair. *Genes Dev.* 2017;31(22):2204-21.
- [219] Escargueil AE, Soares DG, Salvador M, Larsen AK, Henriques JAP. What histone code for DNA repair? *Mutat Res.* 2008;658(3):259-70.
- [220] Agbleke AA, Amitai A, Buenrostro JD, Chakrabarti A, Chu L, Hansen AS, et al. Advances in chromatin and chromosome research: Perspectives from multiple fields. *Mol Cell.* 2020;79(6):881-901.
- [221] Huang Y, Li G-M. DNA mismatch repair in the context of chromatin. *Cell Biosci.* 2020;10(1):10.
- [222] Cleary JD, Pearson CE. Replication fork dynamics and dynamic mutations: The fork-shift model of repeat instability. *Trends Genet.* 2005;21(5):272-80.
- [223] López Castel A, Nakamori M, Tomé S, Chitayat D, Gourdon G, Thornton CA, et al. Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum Mol Genet.* 2011;20(1):1-15.
- [224] Sun JH, Zhou L, Emerson DJ, Phyo SA, Titus KR, Gong W, et al. Disease-associated short tandem repeats co-localize with chromatin domain boundaries. *Cell.* 2018;175(1):224-238.e15.
- [225] Buendía GAR, Leleu M, Marzetta F, Vanzan L, Tan JY, Ythier V, et al. Three-dimensional chromatin interactions remain stable upon CAG/CTG repeat expansion. *Sci Adv.* 2020;6(27):eaaz4012.
- [226] Gorbunova V, Seluanov A, Mittelman D, Wilson JH. Genome-wide demethylation destabilizes CTG-CAG trinucleotide repeats in mammalian cells. *Hum Mol Genet.* 2004;13(23):2979-89.
- [227] Suelves N, Kirkham-McCarthy L, Lahue RS, Ginés S. A selective inhibitor of histone deacetylase 3 prevents cognitive deficits and suppresses striatal CAG repeat expansions in Huntington's disease mice. *Sci Rep.* 2017;7(1):6082.
- [228] House NCM, Yang JH, Walsh SC, Moy JM, Freudenreich CH. NuA4 initiates dynamic histone H4 acetylation to promote high-fidelity sister chromatid recombination at postreplication gaps. *Mol Cell.* 2014;55(6):818-28.
- [229] Zhang M, Xiang S, Joo HY, Wang L, Williams KA, Liu W, et al. HDAC6 deacetylates and ubiquitinates MSH2 to maintain proper levels of MutS α . *Mol Cell.* 2014;55(1):31-46.
- [230] Zhang M, Hu C, Moses N, Haakenson J, Xiang S, Quan D, et al. HDAC6 regulates DNA damage response via deacetylating MLH1. *J Biol Chem.* 2019;294(15):5813-26.
- [231] Radhakrishnan R, Li Y, Xiang S, Yuan F, Yuan Z, Telles E, et al. Histone deacetylase 10 regulates DNA mismatch repair and may involve the deacetylation of MutS homolog 2. *J Biol Chem.* 2015;290(37):22795-804.
- [232] Williams GM, Paschalis V, Ortega J, Muskett FW, Hodgkinson JT, Li G-M, et al. HDAC3 deacetylates the DNA mismatch repair factor MutS β to stimulate triplet repeat expansions. *Proc Natl Acad Sci U S A.* 2020;117(38):23597-605.
- [233] Koch MR, House NCM, Cosetta CM, Jong RM, Salomon CG, Joyce CE, et al. The chromatin remodeler isw1 prevents CAG repeat expansions during transcription in *Saccharomyces cerevisiae*. *Genetics.* 2018;208(3):963-76.
- [234] Jonson I, Ougland R, Klungland A, Larsen E. Oxidative stress causes DNA triplet expansion in Huntington's disease mouse embryonic stem cells. *Stem Cell Res.* 2013;11(3):1264-71.
- [235] Dion V, Lin Y, Price BA, Fyffe SL, Seluanov A, Gorbunova V, et al. Genome-wide demethylation promotes triplet repeat instability independently of homologous recombination. *DNA Repair (Amst).* 2008;7(2):313-20.
- [236] Nakamori M, Panigrahi GB, Lanni S, Gall-Duncan T, Hayakawa H, Tanaka H, et al. A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions *in vivo*. *Nat Genet.* 2020;52(2):146-59.
- [237] Tabrizi SJ, Leavitt BR, Landwehrmeyer GB, Wild EJ, Saft C, Barker RA, et al. Targeting huntingtin expression in patients with Huntington's disease. *N Engl J Med.* 2019;380(24):2307-16.
- [238] Nakamori M, Gourdon G, Thornton CA. Stabilization of expanded (CTG)_n(CAG) repeats by antisense oligonucleotides. *Mol Ther.* 2011;19(12):2222-7.
- [239] Gao Z, Cooper TA. Antisense oligonucleotides: Rising stars in eliminating RNA toxicity in myotonic dystrophy. *Hum Gene Ther.* 2013;24(5):499-507.
- [240] Mosbach V, Poggi L, Viterbo D, Charpentier M, Richard GF. TALEN-induced double-strand break repair of CTG trinucleotide repeats. *Cell Rep.* 2018;22(8):2146-59.
- [241] Richard GF, Viterbo D, Khanna V, Mosbach V, Castelein L, Dujon B. Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. *PLoS One.* 2014;9(4):e95611.
- [242] 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(4-5):871-82.
- [243] Richard GF, Cyncynatus C, Dujon B. Contractions and expansions of CAG/CTG trinucleotide repeats occur during ectopic gene conversion in yeast, by a MUS81-independent mechanism. *J Mol Biol.* 2003;326(3):769-82.
- [244] 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(10):2381-90.
- [245] Mittelman D, Moye C, Morton J, Sykoudis K, Lin Y, Carroll D, et al. Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. *Proc Natl Acad Sci U S A.* 2009;106(24):9607-12.
- [246] Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A.* 1996;93(3):1156-60.
- [247] Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science.* 2009;326(5959):1509-12.
- [248] Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science.* 2009;326(5959):1501.
- [249] Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics.* 2010;186(2):756-61.
- [250] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012;337(6096):816-21.

- [251] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339(6121):819-23.
- [252] Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*. 2009;155(3):733-40.
- [253] van Agtmaal EL, André LM, Willemse M, Cumming SA, van Kessel IDG, van den Broek WJAA, et al. CRISPR/Cas9-Induced (CTG-CAG)_n repeat instability in the myotonic dystrophy type 1 locus: Implications for therapeutic genome editing. *Mol Ther*. 2017;25(1):24-43.
- [254] Mosbach V, Viterbo D, Descorps-Declère S, Poggi L, Vaysse-Zinkhöfer W, Richard G-F. Resection and repair of a Cas9 double-strand break at CTG trinucleotide repeats induces local and extensive chromosomal deletions. *PLoS Genet*. 2020;16(7):e1008924.
- [255] Zhao XN, Usdin K. FAN1 protects against repeat expansions in a Fragile X mouse model. *DNA Repair (Amst)*. 2018;69:1-5.
- [256] Higham CF, Monckton DG. Modelling and inference reveal nonlinear length-dependent suppression of somatic instability for small disease associated alleles in myotonic dystrophy type 1 and Huntington disease. *J R Soc Interface*. 2013;10(88):20130605.
- [257] Higham CF, Morales F, Cobbold CA, Haydon DT, Monckton DG. High levels of somatic DNA diversity at the myotonic dystrophy type 1 locus are driven by ultra-frequent expansion and contraction mutations. *Hum Mol Genet*. 2012;21(11):2450-63.