


## Review Article

# Genetic modifiers of repeat expansion disorders

Sangeerthana Rajagopal<sup>1,2</sup>, Jasmine Donaldson<sup>1,2</sup>, Michael Flower<sup>1,2</sup>, Davina J. Hensman Moss<sup>1,2,3</sup>  
and  Sarah J. Tabrizi<sup>1,2</sup>

<sup>1</sup>UCL Huntington's Disease Centre, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, Queen Square, London WC1N 3BG, U.K.; <sup>2</sup>UK Dementia Research Institute, University College London, London WCC1N 3BG, U.K.; <sup>3</sup>St George's University of London, London SW17 0RE, U.K.

**Correspondence:** Sarah J. Tabrizi (s.tabrizi@ucl.ac.uk)



Repeat expansion disorders (REDs) are monogenic diseases caused by a sequence of repetitive DNA expanding above a pathogenic threshold. A common feature of the REDs is a strong genotype–phenotype correlation in which a major determinant of age at onset (AAO) and disease progression is the length of the inherited repeat tract. Over a disease-gene carrier's life, the length of the repeat can expand in somatic cells, through the process of somatic expansion which is hypothesised to drive disease progression. Despite being monogenic, individual REDs are phenotypically variable, and exploring what genetic modifying factors drive this phenotypic variability has illuminated key pathogenic mechanisms that are common to this group of diseases. Disease phenotypes are affected by the cognate gene in which the expansion is found, the location of the repeat sequence in coding or non-coding regions and by the presence of repeat sequence interruptions. Human genetic data, mouse models and *in vitro* models have implicated the disease-modifying effect of DNA repair pathways via the mechanisms of somatic mutation of the repeat tract. As such, developing an understanding of these pathways in the context of expanded repeats could lead to future disease-modifying therapies for REDs.

## Introduction

Repeat expansion disorders (REDs) are monogenic diseases caused by an expanded sequence of repetitive DNA. Short tandem repeats (STRs) are a normal constituent of the human genome. They are repeated units of 1–6 base pairs, also termed microsatellites, and make up 3% of our genetic material [1]. STRs are important for centromere and telomere function and are found in intergenic regions, exons, introns and promoters [2]. Only certain STRs are linked with disease [3]. The length of STRs are variable in human populations and STRs expanded over a pathogenic threshold lead to over 50 REDs causing predominantly neurological and developmental disorders. REDs include the polyglutamine diseases which are caused by expanded cytosine–adenine–guanine (CAG) repeats encoding glutamine e.g. Huntington's disease (HD), spinocerebellar ataxia (SCA) 1, 2, 3, 6, 7, and 17, spinal-bulbar muscular atrophy (SBMA), those with other exonic repeats, and many with non-coding expansions e.g. Fragile X disease (FXD) and Myotonic Dystrophy (DM1 & 2) (Table 1) [5]. REDs can arise from families without a history of disease if a germline repeat expansion occurs, leading the repeat to enter the pathogenic range in offspring [4]. Despite being monogenic, individual REDs are phenotypically variable, and it is through exploring what genetic modifying factors drive this phenotypic variability that has illuminated key pathogenic mechanisms that are common to this group of diseases. REDs often have a long pre-symptomatic phase and human studies have shown that the pathogenic process may be influenced during this time by genetic modifiers and environmental factors [5].

There are various mechanisms through which the expanded repeat drives pathology, for example small changes in length can alter gene expression through changing methylation and splicing patterns as well as binding of transcription factors [6, 7]. Translated proteins from expanded STRs often have the propensity to aggregate leading to cellular dysfunction. A common feature of the REDs is a strong genotype–phenotype correlation: longer repeats are correlated with earlier age at onset (AAO) and rate of progression. During a pathogenic gene carrier's lifetime, the inherited repeat tract lengthens in

Received: 19 May 2023  
Revised: 20 September 2023  
Accepted: 9 October 2023

Version of Record published:  
20 October 2023

**Table 1** Examples of repeat expansion disorders in coding and non-coding regions

Disease	Gene	Repeat motif	Non-disease repeat	Pathogenic repeat	Location
Huntington disease	<i>HTT</i>	CAG	6–35	36–180	Exon 1
Huntington disease like 2	<i>JPH3</i>	CAG	<50	>50	Exon 2A
Spinocerebellar ataxia 1	<i>ATXN1</i>	CAG	6–35	>39	Exon 8
Spinocerebellar ataxia 2	<i>ATXN2</i>	CAG	14–31	37–270	Exon 1
Spinocerebellar ataxia 3	<i>ATXN3</i>	CAG	12–44	~60–87	Exon 10
Spinocerebellar ataxia 6	<i>CACNA1A</i>	CAG	≤18	19–33	Exon 47
Spinocerebellar ataxia 7	<i>ATXN7</i>	CAG	7–27	37–460	Exon 1
Spinocerebellar ataxia 17	<i>TBP</i>	CAG	25–40	49–66	Exon 3
Oculopharyngeal muscular dystrophy	<i>PABPN1</i>	GCG	≤10 [16]	11–18	Exon 1
Dentatorubral-pallidolusian atrophy	<i>ATN1</i>	CAG	6–35	48–93	Exon 5
Spinobulbar muscular atrophy	<i>AR</i>	CAG	<34	38–68	Exon 1
Spinocerebellar ataxia 10	<i>ATXN10</i>	ATTCT	10–32	280–4500	Intron 9
Spinocerebellar ataxia 31	<i>BEAN1</i>	TGGAA	8–20 [17]	500–760	Intron/ intergenic region
Spinocerebellar ataxia 36	<i>NOP56</i>	GGCCTG	5–14 [18]	650–2500	Intron 1
Myotonic dystrophy type 2	<i>CNBP</i>	CCTG	<30 [9]	50–11 000	Intron 1
Familial adult myoclonic epilepsy 1	<i>SAMD12</i>	TTTCA	<100 [19]	105–3680	Intron 4
Familial adult myoclonic epilepsy 2	<i>STARD7</i>	ATTTC	Undetermined [19]	150–460	Intron 1
Familial adult myoclonic epilepsy 3	<i>MARCHF6</i>	TTTCA	9–20 [19]	700–1035	Intron 1
Friedreich's ataxia	<i>FXN</i>	GAA	<40	Pathological threshold >70 but 600–900 common	Intron 1
Chromosome 9 open reading frame 72 (C9orf72)	<i>C9orf72</i>	GGGCC	<20 [20]	Pathological threshold >30 but >300 common	Intron 1
Cerebellar ataxia, neuropathy and vestibular areflexia syndrome (CANVAS)	<i>RFC1</i>	AAGGG	11 repeats (allele frequency (AF) = 0.75), 12–200 (AF = 0.13), 40–1000 (AF = 0.08) [21]	400–2000 [22]	Intron 2
X-linked dystonia Parkinsonsim	<i>TAF1</i>	AGAGGG (retrotransposon insertion)	Repeat not usually present	>35	Intron 32 [23]
Myotonic dystrophy1(DM 1)	<i>DMPK</i>	CTG	5–37	50–10 000	3'UTR
Spinocerebellar ataxia 8	<i>ATXN8</i>	CAG	15–50 [24]	74–1300	3'UTR
Fragile X syndrome	<i>FMR1</i>	CCG	<45	>200	5'UTR
Neuronal intranuclear inclusion disease	<i>NOTCH2NLC</i>	CGG	4–41 [25]	66–517	5'UTR

Pathogenic expansions in coding regions are in general shorter than those in non-coding regions. This list is not exhaustive but is rather to illustrate that larger pathogenic repeat ranges are seen in conditions where the mutation is non-coding compared with those located in coding regions. Adapted from Chintalaphani et al. [26] and Bunting et al. [27].

somatic cells in an age-dependent and tissue-specific manner. For example in HD, repeat expansion in post-mitotic tissue, such as striatum, is associated with earlier AAO and increases symptom severity [8–11]. Somatic expansion appears to be a pathogenic rate driver, driving increased toxicity of the resulting cognate protein or RNA [6]. Expansion in germ cells sometimes leads to anticipation, the phenomenon of increasing disease severity and earlier AAO in successive generations. However, there is great variation in how much the AAO can be attributed to repeat length [12] (See Table 2), and a study found that disease duration is independent of repeat length [13]. Studies have identified residual heritability in disease onset and progression after accounting

**Table 2 Genes implicated as causing an altered disease course in repeat expansion disorders**

Part 1 of 2

Disorder	Genetic modifier	Effect	Method used to identify genetic modifier
HD	FAN 1	Variants that delay and hasten AAO	GWAS [15] and candidate gene study [12]
		Rare variant hastens AAO	Candidate exome sequencing analysis [39]
	MSH3	Increased expression associated with delayed AAO and stabilises repeats in <i>HTT</i> cell line	Transcriptome-wide association study and <i>in vitro</i> model [60]
		Variant slows disease progression	GWAS [55]
		Increased expression associated with faster progression	Transcriptome-wide association study [58]
	MLH1	Increases somatic CAG expansion in striatum	HD homologue mouse model [61]
		Variant affecting HD progression	GWAS and candidate genotyping [62]
	PMS1, PMS2, LIG1	Variants altered AAO	GWAS [15]
DM1	MLH3	Variants that delay and hasten disease	Candidate gene study [38]
	MSH6	Reduces intergenerational contractions	HD homologue mouse model [61]
	MSH2	Deficiency tended towards contractions	Mouse HD model [63]
	MSH3	Variant reduced somatic expansion	Transcriptome-wide association study [58]
		Variant associated with increased somatic instability	Genotyping candidate genes in DM1 cohort [49]
		Deficiency of MSH3 reduced somatic instability	Humanised DM1 mouse model [64]
		Deficiency of MSH3 reduced intergenerational expansion	Humanised DM1 mouse model [65]
MSH6	Deficiency increased CTG repeat instability in mice	Humanised DM1 mouse model [64]	
MSH2	Knock out shifted instability towards contractions	MSH2 deficient mouse [66]	
SCA 1	TBP, ATN1 & <i>HTT</i>	MSH2 knock out prevent germline expansions	MSH2 deficient mouse [67]
		Variants that delay and hasten AAO	Candidate gene study [12]
		Variants that delay and hasten AAO	Candidate gene study [12]
		Associated with altered AAO	Genotyping of CAG genes in SCA1 cohort [68]
SCA2	FAN1	Variants that delay and hasten AAO	Candidate gene study [12]
	PMS2	Variants that delay and hasten AAO	Candidate gene study [12]
SCA3	MLH1	Associated with altered AAO	GWAS [69]
		Variant associated with later AAO but did not reach significance threshold	GWAS [69]
	RAG	Variant associated with hastening AAO	GWAS [69]
	TRIM29	Variant associated with hastening AAO	GWAS [69]
	FAN1	Variants that delay and hasten AAO	Candidate gene study [12]
		Variant hastened AAO	Candidate gene study [57]
		Variants that delay and hasten AAO	Candidate gene study [12]
PMS2	Variants that delay and hasten AAO	Candidate gene study [12]	
SCA6	FAN1	Variants that delay and hasten AAO	Candidate gene study [12]
	PMS2	Variants that delay and hasten AAO	Candidate gene study [12]
	RRM2B	Variants that delay and hasten AAO	Candidate gene study [12]
SCA7	FAN1	Variants that delay and hasten AAO	Candidate gene study [12]
	PMS2	Variants that delay and hasten AAO	Candidate gene study [12]
SCA17	FAN1	Variants that delay and hasten AAO	Candidate gene study [12]
	PMS2	Variants that delay and hasten AAO	Candidate gene study [12]
Fragile X	FAN1	Knock out shows increased somatic expansion	FXS mouse model with FAN1 knock out [70]
		SNP associated with increased somatic expansion	Candidate SNP analysis [59]
	MLH3	Point mutation in endonuclease domain reduces somatic expansion	FXS embryonic mouse stem cell model [71]
		Required for germline and somatic expansion	FXS mouse model [58, 72]

Continued

Downloaded from <http://portlandpress.com/emergtoplifesci/article-pdf/doi/10.1042/ETLS20230015/950992/etls-2023-0015c.pdf> by University College London user on 06 November 2023

**Table 2 Genes implicated as causing an altered disease course in repeat expansion disorders**

Part 2 of 2

Disorder	Genetic modifier	Effect	Method used to identify genetic modifier
	EXO1	Mutation results in increased germline and somatic expansions	FXS mouse model [73]
	MSH3	SNP associated with altered levels of somatic expansion	Candidate SNP analysis [59]
	MSH2	Mutation associated with increased somatic instability. Loss of one MSH2 allele reduced intergenerational expansion	FXS mouse with MSH2 mutation [74]
	MutL $\alpha$ , MutL $\gamma$ & MutL $\beta$	All required for somatic expansion	FXS mouse embryonic stem cell model [75]
	Pol $\beta$	Heterozygous mutation in Pol $\beta$ reduced expansion in paternal gametes and somatic expansion	FXS mouse model [76]
	CSB	Mutation in CSB suggests that transcription coupled repair may have a role in tissue instability	FXS mouse model [77]
FRDA	MSH2 & MSH6	Absent MSH2 & MSH6 in mice reduced somatic instability	FRDA transgenic mouse model [78]
	PMS2	Reduces large expansions	FRDA transgenic mouse model [78]
	MLH1	Loss of MLH1 activity reduced somatic and intergenerational expansion	MLH1 deficient FRDA mouse model [79]
X-linked dystonia Parkinsonism	MSH3, PMS2, DHFR, ANKRD6, EIF2AK1	Variants associated with altered AAO	GWAS [80]

This table shows genetic modifiers identified in many REDs that effect somatic instability. At many loci there are signals with different directions of effect.

for repeat length. There are genetic modifiers both at the site of the repeat (cis modifiers) and across the genome (trans modifiers). In multiple HD, FXD and DM1 model systems DNA damage repair (DDR) proteins have been shown to drive somatic expansion and recent genome-wide association (GWAS) studies have converged on DDR loci as modifiers of clinical outcomes in many REDs [14, 15]. This review discusses the identification of genetic modifiers of REDs, particularly those affecting somatic expansion, the mechanisms through which they operate and how their study may lead to novel therapies for these devastating and currently untreatable conditions. There are other modifiers that are not yet understood mechanistically or there is insufficient evidence of their significance however this review will not discuss these.

## Tandem repeat length

HD shows autosomal dominant inheritance and is caused by a CAG expansion in exon 1 of the *huntingtin* (*HTT*) gene. Like other REDs, inherited CAG repeat length is the main determinant of AAO. Epidemiological studies in HD have established complete penetrance for clinical manifestations of HD with an inherited CAG repeat tract of  $\geq 40$  [28]. Those with  $\geq 56$  repeats usually develop juvenile HD with a more Parkinsonian phenotype, seizures and rapid progression [29]. 36–39 CAGs shows incomplete penetrance where 38 and 39 repeats confer around 60% and 90% risk of penetrance by the age of 82, respectively [28]. 27–35 CAG repeats are termed intermediate alleles where carriers show no clinical symptoms but are at risk of germline expansion in their offspring [29, 30].

Pathogenic expansions are longer in diseases where expansions are non-coding [31, 32] (Table 1). FXD is an X-linked dominant disorder caused by a CGG non-coding repeat in the 5'UTR of the *FMR1* gene which in health has an unmethylated repeat of up to 45 repeats [33]. Expansions over 200 repeats leads to hypermethylation and silencing of the gene. The FMR protein is important for CNS synaptic plasticity and its loss leads to a syndrome of intellectual disability and autism. However, an allele with a premutation of 55–200 repeats leads to a late onset Fragile X-tremor/ataxia syndrome in men while women may have primary ovarian insufficiency. This premutation allele shows enhanced promoter activity creating toxic RNA products [34]. Therefore, repeat size determines downstream cellular cascades with vastly divergent clinical phenotypes.

Determining the pathogenic threshold is more uncertain in some REDs. C9orf72 is caused by an intronic GGGGCC repeat and can lead to a frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS). It

is the most prevalent genetic cause in both disorders. Generally, the disease range for C9orf72 is >30 repeats with normal being <20 [20], though pathogenic repeats are greater than 400 and often in the 1000s in the disease population. Expansions have been found in 0.15% of controls in a UK 1958 birth cohort and >32 repeats have been shown in those with disease and in controls [35] raising the question of penetrance. However, sizing of larger repeat units is difficult using conventional PCR based methods leading to difficulties in establishing length-phenotype relationships.

## Repeat tract structure

In REDs, the repeat length is the primary determinant of AAO but the pathogenic tract purity is also important. Around 3–5% of DM1 families have GGC, CCG and CTC interruptions in the CTG repeat tract which can lead to slower disease progression with a lack of typical clinical features of DM1 [36]. These interruptions are thought to stabilise the expansion and demonstrated to be associated with reduced somatic instability [28, 37].

In HD, the CAG repeat typically ends with a CAACAG cassette (both CAG and CAA encode glutamine). In a cohort of over 700 HD gene carriers, ultra-deep sequencing showed that over 95% of *HTT* alleles had the typical structure [38]. Loss of CAA interruptions are associated with earlier AAO, whilst additional interruptions (CAA/CAC) are associated with later AAO than predicted by their polyglutamine repeat lengths [26, 39, 40]. After correcting for pure CAG length, deletion of the CAACAG cassette was associated with AAO 10 years earlier than the median HD onset for a typical HD allele. When this occurs in those with intermediate alleles (CAG of 36–39), the chance of clinical symptoms is increased. Conversely, most with duplication of the cassette showed slower disease progression scores. Therefore AAO correlates better with length of the uninterrupted CAG sequence than the number of coded glutamines. It was established that these variants in the repeat tract underly the AAO modifying loci on chromosome 4 proximal to the *HTT* locus identified in a genome-wide association study [15]. When uncorrected for pure CAG length, the loss of interruption variant was associated with increased somatic expansion in blood, likely due to altered stability of secondary structures formed by the expanded repeats such as hairpins [40]. Such structures may form when DNA is single stranded e.g. during transcription and occur when complementary bases on the same DNA strand pair within the repeat sequence. These structures are prone to single-strand breaks, predisposing to errors during DNA repair. This may lead to somatic expansion of repeats, discussed below. Emerging data in HD also suggests that the sequence variation around the repeat in the adjacent polyproline tract may also exert phenotypic effects [41, 42].

Repeat tract interruptions are associated with altered AAO in other REDs including FXS, Friedreich ataxia (FRDA) and SCA 1 and 2 [43–45]. 95% of the wild-type *FMR1* alleles have one or two AGG interruptions in the CGG repeat, both non-coding [46]. A study involving 1471 maternal premutation alleles found that 97.5% of those with intergenerational expansion had no AGG interruptions while in those with two or more AGGs, 46.7% showed intergenerational expansion. A similar pattern was found in those with pathogenic expansions [47].

## Somatic instability of repeats

Somatic expansion describes expansion of the repeat tract in somatic tissue [10, 48]. In a study which examined post-mortem brain tissue of 48 HD patients, after accounting for inherited repeat length, greater somatic expansion of CAG repeats was seen in those with earlier onset [9]. This is concordant with HD blood DNA data where higher degrees of somatic instability has been linked directly to earlier AAO and faster disease progression [38]. In DM1, the *DMPK* CTG repeat exhibits somatic expansions in an age-dependent, tissue-specific (e.g. muscle) and expansion-biased manner and is associated with worse clinical outcomes [49, 50]. C9orf72 also displays similar characteristics with somatic expansions in both neuronal and non-neuronal tissues though correlation with clinical phenotypes is not well delineated [9].

In HD, somatic instability occurs in post mitotic neurons particularly in the striatum. Neuronal tissue is susceptible to DNA damage as it is highly metabolically active and high levels of free radicals are generated [42]. HD has a predilection for the striatum which shows particularly high metabolic demand [9, 27]. Interestingly, the tissue-specific patterns of CAG expansion do not always predict pathology. High instability was observed in the cortex and caudate, which demonstrate vulnerability in HD but are relatively unaffected in SCA [27]. Particular cell types may have unique pathogenic thresholds for different toxic species. Recent single-cell RNA sequencing has demonstrated differential vulnerability of CAG expansion of different cortical cell layers in HD [52].



## Human genetic data and repeat expansion disease

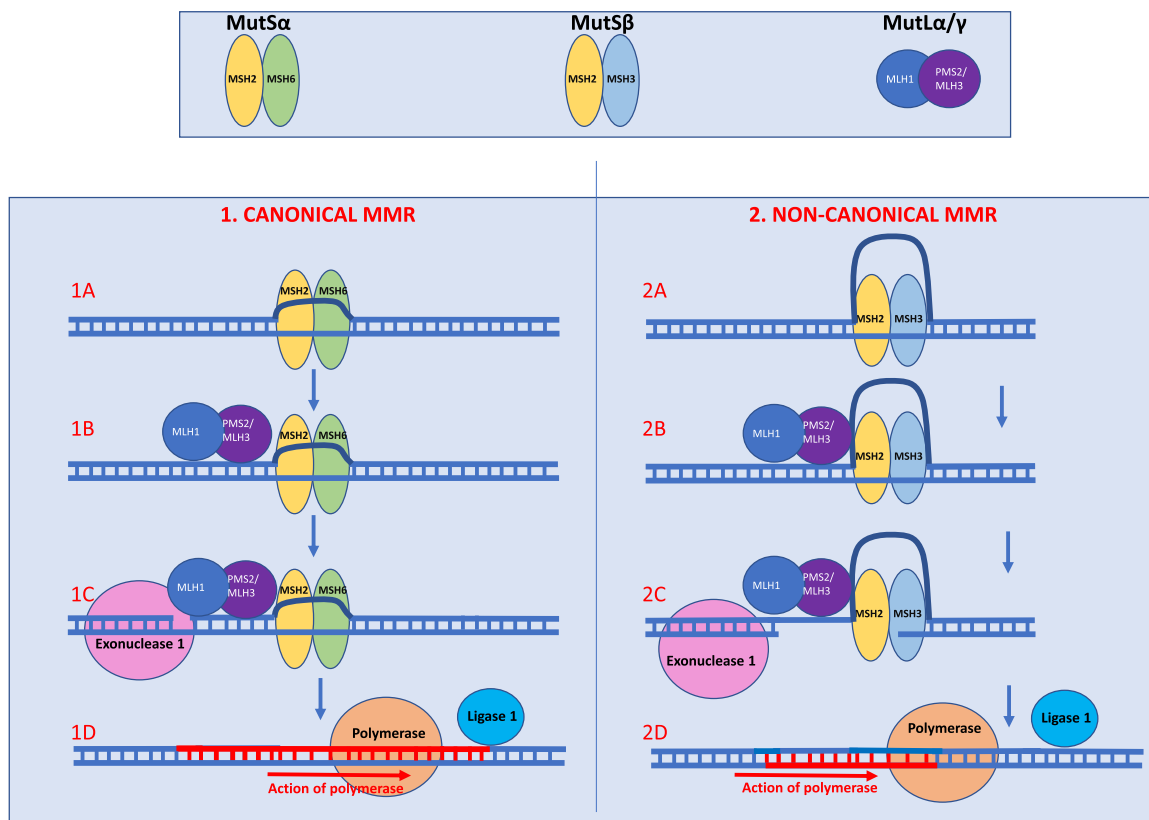
While the existence of somatic instability of the repeat tract of REDs had been well established, more recent human genetic studies and genome wide association studies (GWAS) have had a pivotal role in understanding what genetic modifiers may drive this. These, among other studies, have highlighted the importance of mismatch repair (MMR) and FAN1 in somatic instability (Table 2). The Genetic Modifiers of Huntington's Disease Consortium (GeM) conducted a series of GWAS, first examining the residual age of onset of HD after controlling for CAG repeat length [53]. A significant association at a chromosome 15 locus linked to the DNA repair gene FAN1 displayed two independent effects that accelerate or delay onset by 6.1 years and 1.4 years, respectively. The 2019 GeM consortium strengthened previous findings with significant GWAS signals underpinned by at least six identified candidate modifier loci, many of which contained genes involved in DNA maintenance [15]: PMS1, MLH1, MSH3, PMS2, FAN1, LIG1. The identification of such genes strongly point towards the MMR pathway as an important genetic modifier, likely through involvement in somatic expansion. Other signals were found such as at TCERG1, RRM2B, CCDC82, SYT9, GSG1L that may be related to other mechanisms, or more indirectly involved in DNA maintenance. Subsequent detailed analysis of FAN1 locus modifiers indicates the presence of two onset-hastening missense changes and an onset-delaying variant associated with greater FAN1 expression [54]. Furthermore, an MSH3 SNP in HD with genome-wide significance in meta-analysis was associated with a slower progression score in 216 subjects [55]. An association with MMR in a pathway analysis suggested shared mechanisms influence AAO and disease progression, supported by the finding that about 2/3 of the rate of functional, motor and cognitive progression in HD is determined by the same factors that also determine AAO [56]. While DNA maintenance modifier mechanisms are common to different sub-phenotypes of HD, individual modifier effects act preferentially in the motor or cognitive domains [14].

These HD studies raised the question whether DNA damage response mechanisms were specific to *HTT*, or was a broader mechanism relating to repeat expansions. A candidate gene study found an association between 22 DNA repair loci SNPs and AAO in 1462 subjects with polyglutamine SCAs and HD ( $P = 1.43 \times 10^{-5}$ ) and significant associations at individual SNPs [12]. Additionally, significant association between a FAN1 SNP rs3512 and onset in SCA3 was demonstrated [57]. Supporting this, an allele in *MSH3* exon 1 at the site of the genome-wide signal [55] is associated with a relatively lower rate of somatic CAG-CTG expansion in blood and delayed disease age of onset for HD and DM1 [58]. SNPs in both *MSH3* and *FAN1* are also significantly associated with somatic expansion risk in FXS [59].

## DNA repair mechanisms in REDs

Evidence suggests MMR pathways drive somatic expansion of repeats in HD (CAG), DM1 (CTG), FRDA (GAA) and FXS (CGG) (Table 2). MMR driven mutability is however not always pathogenic as it is harnessed for immunoglobulin hypermutations for the production of high affinity antibodies [81]. Deficiency of core MMR factors, such as MSH2, MSH6, MLH1 and PMS2 have been associated with cancer and can result in genome-wide microsatellite instability (MSI). REDs demonstrate locus-specific expansions at the site of a pathogenic tandem repeats [82, 83]. The mechanisms mediating REDs and cancer are different and highlight the double-edged sword of MMR. The generation of cancer may involve dysfunctional MMR proteins such as loss-of-function mutations in MSH2, MSH6, MLH1 and PMS2 which may be sporadic, or inherited such as in Lynch syndrome [84]. Meanwhile in REDs, functional MMR machinery acts in an error-prone manner on pathogenic repeat sequences, resulting in expansion [89]. The DNA damage response is not the only cellular machinery exhibiting altered function due to the presence of pathogenic repeats as these repeats also trigger non-AUG translation [85].

In dividing cells, MMR operates during DNA replication and is intimately involved with the replication fork as a signal for strand specificity. In the context of DNA damage, the canonical MMR pathway consists of base mismatch recognition, strand specific endonuclease mediated nicking at a site downstream of the mismatch, single-strand degradation and finally single strand synthesis by a polymerase (Figure 1) [86, 87]. Expanded repeat sequences can form secondary structures e.g. hairpins and R-loops and it is thought that these structures act as the substrate for MMR acting erroneously leading to somatic expansion of pathogenic repeats. MutS $\beta$  (MSH2–MSH3) likely recognise these structures and recruit MutL $\alpha$  (MLH1–PMS2) or MutL $\gamma$  (MLH1–MLH3) endonucleases to co-ordinate excision. MutS $\beta$  is thought to compete with FAN1 for loop-out structures resulting in expansion or contractions, respectively [88, 89].



**Figure 1. Normal MMR vs mechanisms that may lead to expansion of repeats in non-dividing cells.**

The figure illustrates the working model based on the current strongest evidence. **The MutS and MutL complexes:** There are two MutS complexes; MutS $\alpha$ , which contains MSH2/MSH6 or MutS $\beta$  comprising MSH2/MSH3. MutS $\alpha$  preferentially recognises and targets mismatches and 1–4 base INDELS while MutS $\beta$  targets medium-sized loop repair [91]. There are three MutL endonuclease complexes; MLH1 complexed with either PMS2, MLH3 or PMS1 creating MutL $\alpha$ , MutL $\gamma$  and MutL $\beta$ , respectively. MLH3 is involved in meiotic repair but can compensate a small amount for MutL $\alpha$  while PMS1's role is unclear and cannot compensate for MutL $\alpha$ . When DNA damage occurs, in canonical MMR, the MutL complexes' endonuclease is thought to create a DNA break in the strand with a mismatch (which is assumed to be the strand carrying an incorrect base). MutL induced breaks are the initiation sites for strand excision performed by exonuclease 1. Canonical and Non-Canonical MMR Hypothesis in non-dividing cells **(1) Canonical MMR: (1A)** The figure illustrates the reaction of MMR to a small loop out. In canonical MMR mismatches, small loops and insertions and deletions (INDELS) are resolved by the recruitment of one of two MutS complexes; MutS $\alpha$  or MutS $\beta$ . MutS $\alpha$  preferentially recognises and targets mismatches and 1–4 base INDELS while MutS $\beta$  targets medium loop repair [33, 91]. This occurs when cells are dividing or the DNA is transcriptionally active causing it to unwind and become single stranded. **(1B)** Once bound, the MutS complex induces recruitment of MutL endonuclease complexes. MutL $\alpha$  is the principle MutL complex for most MMR. The MutL complex creates a DNA break in the strand with an existing break. **(1C)** Excision is then performed by exonucleases e.g. exo 1. **(1D)** There is then faithful repair involving DNA polymerase using the opposite strand as the template strand. **(2) Hypothesised Non-Canonical MMR: (2A)** Strand separation during transcription (or replication) permits pathogenic repeat sequences to form secondary structures e.g. hairpins, R-loops and cruciforms. This figure illustrates the potential downstream effects of large loops. It is thought that these structures act as the substrate for non-canonical MMR. Large loops of 2–10 bases can only be resolved through recognition by MSH3 in MutS $\beta$ . **(2B)** The MutL complex is then recruited and unlike in canonical MMR, MutL complex erroneously creates a break in the strand opposite the loop. **(2C)** One hypothesis is that FAN1 nuclease cuts the strand opposite the loop i.e. the complementary strand (though the location of its action has not been fully elucidated). Therefore the strand with the loop is used as the template strand. **(2D)** There is then erroneous resolution of the loop resulting in elongation of the repeat sequence as Polymerase uses the strand with the loop as the template strand, thus incorporating new repeats into the gene.

Elucidation of MMR in REDs has been aided by genetic knock-outs. Knock-out of both *Msh2* and *Msh3* ablated repeat instability in mouse models of HD [61, 92–94], DM1 [64, 65], FXD [74] and FRDA [78], implicating MutS $\beta$  as an essential driver of somatic expansion. Polymorphisms in *Msh3* found between HD mouse strains appeared to mediate somatic expansion rates via differential *Msh3* expression [92]. In DM1, a human polymorphism proximal to the ATPase domain of *MSH3* was associated with reduced somatic instability in blood [49] and cellular models of HD revealed the ATPase of *MSH3* to be critical in driving repeat expansions [95]. The ATPase catalysis leads to conformational change enabling the sliding of *MSH3*. Therefore, the MutS $\beta$  complex appears to be significant in somatic instability and is known to be important for large loop repair (Figure 1). Interestingly, *MSH3* is known to tolerate loss-of-function variation and therefore represents an attractive therapeutic target [96].

Evidence shows variants in *MLH1* are associated with altered AAO in HD, SCA3 and FRDA. Concordant with hypotheses, *MLH1* promotes the expansion of both *HTT* CAG [60, 97, 98] and *FXN* GAA repeats [79, 99]. MutL cofactors *PMS1* and *PMS2*, but not *MLH3* have been implicated by HD GWAS, though experimental data exploring their role in somatic expansion is inconsistent [14, 15]. Variants in *MLH3* are associated with somatic instability in blood of people with HD [38], and is required for *HTT* CAG repeat expansions in *Hdh*<sup>Q111</sup> mice [97, 100] and *FMR1* CCG expansions in *FXD* mice and embryonic stem cells [70, 71, 75].

Although *PMS1* modifies AAO in HD, the molecular functions of *PMS1* or the heterodimeric complex it forms, MutL $\beta$  (*MLH1*–*PMS1*) are not well understood. Only one study has assessed the effect of *PMS1* on somatic instability, where *PMS1* knockout led to repeat CCG repeat stability in an *FXD* mouse embryonic stem cell (mESC) model [75]. Loss of *PMS2* attenuated repeat instability DM1 mice and *FXD* mESCs, yet potentiated large GAA *FXN* expansions in an *FRDA* mouse model [78].

The most significant modifiers of HD AAO are at *FAN1* loci which, although not a canonical MMR factor, it is a structure-specific 5' exo/endonuclease involved in DNA repair, particularly inter-strand cross-link repair and replication fork recovery [14, 53, 55]. *FAN1* likely modifies AAO in HD and SCAs via an effect on somatic instability [12, 53]. Indeed, *FAN1* protects against repeat expansions in multiple models of HD [39, 60, 98] and a mouse model of *FXD* [70]. Knock-out of *Mlh1* and *Fan1* ablated somatic expansion in *Hdh*<sup>Q111</sup> mice suggesting that *FAN1* works in an MMR-dependent manner to protect against repeat expansions [98].

*FAN1* is hypothesised to stabilise the repeat via two distinct functions. The first function of *FAN1* is its interaction with *MLH1* through competition with MutS $\beta$  for interaction with MutL $\alpha$  [89, 101]. Recent data indicates a protein motif on *FAN1*, dubbed the SPYF motif, is critical for *MLH1* binding and conferring *HTT* CAG repeat stability, likely through its competition with *MSH3* [89]. Secondly, via its nuclease activity, it is thought to promote accurate repair at the repeat [39, 102]. Supporting this, exome sequencing revealed rare non-synonymous coding variants clustering to the nuclease domain associated with an earlier AAO [39]. Candidate gene studies in the SCAs and *FXD* have also implicated a role for *FAN1* but further work in model systems is required [12].

These investigations implicate non-canonical DDR directing somatic instability. Of therapeutic interest are strategies to target *MSH3* since its knockdown appears not to be particularly oncogenic. Interventions could include *MSH3* knockdown using siRNA or antisense oligonucleotides, ATPase inhibition, or potentially involve histone deacetylase 3 (HDAC) inhibition of *MSH3* whose activity potentiates somatic expansions in HD mice [103]. Further work is needed to understand how these processes protect against sequence elongation.

## Conclusions

Repetitive DNA sequences are an important source of genetic diversity in healthy human populations [104]. The REDs are a set of monogenic diseases caused by expanded repeat sequences leading to a group of predominantly developmental and neurological abnormalities. The cognate gene in which the pathogenic expansion exists, and whether repeats are intronic or exonic, exerts an effect on pathogenic repeat length threshold and phenotype. Inherited repeat length is a key factor influencing AAO but does not completely explain variability in disease course. GWA studies probing this residual phenotypic variability have identified other genetic modifiers, including an important role for DNA repair, and specifically the mismatch repair pathway. Biochemical, cell and animal models show variants in DNA repair genes underlie repeat instability and influence disease course in numerous REDs, including HD, *FXD*, SCAs and DM1, suggesting therapeutic intervention could benefit a range of diseases. Recent studies suggest that abnormal DNA structures, such as hairpins, can form at repeat tracts and act as substrates for error-prone non-canonical DDR, resulting in repeat instability. There are not yet any disease-modifying therapies for REDs, but the identification of genetic modifiers demonstrates that



disease course can be altered, and give considerable hope that therapeutics harnessing DNA repair, which are currently in clinical development, can delay onset and slow progression in a whole range of REDs. Future therapy design will need to balance the oncogenic risks with RED disease modification but targeting MSH3 is likely to hold some promise.

## Summary

- The REDs are a group of monogenic diseases that lead to a heterogeneous group of predominantly developmental, and neurological diseases.
- Repeat lengths in REDs are key to determining the age of symptom onset: larger repeats lead to earlier onset and worse disease severity.
- The purity of repeat tracts impacts on age of onset: the presence of interruptions in repeat sequences leads to less severe disease phenotypes.
- DNA damage repair (DDR) pathways modify disease course by acting through somatic instability of repeat sequences during the lifetime of gene carriers.
- There is a huge therapeutic focus on DDR pathways in REDs which will likely lead to new clinical trials and therapies in the REDs.

## Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

## Open Access

Open access for this article was enabled by the participation of University College London in an all-inclusive *Read & Publish* agreement with Portland Press and the Biochemical Society under a transformative agreement with JISC.

## Acknowledgements

Thank you to Darren Monkton for his helpful criticism of this paper.

## Abbreviations

AAO, age at onset; DDR, DNA damage repair; FRDA, Friedreich ataxia; FXD, Fragile X disease; GWAS, genome-wide association; HD, Huntington's disease; HTT, huntingtin; MMR, mismatch repair; REDs, repeat expansion disorders; SCA, spinocerebellar ataxia; STRs, short tandem repeats.

## References

- 1 Gymrek, M. (2017) A genomic view of short tandem repeats. *Curr. Opin. Genet. Dev.* **44**, 9–16 <https://doi.org/10.1016/j.gde.2017.01.012>
- 2 Malik, I., Kelley, C.P., Wang, E.T. and Todd, P.K. (2021) Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.* **22**, 589–607 <https://doi.org/10.1038/s41580-021-00382-6>
- 3 Mousavi, N., Shleizer-Burko, S., Yanicky, R. and Gymrek, M. (2019) Profiling the genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res.* **47**, e90 <https://doi.org/10.1093/nar/gkz501>
- 4 Paulson, H. (2018) Repeat expansion diseases. *Handb. Clin. Neurol.* **147**, 105–123 <https://doi.org/10.1016/B978-0-444-63233-3.00009-9>
- 5 Scahill, R.I., Zeun, P., Osborne-Crowley, K., Johnson, E.B., Gregory, S., Parker, C. et al. (2020) Biological and clinical characteristics of gene carriers far from predicted onset in the Huntington's disease young adult study (HD-YAS): a cross-sectional analysis. *Lancet Neurol.* **19**, 502–512 [https://doi.org/10.1016/S1474-4422\(20\)30143-5](https://doi.org/10.1016/S1474-4422(20)30143-5)
- 6 Hannan, A.J. (2018) Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev. Genet.* **19**, 286–298 <https://doi.org/10.1038/nrg.2017.115>
- 7 Xiao, X., Zhang, C.Y., Zhang, Z., Hu, Z., Li, M. and Li, T. (2022) Revisiting tandem repeats in psychiatric disorders from perspectives of genetics, physiology, and brain evolution. *Mol. Psychiatry* **27**, 466–475 <https://doi.org/10.1038/s41380-021-01329-1>

- 8 Donaldson, J., Powell, S., Rickards, N., Holmans, P. and Jones, L. (2021) What is the pathogenic CAG expansion length in Huntington's disease? *J. Huntingtons Dis.* **10**, 175–202 <https://doi.org/10.3233/JHD-200445>
- 9 Swami, M., Hendricks, A.E., Gillis, T., Massood, T., Mysore, J., Myers, R.H. et al. (2009) Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.* **18**, 3039–3047 <https://doi.org/10.1093/hmg/ddp242>
- 10 Gonitell, R., Moffitt, H., Sathasivam, K., Woodman, B., Detloff, P.J., Faull, R.L.M. et al. (2008) DNA instability in postmitotic neurons. *Proc. Natl Acad. Sci. U.S.A.* **105**, 3467–3472 <https://doi.org/10.1073/pnas.0800048105>
- 11 Kaplan, S., Itzkovitz, S. and Shapiro, E. (2007) A universal mechanism ties genotype to phenotype in trinucleotide diseases. *PLoS Comput. Biol.* **3**, e235 <https://doi.org/10.1371/journal.pcbi.0030235>
- 12 Bettencourt, C., Hensman-Moss, D., Flower, M., Wiethoff, S., Brice, A., Goizet, C. et al. (2016) DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann. Neurol.* **79**, 983–990 <https://doi.org/10.1002/ana.24656>
- 13 Keum, J.W., Shin, A., Gillis, T., Mysore, J.S., Abu Elneel, K., Lucente, D. et al. (2016) The *HTT* CAG-expansion mutation determines age at death but not disease duration in Huntington disease. *Am. J. Hum. Genet.* **98**, 287–298 <https://doi.org/10.1016/j.ajhg.2015.12.018>
- 14 Lee, J.M., Huang, Y., Orth, M., Gillis, T., Siciliano, J., Hong, E. et al. (2022) Genetic modifiers of Huntington disease differentially influence motor and cognitive domains. *Am. J. Hum. Genet.* **109**, 885–899 <https://doi.org/10.1016/j.ajhg.2022.03.004>
- 15 Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Electronic address: gusella@helix.mgh.harvard.edu, Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. (2019) CAG repeat Not polyglutamine length determines timing of Huntington's disease onset. *Cell* **178**, 887–900.e14 <https://doi.org/10.1016/j.cell.2019.06.036>
- 16 Banerjee, A., Apponi, L.H., Pavlath, G.K. and Corbett, A.H. (2013) PABPN1: molecular function and muscle disease. *FEBS J.* **280**, 4230–4250 <https://doi.org/10.1111/febs.12294>
- 17 Ishiguro, T., Nagai, Y. and Ishikawa, K. (2021) Insight into spinocerebellar ataxia type 31 (SCA31) from Drosophila model. *Front. Neurosci.* **15**, 648133 <https://doi.org/10.3389/fnins.2021.648133>
- 18 García-Murias, M., Quintáns, B., Arias, M., Seixas, A.I., Cacheiro, P., Tarrío, R. et al. (2012) “Costa da Morte” ataxia is spinocerebellar ataxia 36: clinical and genetic characterization. *Brain* **135**, 1423–1435 <https://doi.org/10.1093/brain/aws069>
- 19 Loureiro, J.R., Castro, A.F., Figueiredo, A.S. and Silveira, I. (2022) Molecular mechanisms in pentanucleotide repeat diseases. *Cells* **11**, 205 <https://doi.org/10.3390/cells11020205>
- 20 Bram, E., Javanmardi, K., Nicholson, K., Culp, K., Thibert, J.R., Kemppainen, J. et al. (2019) Comprehensive genotyping of the C9orf72 hexanucleotide repeat region in 2095 ALS samples from the NINDS collection using a two-mode, long-read PCR assay. *Amyotroph. Lateral Scler. Frontotemporal. Degener.* **20**, 107–114 <https://doi.org/10.1080/21678421.2018.1522353>
- 21 Cortese, A., Reilly, M.M. and Houlden, H. (1993) RFC1 CANVAS/Spectrum Disorder
- 22 Cortese, A., Simone, R., Sullivan, R., Vandrovicova, J., Tariq, H., Yau, W.Y. et al. (2019) Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat. Genet.* **51**, 649–658 <https://doi.org/10.1038/s41588-019-0372-4>
- 23 Bragg, D.C., Mangkalaphiban, K., Vaine, C.A., Kulkarni, N.J., Shin, D., Yadav, R. et al. (2017) Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. *Proc. Natl Acad. Sci. U.S.A.* **114**, E11020–E11028 <https://doi.org/10.1073/pnas.1712526114>
- 24 Soong, B.W. and Morrison, P.J. (2018) Spinocerebellar ataxias. *Handb. Clin. Neurol.* **155**, 143–174 <https://doi.org/10.1016/B978-0-444-64189-2.00010-X>
- 25 Fan, Y., Xu, Y. and Shi, C. (2022) NOTCH2NLC-related disorders: the widening spectrum and genotype-phenotype correlation. *J. Med. Genet.* **59**, 1–9 <https://doi.org/10.1136/jmedgenet-2021-107883>
- 26 Chintalaphani, S.R., Pineda, S.S., Deveson, I.W. and Kumar, K.R. (2021) An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathol. Commun.* **9**, 98 <https://doi.org/10.1186/s40478-021-01201-x>
- 27 Bunting, E.L., Hamilton, J. and Tabrizi, S.J. (2022) Polyglutamine diseases. *Curr. Opin. Neurobiol.* **72**, 39–47 <https://doi.org/10.1016/j.conb.2021.07.001>
- 28 Tabrizi, S.J., Schobel, S., Gantman, E.C., Mansbach, A., Borowsky, B., Konstantinova, P. et al. (2022) A biological classification of Huntington's disease: the integrated staging system. *Lancet Neurol.* **21**, 632–644 [https://doi.org/10.1016/S1474-4422\(22\)00120-X](https://doi.org/10.1016/S1474-4422(22)00120-X)
- 29 Cubo, E., Ramos-Arroyo, M.A., Martínez-Horta, S., Martínez-Descalls, A., Calvo, S., Gil-Polo, C. et al. (2016) Clinical manifestations of intermediate allele carriers in Huntington disease. *Neurology.* **87**, 571–578 <https://doi.org/10.1212/WNL.0000000000002944>
- 30 Ha, A.D. and Jankovic, J. (2011) Exploring the correlates of intermediate CAG repeats in Huntington disease. *Postgrad Med.* **123**, 116–121 <https://doi.org/10.3810/pgm.2011.09.2466>
- 31 Ishiura, H., Shibata, S., Yoshimura, J., Suzuki, Y., Qu, W., Doi, K. et al. (2019) Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat. Genet.* **51**, 1222–1232 <https://doi.org/10.1038/s41588-019-0458-z>
- 32 Zhou, Z.D., Jankovic, J., Ashizawa, T. and Tan, E.K. (2022) Neurodegenerative diseases associated with non-coding CGG tandem repeat expansions. *Nat. Rev. Neurol.* **18**, 145–157 <https://doi.org/10.1038/s41582-021-00612-7>
- 33 Salcedo-Arellano, M.J., Hagerman, R.J. and Martínez-Cerdeño, V. (2020) Fragile X syndrome: clinical presentation, pathology and treatment. *Gac. Med. Mex.* **156**, 60–66 <https://doi.org/10.24875/GMM.19005275>
- 34 Deng, P.Y. and Klyachko, V.A. (2021) Channelopathies in fragile X syndrome. *Nat. Rev. Neurosci.* **22**, 275–289 <https://doi.org/10.1038/s41583-021-00445-9>
- 35 Beck, J., Poulter, M., Hensman, D., Rohrer, J.D., Mahoney, C.J., Adamson, G. et al. (2013) Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am. J. Hum. Genet.* **92**, 345–353 <https://doi.org/10.1016/j.ajhg.2013.01.011>
- 36 Braida, C., Stefanatos, R.K.A., Adam, B., Mahajan, N., Smeets, H.J.M., Niel, F. et al. (2010) Variant CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients. *Hum. Mol. Genet.* **19**, 1399–1412 <https://doi.org/10.1093/hmg/ddq015>
- 37 Cumming, S.A., Hamilton, M.J., Robb, Y., Gregory, H., McWilliam, C., Cooper, A. et al. (2018) *De novo* repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *Eur. J. Hum. Genet.* **26**, 1635–1647 <https://doi.org/10.1038/s41431-018-0156-9>

- 38 Ciosi, M., Maxwell, A., Cumming, S.A., Hensman Moss, D.J., Alshammari, A.M., Flower, M.D. et al. (2019) A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine* **48**, 568–580 <https://doi.org/10.1016/j.ebiom.2019.09.020>
- 39 McAllister, B., Donaldson, J., Binda, C.S., Powell, S., Chughtai, U., Edwards, G. et al. (2022) Exome sequencing of individuals with Huntington's disease implicates FAN1 nuclease activity in slowing CAG expansion and disease onset. *Nat. Neurosci.* **25**, 446–457 <https://doi.org/10.1038/s41593-022-01033-5>
- 40 Wright, G.E.B., Collins, J.A., Kay, C., McDonald, C., Dolzhenko, E., Xia, Q. et al. (2019) Length of uninterrupted CAG, independent of polyglutamine size, results in increased somatic instability, hastening onset of Huntington disease. *Am. J. Hum. Genet.* **104**, 1116–1126 <https://doi.org/10.1016/j.ajhg.2019.04.007>
- 41 Dawson, J., Baine-Savanhu, F.K., Ciosi, M., Maxwell, A., Monckton, D.G. and Krause, A. (2022) A probable cis-acting genetic modifier of Huntington disease frequent in individuals with African ancestry. *HGG Adv.* **3**, 100130 <https://doi.org/10.1016/j.xhgg.2022.100130>
- 42 Xu, P., Pan, F., Roland, C., Sagui, C. and Wenginger, K. (2020) Dynamics of strand slippage in DNA hairpins formed by CAG repeats: roles of sequence parity and trinucleotide interrupts. *Nucleic Acids Res.* **48**, 2232–2245 <https://doi.org/10.1093/nar/gkaa036>
- 43 Chen, L., Hadd, A., Sah, S., Filipovic-Sadic, S., Krosting, J., Sekinger, E. et al. (2010) An information-rich CGG repeat primed PCR that detects the full range of fragile X expanded alleles and minimizes the need for southern blot analysis. *J. Mol. Diagn.* **12**, 589–600 <https://doi.org/10.2353/jmol.2010.090227>
- 44 Menon, R.P., Nethisinghe, S., Faggiano, S., Vannocci, T., Rezaei, H., Pemble, S. et al. (2013) The role of interruptions in polyQ in the pathology of SCA1. *PLoS Genet.* **9**, e1003648 <https://doi.org/10.1371/journal.pgen.1003648>
- 45 Charles, P., Camuzat, A., Benammar, N., Sellal, F., Destée, A., Bonnet, A.M. et al. (2007) Are interrupted SCA2 CAG repeat expansions responsible for parkinsonism? *Neurology* **69**, 1970–1975 <https://doi.org/10.1212/01.wnl.0000269323.21969.db>
- 46 Villate, O., Ibarluzea, N., Maortua, H., de la Hoz, A.B., Rodriguez-Revenga, L., Izquierdo-Álvarez, S. et al. (2020) Effect of AGG interruptions on FMR1 maternal transmissions. *Front. Mol. Biosci.* **7**, 135 <https://doi.org/10.3389/fmolb.2020.00135>
- 47 Domniz, N., Ries-Levavi, L., Cohen, Y., Marom-Haham, L., Berkenstadt, M., Pras, E. et al. (2018) Absence of AGG interruptions is a risk factor for full mutation expansion among Israeli FMR1 premutation carriers. *Front. Genet.* **9**, 606 <https://doi.org/10.3389/fgene.2018.00606>
- 48 Kennedy, L., Evans, E., Chen, C.M., Craven, L., Detloff, P.J., Ennis, M. et al. (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.* **12**, 3359–3367 <https://doi.org/10.1093/hmg/ddg352>
- 49 Morales, F., Vásquez, M., Santamaría, C., Cuenca, P., Corrales, E. and Monckton, D.G. (2016) 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)* **40**, 57–66 <https://doi.org/10.1016/j.dnarep.2016.01.001>
- 50 Morales, F., Couto, J.M., Higham, C.F., Hogg, G., Cuenca, P., Braidá, C. et al. (2012) Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. *Hum Mol Genet.* **21**, 3558–3567 <https://doi.org/10.1093/hmg/dds185>
- 51 Nordin, A., Akimoto, C., Wuolikainen, A., Alstermark, H., Jonsson, P., Birve, A. et al. (2015) Extensive size variability of the GGGGCC expansion in C9orf72 in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD. *Hum. Mol. Genet.* **24**, 3133–3142 <https://doi.org/10.1093/hmg/ddv064>
- 52 Pressl, C., Mätlik, K., Kus, L., Darnell, P., Luo, J.D., Weiss, A.R. et al. (2023) Layer 5a corticostriatal projection neurons are selectively vulnerable in Huntington's disease. *bioRxiv* <https://doi.org/10.1101/2023.04.24.538096>
- 53 Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. (2015) Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell* **162**, 516–526 <https://doi.org/10.1016/j.cell.2015.07.003>
- 54 Kim, K.H., Hong, E.P., Shin, J.W., Chao, M.J., Loupe, J., Gillis, T. et al. (2020) Genetic and functional analyses point to FAN1 as the source of multiple Huntington disease modifier effects. *Am. J. Hum. Genet.* **107**, 96–110 <https://doi.org/10.1016/j.ajhg.2020.05.012>
- 55 Moss, D.J.H., Pardiñas, A.F., Langbehn, D., Lo, K., Leavitt, B.R., Roos, R. et al. (2017) Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *Lancet Neurol.* **16**, 701–711 [https://doi.org/10.1016/S1474-4422\(17\)30161-8](https://doi.org/10.1016/S1474-4422(17)30161-8)
- 56 Aziz, N.A., van der Burg, J.M.M., Tabrizi, S.J. and Landwehrmeyer, G.B. (2018) Overlap between age-at-onset and disease-progression determinants in Huntington disease. *Neurology* **90**, e2099–e2106 <https://doi.org/10.1212/WNL.0000000000005690>
- 57 Mergener, R., Furtado, G.V., de Mattos, E.P., Leotti, V.B., Jardim, L.B. and Saraiva-Pereira, M.L. (2020) Variation in DNA repair system gene as an additional modifier of age at onset in spinocerebellar ataxia type 3/Machado-Joseph disease. *Neuromolecular Med.* **22**, 133–138 <https://doi.org/10.1007/s12017-019-08572-4>
- 58 Flower, M., Lomeikaite, V., Ciosi, M., Cumming, S., Morales, F., Lo, K. et al. (2019) MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain* **142**, 1876–1886 <https://doi.org/10.1093/brain/awz115>
- 59 Hwang, Y.H., Hayward, B.E., Zafarullah, M., Kumar, J., Durbin Johnson, B., Holmans, P. et al. (2022) Both cis and trans-acting genetic factors drive somatic instability in female carriers of the FMR1 premutation. *Sci. Rep.* **12**, 10419 <https://doi.org/10.1038/s41598-022-14183-0>
- 60 Goold, R., Flower, M., Moss, D.H., Medway, C., Wood-Kaczmar, A., Andre, R. et al. (2019) FAN1 modifies Huntington's disease progression by stabilizing the expanded *HTT* CAG repeat. *Hum. Mol. Genet.* **28**, 650–661 <https://doi.org/10.1093/hmg/ddy375>
- 61 Dragileva, E., Hendricks, A., Teed, A., Gillis, T., Lopez, E.T., Friedberg, E.C. et al. (2009) Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol. Dis.* **33**, 37–47 <https://doi.org/10.1016/j.nbd.2008.09.014>
- 62 Lee, J.M., Chao, M.J., Harold, D., Abu Elneel, K., Gillis, T., Holmans, P. et al. (2017) A modifier of Huntington's disease onset at the MLH1 locus. *Hum. Mol. Genet.* **26**, 3859–3867 <https://doi.org/10.1093/hmg/ddx286>
- 63 Manley, K., Shirley, T.L., Flaherty, L. and Messer, A. (1999) Msh2 deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.* **23**, 471–473 <https://doi.org/10.1038/70598>
- 64 van den Broek, W.J.A.A., Nelen, M.R., Wansink, D.G., Coerwinkel, M.M., te Riele, H., Groenen, P.J.T.A. et al. (2002) Somatic expansion behaviour of the (CTG)<sub>n</sub> repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum. Mol. Genet.* **11**, 191–198 <https://doi.org/10.1093/hmg/11.2.191>
- 65 Foiry, L., Dong, L., Savouret, C., Hubert, L., te Riele, H., Junien, C. et al. (2006) Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Hum. Genet.* **119**, 520–526 <https://doi.org/10.1007/s00439-006-0164-7>

- 66 Savouret, C., Brisson, E., Essers, J., Kanaar, R., Pastink, A., te Riele, H. et al. (2003) CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J.* **22**, 2264–2273 <https://doi.org/10.1093/emboj/cdg202>
- 67 Savouret, C., Garcia-Cordier, C., Megret, J., te Riele, H., Junien, C. and Gourdon, G. (2004) MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Mol. Cell. Biol.* **24**, 629–637 <https://doi.org/10.1128/MCB.24.2.629-637.2004>
- 68 Wang, P., Chen, Z., Peng, Y., Cao, L., Li, X., Wang, C. et al. (2019) (CAG)<sub>n</sub> loci as genetic modifiers of age at onset in patients with spinocerebellar ataxia type 1 from mainland China. *Eur. J. Neurol.* **26**, 1130–1136 <https://doi.org/10.1111/ene.13954>
- 69 Akçimen, F., Martins, S., Liao, C., Bourassa C, V., Catoire, H., Nicholson, G.A. et al. (2020) Genome-wide association study identifies genetic factors that modify age at onset in Machado-Joseph disease. *Aging* **12**, 4742–4756 <https://doi.org/10.18632/aging.102825>
- 70 Zhao, X.N. and Usdin, K. (2018) FAN1 protects against repeat expansions in a Fragile X mouse model. *DNA Repair (Amst)* **69**, 1–5 <https://doi.org/10.1016/j.dnarep.2018.07.001>
- 71 Hayward, B.E., Steinbach, P.J. and Usdin, K. (2020) A point mutation in the nuclease domain of MLH3 eliminates repeat expansions in a mouse stem cell model of the Fragile X-related disorders. *Nucleic Acids Res.* **48**, 7856–7863 <https://doi.org/10.1093/nar/gkaa573>
- 72 Zhao, X.N., Kumari, D., Gupta, S., Wu, D., Evanitsky, M., Yang, W. et al. (2015) Mutsβ generates both expansions and contractions in a mouse model of the Fragile X-associated disorders. *Hum. Mol. Genet.* **24**, 7087–7096 <https://doi.org/10.1093/hmg/ddv408>
- 73 Zhao, X., Zhang, Y., Wilkins, K., Edelmann, W. and Usdin, K. (2018) MutLγ promotes repeat expansion in a Fragile X mouse model while EXO1 is protective. *PLoS Genet.* **14**, e1007719 <https://doi.org/10.1371/journal.pgen.1007719>
- 74 Lokanga, R.A., Zhao, X.N. and Usdin, K. (2014) The mismatch repair protein MSH2 is rate limiting for repeat expansion in a fragile X premutation mouse model. *Hum. Mutat.* **35**, 129–136 <https://doi.org/10.1002/humu.22464>
- 75 Miller, C.J., Kim, G.Y., Zhao, X. and Usdin, K. (2020) All three mammalian MutL complexes are required for repeat expansion in a mouse cell model of the Fragile X-related disorders. *PLoS Genet.* **16**, e1008902 <https://doi.org/10.1371/journal.pgen.1008902>
- 76 Lokanga, R.A., Senejani, A.G., Sweasy, J.B. and Usdin, K. (2015) Heterozygosity for a hypomorphic Polβ mutation reduces the expansion frequency in a mouse model of the Fragile X-related disorders. *PLoS Genet.* **11**, e1005181 <https://doi.org/10.1371/journal.pgen.1005181>
- 77 Zhao, X.N. and Usdin, K. (2014) Gender and cell-type-specific effects of the transcription-coupled repair protein, ERCC6/CSB, on repeat expansion in a mouse model of the fragile X-related disorders. *Hum. Mutat.* **35**, 341–349 <https://doi.org/10.1002/humu.22495>
- 78 Bourn, R.L., De Biase, I., Pinto, R.M., Sandi, C., Al-Mahdawi, S., Pook, M.A. et al. (2012) Pms2 suppresses large expansions of the (GAA-TTC)<sub>n</sub> sequence in neuronal tissues. *PLoS ONE* **7**, e47085 <https://doi.org/10.1371/journal.pone.0047085>
- 79 Ezzatizadeh, V., Sandi, C., Sandi, M., Anjomani-Virmouni, S., Al-Mahdawi, S. and Pook, M.A. (2014) MutLα heterodimers modify the molecular phenotype of Friedreich ataxia. *PLoS ONE* **9**, e100523 <https://doi.org/10.1371/journal.pone.0100523>
- 80 Laabs, B.H., Klein, C., Pozojevic, J., Domingo, A., Brüggemann, N., Grütz, K. et al. (2021) Identifying genetic modifiers of age-associated penetrance in X-linked dystonia-parkinsonism. *Nat. Commun.* **12**, 3216 <https://doi.org/10.1038/s41467-021-23491-4>
- 81 Xantotti, K.J. and Gearhart, P.J. (2016) Antibody diversification caused by disrupted mismatch repair and promiscuous DNA polymerases. *DNA Repair (Amst)* **38**, 110–116 <https://doi.org/10.1016/j.dnarep.2015.11.011>
- 82 De' Angelis, G.L., Bottarelli, L., Azzoni, C., De' Angelis, N., Leandro, G., di Mario, F. et al. (2018) Microsatellite instability in colorectal cancer. *Acta Biomed.* **89**, 97–101 <https://doi.org/10.23750/abm.v89i9-S.7960>
- 83 Sud, A., Kinnersley, B. and Houlston, R.S. (2017) Genome-wide association studies of cancer: current insights and future perspectives. *Nat. Rev. Cancer* **17**, 692–704 <https://doi.org/10.1038/nrc.2017.82>
- 84 Cerretelli, G., Ager, A., Arends, M.J. and Frayling, I.M. (2020) Molecular pathology of Lynch syndrome. *J. Pathol.* **250**, 518–531 <https://doi.org/10.1002/path.5422>
- 85 Kearse, M.G., Green, K.M., Krans, A., Rodriguez, C.M., Linsalata, A.E., Goldstrohm, A.C. et al. (2016) CGG repeat-associated non-AUG translation utilizes a cap-dependent scanning mechanism of initiation to produce toxic proteins. *Mol. Cell* **62**, 314–322 <https://doi.org/10.1016/j.molcel.2016.02.034>
- 86 Gupta, S., Gellert, M. and Yang, W. (2011) Mechanism of mismatch recognition revealed by human MutSβ bound to unpaired DNA loops. *Nat. Struct. Mol. Biol.* **19**, 72–78 <https://doi.org/10.1038/nsmb.2175>
- 87 Genschel, J. and Modrich, P. (2003) Mechanism of 5'-directed excision in human mismatch repair. *Mol. Cell* **12**, 1077–1086 [https://doi.org/10.1016/s1097-2765\(03\)00428-3](https://doi.org/10.1016/s1097-2765(03)00428-3)
- 88 Phadte, A.S., Bhatia, M., Ebert, H., Abdullah, H., Elrazaq, E.A., Komolov, K.E. et al. (2023) FAN1 removes triplet repeat extrusions via a PCNA- and RFC-dependent mechanism. *Proc. Natl Acad. Sci. U.S.A.* **120**, e2302103120 <https://doi.org/10.1073/pnas.2302103120>
- 89 Goold, R., Hamilton, J., Menneteau, T., Flower, M., Bunting, E.L., Aldous, S.G. et al. (2021) FAN1 controls mismatch repair complex assembly via MLH1 retention to stabilize CAG repeat expansion in Huntington's disease. *Cell Rep.* **36**, 109649 <https://doi.org/10.1016/j.celrep.2021.109649>
- 90 Iyer, R.R. and Pluciennik, A. (2021) DNA mismatch repair and its role in Huntington's disease. *J. Huntingtons Dis.* **10**, 75–94 <https://doi.org/10.3233/JHD-200438>
- 91 Liu, D., Keijzers, G. and Rasmussen, L.J. (2017) DNA mismatch repair and its many roles in eukaryotic cells. *Mutat. Res. Rev. Mutat. Res.* **773**, 174–187 <https://doi.org/10.1016/j.mrrev.2017.07.001>
- 92 Tomé, S., Manley, K., Simard, J.P., Clark, G.W., Slean, M.M., Swami, M. et al. (2013) MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. *PLoS Genet.* **9**, e1003280 <https://doi.org/10.1371/journal.pgen.1003280>
- 93 Wheeler, V.C., Lebel, L.A., Vrbanac, V., Teed, A., te Riele, H. and MacDonald, M.E. (2003) Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. *Hum. Mol. Genet.* **12**, 273–281 <https://doi.org/10.1093/hmg/ddg056>
- 94 Kovalenko, M., Dragileva, E., St Claire, J., Gillis, T., Guide, J.R., New, J. et al. (2012) 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* **7**, e44273 <https://doi.org/10.1371/journal.pone.0044273>
- 95 Keogh, N., Chan, K.Y., Li, G.M. and Lahue, R.S. (2017) Mutsβ abundance and Msh3 ATP hydrolysis activity are important drivers of CTG•CAG repeat expansions. *Nucleic Acids Res.* **45**, 10068–10078 <https://doi.org/10.1093/nar/gkx650>
- 96 Valle, L., de Voer, R.M., Goldberg, Y., Sjursen, W., Försti, A., Ruiz-Ponte, C. et al. (2019) Update on genetic predisposition to colorectal cancer and polyposis. *Mol. Aspects Med.* **69**, 10–26 <https://doi.org/10.1016/j.mam.2019.03.001>

- 97 Pinto, R.M., Dragileva, E., Kirby, A., Lloret, A., Lopez, E., St Claire, J. et al. (2013) Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. *PLoS Genet.* **9**, e1003930 <https://doi.org/10.1371/journal.pgen.1003930>
- 98 Loupe, J.M., Pinto, R.M., Kim, K.H., Gillis, T., Mysore, J.S., Andrew, M.A. et al. (2020) 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.* **29**, 3044–3053 <https://doi.org/10.1093/hmg/ddaa196>
- 99 Halabi, A., Fuselier, K.T.B. and Grabczyk, E. (2018) GAA•TTC repeat expansion in human cells is mediated by mismatch repair complex MutL $\gamma$  and depends upon the endonuclease domain in MLH3 isoform one. *Nucleic Acids Res.* **46**, 4022–4032 <https://doi.org/10.1093/nar/gky143>
- 100 Roy, J.C.L., Vitalo, A., Andrew, M.A., Mota-Silva, E., Kovalenko, M., Burch, Z. et al. (2021) Somatic CAG expansion in Huntington's disease is dependent on the MLH3 endonuclease domain, which can be excluded via splice redirection. *Nucleic Acids Res.* **49**, 3907–3918 <https://doi.org/10.1093/nar/gkab152>
- 101 Porro, A., Mohiuddin, M., Zurfluh, C., Spegg, V., Dai, J., Iehl, F. et al. (2021) FAN1-MLH1 interaction affects repair of DNA interstrand cross-links and slipped-CAG/CTG repeats. *Sci. Adv.* **7**, eabf7906 <https://doi.org/10.1126/sciadv.abf7906>
- 102 Zhao, X., Lu, H. and Usdin, K. (2021) FAN1's protection against CGG repeat expansion requires its nuclease activity and is FANCD2-independent. *Nucleic Acids Res.* **49**, 11643–11652 <https://doi.org/10.1093/nar/gkab899>
- 103 Williams, G.M., Paschalis, V., Ortega, J., Muskett, F.W., Hodgkinson, J.T., Li, G.M. et al. (2020) HDAC3 deacetylates the DNA mismatch repair factor MutS $\beta$  to stimulate triplet repeat expansions. *Proc. Natl Acad. Sci. U.S.A.* **117**, 23597–23605 <https://doi.org/10.1073/pnas.2013223117>
- 104 Marshall, J.N., Lopez, A.I., Pfaff, A.L., Koks, S., Quinn, J.P. and Bubbs, V.J. (2021) Variable number tandem repeats: their emerging role in sickness and health. *Exp. Biol. Med. (Maywood)* **246**, 1368–1376 <https://doi.org/10.1177/15353702211003511>