Review

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’s Disease

Xiaonan Zhaoa, Daman Kumaria, Carson J. Millera, Geum-Yi Kimd, Bruce Haywardf, Antonia G. Vitalob,c, Ricardo Mouropintob,c,d,∗ and Karen Usdina,∗
aLaboratory of Cell and Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA
bCenter for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
cDepartment of Neurology, Harvard Medical School, Boston, MA, USA
dProgram in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Abstract. Huntington’s disease (HD) is one of a large group of human disorders that are caused by expanded DNA repeats. These repeat expansion disorders can have repeat units of different size and sequence that can be located in any part of the gene and, while the pathological consequences of the expansion can differ widely, there is evidence to suggest that the underlying mutational mechanism may be similar. In the case of HD, the expanded repeat unit is a CAG trinucleotide located in exon 1 of the huntingtin (HTT) gene, resulting in an expanded polyglutamine tract in the huntingtin protein. Expansion results in neuronal cell death, particularly in the striatum. Emerging evidence suggests that somatic CAG expansion, specifically expansion occurring in the brain during the lifetime of an individual, contributes to an earlier disease onset and increased severity. In this review we will discuss mouse models of two non-CAG repeat expansion diseases, specifically the Fragile X-related disorders (FXDs) and Friedreich ataxia (FRDA). We will compare and contrast these models with mouse and patient-derived cell models of various other repeat expansion disorders and the relevance of these findings for somatic expansion in HD. We will also describe additional genetic factors and pathways that modify somatic expansion in the FXD mouse model for which no comparable data yet exists in HD mice or humans. These additional factors expand the potential druggable space for diseases like HD where somatic expansion is a significant contributor to disease impact.

Keywords: Huntington’s disease, Friedreich ataxia, Fragile X-related disorders, FMR1-associated disorders, trinucleotide repeat instability, mismatch repair, base excision repair, double-strand break repair, non-homologous end-joining

∗Correspondence to: Karen Usdin, Gene Structure and Disease Section, Laboratory of Cell and Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. E-mail: karenu@nih.gov; Ricardo Mouropinto, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. E-mail: rmouropinto@mgh.harvard.edu.
INTRODUCTION

Huntington's disease (HD) is a neurodegenerative disorder characterized most typically by abnormal involuntary movements or chorea, together with progressive and debilitating motor, behavioral and cognitive impairments (reviewed in [1]). HD is inherited in an autosomal dominant manner and is invariably caused by the expansion of a CAG repeat located in exon 1 of the HTT gene [2]. Expansion results in the production of a mutant HTT protein with an expanded polyglutamine tract, aberrant HTT splicing isoforms [3], novel HTT antisense transcripts [4], elevated levels of CAG-containing microRNAs [5], and a number of abnormal proteins generated by repeat-associated non-AUG (RAN) translation [6], with expansion ultimately resulting in neuronal dysfunction and death. Pathology is seen in carriers of alleles with >35 CAG repeats, with alleles having 36–39 CAGs showing reduced penetrance [2]. The length of the expanded repeat is a major determinant of both the likelihood of further expansion and the age at disease onset [7–11]. Progressive increases in repeat number on germline transmission account for the genetic anticipation seen in HD families [11, 12], as well as transitions from high normal alleles (27–35 CAGs) to disease alleles (i.e., de novo mutations) and transitions from alleles associated with incomplete penetrance to those causing completely penetrant disease [13–15].

The CAG repeat is also highly unstable in some somatic tissues, expanding progressively over time in a length-dependent and cell type/tissue-specific manner [16]. Expansions occur in postmitotic neurons [17, 18], with expansions in some brain regions like cortex and striatum being typically more extensive than expansions in blood [19]. A number of lines of evidence support the idea that somatic expansion is an important disease modifier. This includes the fact that larger somatic expansions in HD postmortem brain are associated with an earlier disease onset [20], and the observation that HD patients with higher levels of somatic expansion measured in blood have worse HD outcomes [21]. Furthermore, individuals with pure CAG repeat tracts have an earlier age at onset than individuals with CAA interruptions at the 3′ end of the repeat tract [21–23]. While such interruptions do not change the number of glutamines in the PolyQ tract, they result in a reduction in somatic expansion [23]. This suggests that the rate of further repeat expansion during an individual’s lifetime is an important contributor to HD onset. Finally, as will be discussed later in this review, compelling evidence from recent genome-wide association (GWA) and transcriptome-wide association (TWA) studies strongly implicates known genetic modifiers of somatic expansion as significant modifiers of the onset of HD motor symptoms, accounting for 30–50% of the variation in age at onset [21, 22, 24–27].

A large number of other degenerative diseases are also caused by repeat expansion, with the largest group being those that are also caused by expanded CAG/CTG repeats such as in Myotonic Dystrophy type 1 (DM1) and numerous Spinocerebellar ataxias (SCAs) (see Fig. 1). In addition, many other disorders are caused by the expansion of trinucleotide repeats other than CAG/CTG or repeats with a variety of other unit sizes and sequences (reviewed in [28]). Furthermore, there is some evidence to suggest that tandem repeats may contribute to autism spectrum disorders [29], and in addition to those cases where the repeat is sufficient to cause disease, variations in the number of tandem repeats can also be a modifier of disease severity or age at disease onset, as in X-linked dystonia parkinsonism [30]. The disease-associated repeats share some common features including the ability to form non-canonical nucleic acid secondary structures as illustrated at the bottom of Fig. 1 (reviewed in [31]). Notably, despite the wide range of structures formed, these structures all have regions of single-strandedness that may make them prone to DNA damage. The repetitive nature of the repeat tract, coupled with the ability to form secondary structures, also increases the possibility of out-of-register reannealing during replication, transcription or repair that could potentially generate a substrate upon which the expansion process can act.

The expansion process in the non-CAG/CTG repeat expansion disorders has been most intensively studied in Friedreich ataxia (FRDA) and the Fragile X-related disorders (FXDs; aka the FMR1-associated disorders). FRDA is the most common hereditary ataxia and is typically caused by homozygosity for an expanded GAA-repeat tract in intron 1 of the frataxin (FXN) gene [32]. The FXDs are caused by the expansion of a CGG-repeat tract in the 5′ untranslated region of the X-linked FMR1 gene. This group of disorders includes the neurodegenerative disorder, Fragile X associated tremor/ataxia, a form of female infertility, Fragile X-associated primary ovarian insufficiency, and Fragile X syndrome, the most common inherited cause of intellectual disabilities and monogenic cause of autism worldwide [33]. Expansions in these diseases share a number of
Fig. 1. The repeat expansion diseases.

From mouse models of FRDA and the FXDs. For a broader perspective, the reader is referred to an excellent recent review that discusses findings from other model systems [44].

**FRDA AND FXD MOUSE MODELS**

A number of FRDA mouse models with expanded GAA repeats have been generated including a knock-in (KI) mouse containing a (GAA)$_{230}$ repeat expansion in the first intron of the endogenous $Fxn$ gene [45] and two yeast artificial chromosome (YAC) transgenic mouse lines, YG8R and YG22R, containing different numbers of copies of a randomly integrated human $Fxn$ transgene (370kb of human genomic sequence) with 90–190 GAA repeats [46]. More recently, through natural breeding of the YG8R line, a new line, YG8sR, has been developed harboring a single copy of the $Fxn$ transgene and a single common features with expansions in HD. The FXDs and FRDA, like HD, show expansion both on germline transmission and somatically during the lifetime of the individual [34–42]. Although the magnitude of the changes in repeat number that are generally associated with HD is much smaller than those seen in the FXDs and FRDA, the starting repeat sizes are also much smaller (Fig. 1). Nonetheless, large expansions are seen in some tissues like the striatum (prior to pathological cell loss), liver and testis [20, 43]. Larger expansions are also seen in multiple tissues in juvenile HD cases where the inherited repeat size is significantly larger [43]. While the significance of somatic instability for FXD and FRDA disease severity is unclear at this time, emerging evidence suggests that the mechanism responsible for somatic expansions in these disorders is relevant for efforts to ameliorate somatic expansion in HD. This review will focus on what can be learned about the mechanism responsible for somatic CAG expansion in HD from mouse models of FRDA and the FXDs.
(GAA)\textsubscript{120} repeat tract [47]. Of these mouse lines, the YAC-based lines have been most extensively studied [46, 47]. These mice also display age-dependent and tissue-specific expansion of the GAA repeat in brain, cerebellum, dorsal root ganglia and liver tissues [46–48], similar to that seen in FRDA patient autopsy tissues [40–42].

The most intensively studied FXD mouse model is one containing 130+ uninterrupted CGG repeats in the endogenous murine \emph{Fmr1} gene [49]. These mice show both germline and somatic instability with a strong expansion bias as seen in humans [35]. The dynamics of somatic expansion in these animals resembles that seen in human carriers of expansion-prone \emph{FMR1} alleles [50]. They are also similar to those seen in HD patients [16–20] and in mouse models of HD [51–55] and individuals with DM1 [56], although with some differences in the extent of expansion in different tissues. The expansion profiles in all cases are consistent with a high frequency of relatively small expansions (1–3 repeats) [54], although larger expansions are also occasionally seen [57, 58]. For expansion-prone cell types like the mucosal cells of the small intestine, a ∼170 repeat allele expands as often as once every 5–6 days in the majority of cells in the population [59].

Several lines of evidence demonstrate that origin-independent chromosomal replication is not required for expansion at the \emph{FMR1} locus in either mice or humans. For example, expansion is seen in mouse oocytes [50], a cell-type that is non-dividing. Expansion in oocytes is consistent with the maternal age effect seen for expansion risk in humans [60]. Furthermore, expansion in somatic cells of the FXD mouse does not correlate with the tissue proliferation rate [35]. In mouse models of other repeat expansion disorders and in humans with such diseases expansion is seen in post-mitotic somatic cells such as neurons [17, 18, 38, 40, 61, 62].

Because the \emph{FMR1} gene is located on the X chromosome, the FXDs provide a particularly clear demonstration that expansion requires transcription or open chromatin. Specifically, expansion in female FXD mice, as in women with an expansion-prone \emph{FMR1} allele, only occurs when the expanded \emph{FMR1} allele is on the active X chromosome [63]. A dependence on transcription for expansion is consistent with correlations between transcriptional activity and expansion that are seen in transgenic mouse models of some of the CAG repeat expansion disorders [51, 64]. Thus, many characteristics of somatic expansion in the FRDA and FXD mice, including their small average size and high frequency, their transcription-dependence and replication-independence, are similar to those of mouse models of HD and other repeat expansion disorders, as well as human patients with these diseases.

Many genetic factors that affect somatic expansion have been identified in the FXD and FRDA mouse models. These include factors that are required for, or play a role in, promoting expansion and factors that protect against expansion. In principle, if there is a single expansion mechanism, then the factors essential for expansion should act in all cell types where expansion is observed. For factors that play an auxiliary role in expansion or those that are protective, their effect in different cell types may reflect the relative levels of other protective or promoting factors, as well as the stoichiometry and functional redundancy of such factors.

**THE ROLE OF MUTS AND MUTL COMPLEXES IN THE FRDA AND FXD MOUSE MODELS**

Somatic expansions in both the FXD KI and FRDA YAC mice involve proteins that are critical for normal mismatch repair (MMR) (Table 1). This includes one or both of the MutS complexes MutSα and MutSβ, that are involved in mismatch recognition; as well as one or more of the three mammalian MutL complexes, MutLα, MutLβ and MutLγ, that are involved in lesion processing. Specifically, MSH2, a constituent of both MutS complexes, plays an important role in repeat expansion in FRDA [65] and is essential for expansion in the FXD mice [66]. Loss of MSH6, the MSH2-binding partner in the MutSα complex, leads to a sharp reduction in expansions in the cerebellum of FRDA mice [65] and a >50% reduction in the extent of expansion in most FXD mouse tissues.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Effect</th>
<th>FXDs</th>
<th>FRDA</th>
<th>HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MutS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSH2</td>
<td>↑</td>
<td>[66]</td>
<td>[65]</td>
<td>[105, 106]</td>
</tr>
<tr>
<td>MSH3</td>
<td>↑</td>
<td>[68]</td>
<td></td>
<td>[62, 108]</td>
</tr>
<tr>
<td>MSH6</td>
<td>↑</td>
<td>[67]</td>
<td>[65]</td>
<td></td>
</tr>
<tr>
<td>MutL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLH1</td>
<td>↑</td>
<td></td>
<td>[69]</td>
<td>[107]</td>
</tr>
<tr>
<td>MLH3</td>
<td>↑</td>
<td></td>
<td>[59]</td>
<td></td>
</tr>
<tr>
<td>PMS1</td>
<td>↑</td>
<td></td>
<td>[70]</td>
<td>[107]</td>
</tr>
<tr>
<td>PMS2</td>
<td>↓</td>
<td></td>
<td>[70]</td>
<td></td>
</tr>
</tbody>
</table>

↑: promotes expansion. ↓: prevents expansion; --: no effect.
MLH1, a protein present in all three MutL complexes, is also an important contributor to the expansion process in the FRDA YAC mice [69]. However, PMS2, the MLH1-binding partner in the MutLα complex, protects against somatic expansion in these animals [65]. Given the importance of MLH1 in expansions, we can infer that one or more of the other MLH1-binding partners, PMS1 and/or MLH3, must play a role in expansions. In the case of the FXD mouse, all three MLH1-binding proteins, PMS1, PMS2 and MLH3, are required for expansion since the loss of any one of these proteins eliminates expansions either in vivo or in embryonic stem cells derived from these animals [59, 70]. Furthermore, a point mutation (D1185N) in the nuclease domain of MLH3 also eliminates all expansions in FXD mouse embryonic stem cells [71], suggesting that the nuclease activity of MutLγ is required.

Since MMR normally acts to prevent mismatches or insertion/deletions, it is generally thought that the role of these proteins in expansion reflects their ability to bind and process the secondary structures or loop-outs formed by the repeats that contain mismatched bases or regions of single-strandedness. However, given that MMR usually acts to prevent instability of similar tandem repeats or microsatellites, the processing of these structures presumably differs from canonical MMR. The specific requirement for MutLγ and its nuclease activity is interesting since it is much less abundant than MutLα [72], which typically plays a much larger role in MMR [73]. This suggests either that expansion involves an intermediate that is preferentially processed by MutLγ, or perhaps that MutLγ cleavage plays a unique and critical role in generating an intermediate that can be processed to generate an expansion. Notably, while MutLγ only plays a minor role in MMR, it plays a critical role in meiosis in resolving Holliday Junctions [74–76], which are cruciform-like structures that are also reminiscent of loop outs that could be formed by intrastrand structure formation by both strands of the repeats or perhaps simply by out-of-register reannealing. The role of MutLβ is also intriguing since it has no known nuclease motifs and, despite its abundance relative to MutLγ, its function is largely unknown [77].

In addition to MutS and MutL proteins, a variety of other genetic modifiers of somatic expansion risk have been identified in the FXD mouse (Table 2). Some of these factors promote expansion, whilst others are protective or neutral. The factors involved in the expansion process include DNA polymerase β (Polβ) [78]. Polβ is a DNA polymerase essential for base excision repair [79], as well as for gap-filling in other repair processes [80–82]. Its importance in the expansion process is evidenced by the fact that even heterozygosity for a hypomorphic allele resulted in a significant decrease in expansions in FXD mice [78]. Cockayne Syndrome B (CSB; aka ERCC6), a protein essential for transcription-coupled repair, contributes to, but is not required for, somatic expansion in older mice [83]. Since it is not essential for expansion, it is presumably acting outside of transcription-coupled repair to facilitate expansions, perhaps via its participation in other DNA processing pathways like base excision repair [84], chromatin remodeling [85–87] or R-loop induced double-strand break repair [88, 89].

Factors protecting against expansion include EXO1 [59], a 5′-3′ exonuclease that is involved in meiosis as well as MMR [90, 91]. Loss of EXO1 caused a significant increase in expansions in the germ line and in the small intestine, but not in the brain [59]. Moreover, a point mutation (D173A) in the active site of EXO1 also significantly increased the extent of expansion in small intestine and germ line, but not quite to the same extent as the EXO1 null mutation [59]. The D173A mutant protein is defective in MMR but retains the structural role of EXO1 in meiosis [92], suggesting that EXO1 protects against expansion both in a nuclease-dependent and a nuclease-independent manner. However, since the loss of EXO1 had no effect on somatic expansion

<table>
<thead>
<tr>
<th>Protein</th>
<th>Repair Pathways</th>
<th>Effect</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polβ</td>
<td>base excision repair/other</td>
<td>↑</td>
<td>[78]</td>
</tr>
<tr>
<td>CSB</td>
<td>transcription coupled repair/other</td>
<td>↑</td>
<td>[83]</td>
</tr>
<tr>
<td>FAN1</td>
<td>Fanconi anemia/other</td>
<td>↓</td>
<td>[58]</td>
</tr>
<tr>
<td>EXO1</td>
<td>MMR/other</td>
<td>↓</td>
<td>[59]</td>
</tr>
<tr>
<td>Lig4</td>
<td>non-homologous end-joining</td>
<td>↓</td>
<td>[57]</td>
</tr>
<tr>
<td>MRE11</td>
<td>homologous recombination/other</td>
<td>–</td>
<td>[57]</td>
</tr>
</tbody>
</table>

↑: promotes expansion. ↓: prevents expansion; –: no effect.
in organs other than the small intestine [59], EXO1 may be more relevant for germline rather than somatic expansion in the FXD mouse.

The loss of FAN1, another nuclease that has both 5′-3′ exonuclease and 5′ flap endonuclease activities [93–97], also causes a significant increase in somatic expansion in multiple tissues of FXD mice, including the brain [58]. Although named for its role in the Fanconi anemia (FA) pathway, FAN1 interacts with MLH1 [98] and has been suggested to be able to substitute for EXO1 in MMR [99].

DNA ligase IV (LIG4) also protects against somatic expansion in liver [57]. LIG4 is required for non-homologous end-joining [100–102], a process that is particularly important for neuronal integrity [103] since it is the major form of double-strand break repair active outside of S-phase. It is also the major repair pathway able to repair double-strand breaks with 5′ overhangs without requiring significant end resectioning. The protective effect of LIG4 suggests that non-homologous end-jointing and the expansion pathway compete for a common substrate and thus expansion likely involves an intermediate with a double-strand break. Other factors involved in other forms of double-strand break repair like MRE11, an exonuclease important for the end-resectioning required to generate the 3′ overhang necessary for homologous recombination and other related forms of double-strand break repair [104], does not affect repeat expansion in FXD mice [57]. In principle, the nucleases EXO1 and FAN1 might be able to compensate for the absence of MRE11. However, since these nucleases are both protective, this seems unlikely. Thus, evidence suggests that expansion in the FXD mouse involves a homologous recombination-independent and non-homologous end-joining-independent processing of a double-strand break intermediate, perhaps one generated by MutLγ cleavage.

The wide variety of different proteins that play a role in modulating repeat expansion presumably reflects the different ways that the same repeat DNA substrates can be processed, with competition between factors that promote expansion and those that protect against them. The relative abundance of these proteins/complexes in different cell types could contribute to the tendency of the repeat to expand more in some cell types and not others [35, 78]. For example, MSH6, which promotes expansion is highly expressed in brain, liver and testes, some of the most expansion prone tissues, whilst MSH2, a protein essential for expansion, is difficult to detect at all in heart, a tissue that shows little or no expansion [35].

**PARALLELS TO MOUSE AND PATIENT-DERIVED CELL MODELS OF OTHER REPEAT EXPANSION DISORDERS**

MutS and MutL complexes have also been shown to play major roles in somatic expansion in HD mouse models [62, 105–109] as well as mouse models of DM1 [110, 111], with a good correlation being observed between the levels of MutSβ and the extent of repeat expansion [112]. A critical role for MutS proteins in expansion has also been reported in cells from FRDA, HD and DM1 patients and mammalian model systems [113–117]. MutLγ has also been implicated in expansion in FRDA fibroblasts, where a role for the MLH3 nuclease has also been proposed [118]. As in the FXD mouse, FAN1 also protects against expansion in HD KI mice [119] and in HD induced pluripotent stem cells [120].

Despite the similarities seen across different disease-associated repeat loci, some differences are seen. For example, loss of MSH6 results in a significant suppression of somatic expansions in both FXD and FRDA mice [65, 67], a phenomenon that is also seen in FRDA patient induced pluripotent stem cells [113]. In contrast, in a HD mouse model, knockout of MSH6 had no obvious effect on expansions in striatum [62], whilst loss of MSH6 increased the expansion frequency in a human cell line carrying a (CAG)800 construct [116]. Similarly, in the DM1 mouse, a protective effect of MSH6 was seen in some organs like liver, but not in others, including brain [111]. Another example of a difference between disease models is PMS2, which promotes repeat expansions in FXD mouse embryonic stem cells [70] and in multiple tissues of DM1 mice [121], yet it seems to protect against GAA expansions in multiple brain regions of FRDA mice [65].

How is it that the same gene can have apparently opposing effects at different repeat expansion loci? Given that all of these diseases share many unusual features, it is possible that the occasional differences do not represent fundamentally different expansion mechanisms. For example, the differential effect of MSH6 may reflect the fact that MutSβ is able to promote MutSβ binding to mismatches [122] and to the hairpins formed by the FX repeats [67], an ability that may only be apparent when MutSβ is rate-limiting.
Alternatively, since typical MMR lesions are bound by two or more MutS dimers [123] and MutSα and MutSβ cofractionate from human cell extracts [124], and can both bind to the same FX hairpin [67], MutSα may be able to contribute to the MutSβ-bound lesion when MutSβ is limiting. The fact that loss of MSH6 has a bigger effect in the liver than in the brain in the FXD mouse model, would be consistent with this idea since liver has less MSH3 and more MSH6 than brain [35]. Similarly, in testis which shows the largest effect of the loss of MSH6, MSH6 is more abundant than it is in either brain or liver [35]. The apparently paradoxical effect of PMS2 on repeat expansion may result from competition between MutLγ and MutLα for binding to the expansion substrate as illustrated in Fig. 2.

The fact that the base excision repair polymerase, Polβ, plays an important role in promoting expansion in the FXD mouse is consistent with the fact that the loss of OGG1 and NEIL1, two DNA glycosylases involved in base excision repair, reduce expansion in HD mouse models [125, 126]. Since base excision repair is the major pathway by which oxidative damage to DNA is repaired in mammalian cells, this would also be consistent with the fact that oxidizing agents increase expansions in FXD mice [127] and in embryonic stem cells derived from HD mice [128]. The transcription-coupled repair protein CSB has also been reported to protect against expansion in HD mice [129]. However, this effect was only seen in an OGG1 background and thus its relationship to what is seen in the FXD mouse [83] is unclear. In contrast, the loss of XPA, a protein involved in the common steps of transcription-coupled repair and global-genome nucleotide excision repair, has been reported to reduce expansions in neuronal tissues of SCA1 mice, but not in gametes or liver [130]. This tissue-specific effect is reminiscent of the effect of CSB in FXD mice [83]. The fact that it is not essential for expansion suggests that, like CSB, it is acting independently of nucleotide excision repair, perhaps via its participation in other repair pathways [131].

Finally, recent work has shown that FAN1 has a clear protective role in HD patient cells [120] and HD KI mice [119], thus representing an additional parallel to the FXD mice [58].

Thus, in spite of some differences, a case can be made that many factors affecting somatic repeat instability in FRDA and FXD mouse models are similar in key respects to the genetic factors shown to affect somatic expansions in mouse and cell models of other repeat expansion disorders.

### The Relevance of These Models to Somatic Expansions in HD Patients and Affected Individuals with Other Repeat Expansion Disorders

As described in more detail elsewhere in this volume [132], recent GWA studies in HD patient cohorts have implicated loci containing some of the same DNA repair genes discussed above—FAN1, MSH3, MLH1, PMS1, and PMS2—as modifiers of somatic expansion, age at symptoms onset and disease progression [21, 22, 24–26]. Polymorphisms in MSH3 have also been shown to modify somatic expansion risk in DM1 patients [26, 133] and patients with CAG-repeat related spinocerebellar ataxias [25], while variants in FAN1 and PMS2 were also associated with variations in age at disease onset in patients with CAG-repeat related spinocerebellar ataxias [25]. Although CSB has not been implicated as a modifier of expansion risk in other diseases, a SNP in the CSB/ERCC6 gene has been shown to be associated with increased somatic instability in SCA3 [134].

In the case of FAN1, SNPs associated with earlier HD onset include missense variants within or near FAN1’s DNA-binding domain, consequently reducing its DNA-binding activity and capacity to rescue mitomycin C-induced cytotoxicity [135]. In addition, SNPs associated with later HD onset are associated with increased FAN1 expression in various brain regions [23, 135]. This is consistent with FAN1 activity protecting against expansions in humans, as in HD and FXD mice [58, 119, 120]. In contrast, again consistent with the results from model systems, SNPs associated with decreased MSH3 expression were associated with later disease onset, whilst SNPs associated with increased expression resulted in earlier ages of onset [22, 136]. TWAS has also shown a correlation between increased PMS2 expression and a later age at disease onset [23] which would be consistent with the protective effect of PMS2 in some mouse models. The situation is less clear for PMS1 where TWAS showed a variable effect, with an increase in cortex PMS1 expression being associated with a later HD onset [22], whilst data from other tissue sources suggest the opposite (P. Holmans, personal communication).

Thus, while loci associated with variations in expansion risk in patient GWA studies contain genes implicated in modifying expansion risk in mouse and patient-derived cell models, more work is needed to fully understand the contribution of these genes.
Fig. 2. MutL competition model for the differential effect of the loss of PMS2 seen in different cell types or disease models. A simple mathematical model was developed for the competition between MutL proteins for binding to the expansion substrates. This model used the following assumptions: 1) Only a small proportion of the total cellular MutL is actually available for binding to the repeat; 2) Any one of the three MutLs can be recruited to a MutS-bound substrate; 3) Three MutLs (a MutL trimer) are required to bind productively to a substrate [123]; 4) The available MutL complexes are distributed across all the substrates in proportion to their levels/binding affinity. 5) Only those MutL trimers that contain at least one MutL\(\gamma\) complex result in an expansion (indicated by a check mark); 6) Trimers that lack MutL\(\gamma\) or lesions that are not bound by at least three MutL complexes do not produce an expansion (indicated by a cross). A) Diagrammatic representation of the model showing MutL binding when the expansion substrates are present at different levels in the presence or absence of PMS2, with tick marks indicating outcomes that lead to expansions and the crosses those that do not. The number of available MutL complexes was set at MutL\(\alpha\) = 10; MutL\(\beta\) = 5 and MutL\(\gamma\) = 2, a ratio similar to that reported in mammalian cells [72]. When expansion substrate levels are low and PMS2 is present, not all MutL\(\gamma\) is bound, since PMS2 competes effectively for binding to the expansion substrate. As a result many MutL trimers formed lack MutL\(\gamma\) and their substrates are repaired without expansion. In the absence of PMS2, more MutL\(\gamma\) is able to bind and MutL\(\beta\) contributes to the formation of additional MutL trimers required for MutL\(\gamma\)-generated expansions. As a result, a net increase in expansions is seen relative to cells with PMS2. At intermediate levels of substrate more MutL\(\gamma\) is able to bind and when PMS2 is absent, the residual MutL\(\beta\) is sufficient for trimer formation at all Mut\(\beta\)-bound sites. This results in no net change in the expansion frequency relative to cells with PMS2. However, at high levels of substrate, MutL\(\beta\) becomes rate-limiting when PMS2 is absent, resulting in a net decrease in expansions. B) Graphical representation of the expansion probabilities across the range of substrate levels in the presence or absence of PMS2 based on the average of 1000 independent tests of the chances of binding of MutL\(\alpha\), MutL\(\beta\) and MutL\(\gamma\) for each of the substrate levels. The Python script used to generate the data upon which the graph is based is provided in the Supplementary Material. As in panel A, the number of available MutL complexes used was MutL\(\alpha\) = 10; MutL\(\beta\) = 5 and MutL\(\gamma\) = 2. However, as shown in the Supplementary Material similar results in terms of the range of effects of the loss of PMS2 are seen with wide range of different proportions of MutL\(\alpha\), MutL\(\beta\) and MutL\(\gamma\) and with a wide range of absolute levels of total MutL.
to expansion in disease-relevant organs like brain in humans. Additional factors that contribute to expansion risk in FXD mice may only become apparent with GWA studies on larger HD patient cohorts. However, given the already established similarities between expansion in the FXD mouse and HD models and patients, it is reasonable to think that many of these factors could well play a role in modulating expansion in HD as well as other repeat expansion disorders.

**IMPLICATIONS FOR THE MECHANISM OF SOMATIC EXPANSION IN HD**

Taken together, the available data suggests that most expansions occur via a transcription-dependent, replication-independent process, that involves components of multiple DNA repair pathways including MMR, base excision repair and some form of double-strand break repair. One way that all of these factors can be accommodated in a single model for repeat expansion is illustrated in Fig. 2.

Oxidative damage has been proposed to initiate repair of the damaged base with strand loop-outs arising by strand-slippage during Long Patch base excision repair [125, 137]. The loop-outs could then bind the MutS proteins to ultimately generate expansions. However, whether oxidative stress can account for the extraordinarily high expansion frequency that is seen in some mouse cells remains unclear; and it is possible that oxidative damage is not the only trigger for expansion. This would be consistent with data showing that antioxidants have a relatively small effect on somatic expansion frequencies in HD mice [138, 139]. In theory, similar loop-outs could be generated any time the repeat region was unpaired, during transcription for example, when hairpins could form on the non-template strand. It is also possible that simple out-of-register reannealing of the template and non-template strand could occur after the transcription complex has moved on. This could be exacerbated by the formation of stable R-loops that are characteristic of many of the repeat expansion loci associated with disease [49, 140–143]. Loop-outs formed on both strands would resemble a Holliday Junction, a preferred MutL binding substrate [144]. These loop-outs would be bound by MutSβ, and perhaps in some cases by MutSo as well [67]. The MutS-bound loop-outs would then be cleaved by MutLβ. This reaction may be facilitated by MutLβ in some way. Since non-homologous end-joining protects against expansion in the FXD mouse [57], it suggests that expansion proceeds via a double-strand break that is then processed by a mechanism that is independent of non-homologous end-joining and homologous recombination. MutLγ is known to cleave the strand opposite loop-outs in vitro [145]. Loop-outs on both strands may result in off-set cleavages as illustrated in Fig. 3. This could result in staggered double-strand break with 5’ overhangs that could then anneal out-of-register resulting in small gaps. Simple gap filling, a reaction that can be carried out by Polβ [146, 147], followed by ligation would generate an expanded allele. CSB and XPA may affect expansion in a number of different ways, including Fig. 3. Double-strand break model for the generation of repeat expansions. Expansion in this model is initiated when the repeat is transiently unpaired, as for example during transcription, replication or DNA damage repair. Out-of-register annealing of the two strands during this process could result in a double loop-out structure that resembles a Holliday Junction, the normal MutLβ meiotic substrate. This process may be exacerbated by the ability of the individual strands of some repeats to form stable intrastrand secondary structures like hairpins. Cleavage by MutLγ on either side of the double loop-out results in a double-strand break that can anneal out of register. Simple gap filling and ligation then results in expansions.
via binding to R-loops [148] or the inhibition of non-homologous end-joining in the case of CSB [149], and the ability to bind to Holliday junctions [150] in the case of XPA. FAN1 and EXO1 may reduce the likelihood that this pathway is used by digesting the broken ends in such a way as to favor their processing to restore the original allele or to generate contractions. EXO1 also has a structural role in determining the orientation of MutLγ cleavage [92, 151] that may explain its nuclease-independent role in preventing expansions [59].

This particular expansion pathway may occur in parallel with other potential expansion pathways that have been described [44, 152–155]. However, since these pathways do not involve the MMR repair proteins implicated by GW AS in patient cohorts, their contribution to somatic expansion in the repeat expansion diseases is unclear.

CONCLUDING REMARKS

The genetic data from FRDA, FXD, and HD mice and human cell models, as well as GWA/TWA studies in HD patients, strongly implicate a subset of MMR components as major contributors to somatic repeat expansions. However, studies in the FXD mice suggest that there are a variety of additional DNA repair factors and pathways that could be targeted to reduce somatic expansion in HD, as well as other repeat expansion disorders in which somatic instability is a contributor to disease burden. This naturally increases the potentially druggable space for somatic expansion in HD patients and increases the chances that a suitable modifier can be identified that is amenable to safe modulation and efficient drug targeting. A common druggable target for all repeat expansion diseases may make the development of a more broadly useful drug a more economically feasible endeavor.

ACKNOWLEDGMENTS

The authors would like to thank Drs Peter Holmans and Tom Massey (Cardiff University) for helpful conversations. This work was made possible by funding from the Intramural Program of the National Institute of Diabetes, Kidney and Digestive Diseases to KU, and the Huntington’s Disease Society of America to RMP.

CONFLICT OF INTEREST

RMP has received sponsored research funding from Pfizer unrelated to the content of this manuscript.

Other authors have no conflict of interest to report.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JHD-200423.

REFERENCES


