



## Disease-associated repeat instability and mismatch repair



Monika H.M. Schmidt<sup>a,b</sup>, Christopher E. Pearson<sup>a,b,\*</sup>

<sup>a</sup> Genetics & Genome Biology, The Hospital for Sick Children, Peter Gilgan Centre for Research & Learning, 686 Bay St., Toronto, Ontario M5G 0A4, Canada

<sup>b</sup> Department of Molecular Genetics, University of Toronto, Medical Sciences Bldg., 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada

### ARTICLE INFO

#### Article history:

Received 16 November 2015

Received in revised form

23 November 2015

Accepted 30 November 2015

Available online 12 December 2015

#### Keywords:

Mismatch repair

Trinucleotide repeat instability

MSH2

MSH3

MSH6

MLH1

PMS2

MLH3

MutS $\beta$

MutS $\alpha$

Slipped-DNA

R-loops

Error-prone repair

Huntington's disease

Myotonic dystrophy

Fragile X

Friedreich's ataxia

Amyotrophic lateral sclerosis

DNA repair gene polymorphism

Tissue-specific

### ABSTRACT

Expanded tandem repeat sequences in DNA are associated with at least 40 human genetic neurological, neurodegenerative, and neuromuscular diseases. Repeat expansion can occur during parent-to-offspring transmission, and arise at variable rates in specific tissues throughout the life of an affected individual. Since the ongoing somatic repeat expansions can affect disease age-of-onset, severity, and progression, targeting somatic expansion holds potential as a therapeutic target. Thus, understanding the factors that regulate this mutation is crucial. DNA repair, in particular mismatch repair (MMR), is the major driving force of disease-associated repeat expansions. In contrast to its anti-mutagenic roles, mammalian MMR curiously drives the expansion mutations of disease-associated (CAG):(CTG) repeats. Recent advances have broadened our knowledge of both the MMR proteins involved in disease repeat expansions, including: MSH2, MSH3, MSH6, MLH1, PMS2, and MLH3, as well as the types of repeats affected by MMR, now including: (CAG):(CTG), (CGG):(CCG), and (GAA):(TTC) repeats. Mutagenic slipped-DNA structures have been detected in patient tissues, and the size of the slip-out and their junction conformation can determine the involvement of MMR. Furthermore, the formation of other unusual DNA and R-loop structures is proposed to play a key role in MMR-mediated instability. A complex correlation is emerging between tissues showing varying amounts of repeat instability and MMR expression levels. Notably, naturally occurring polymorphic variants of DNA repair genes can have dramatic effects upon the levels of repeat instability, which may explain the variation in disease age-of-onset, progression and severity. An increasing grasp of these factors holds prognostic and therapeutic potential.

© 2015 Elsevier B.V. All rights reserved.

### Contents

1. Introduction .....	118
2. Disease-associated repeats tend to expand .....	118
2.1. MMR promotes instability of (CAG):(CTG) repeats <i>in vivo</i> .....	118
2.2. Tissue-specificity of repeat instability— a role for DNA repair? .....	118
2.3. MMR promotes instability of (CGG):(CCG) and (GAA):(TTC) repeats .....	119

**Abbreviations:** MMR, mismatch repair; MSH, mutS homolog; MLH, mutL homolog; BER, base excision repair; NER, nucleotide excision repair; TNR, trinucleotide repeat; DM1, myotonic dystrophy type 1; HD, huntington's disease; FRDA, Friedreich's ataxia; FXD, fragile X disease; FXS, fragile X syndrome; FRAXA, Fragile X type A; hESC, human embryonic stem cell(s); B6, C57BL/6j mouse background; CBy, BALB/cByJ mouse background; Pol $\beta$ , polymerase beta; Pol $\delta$ , polymerase delta.

\* Corresponding author.

E-mail address: [cepearson.sickkids@gmail.com](mailto:cepearson.sickkids@gmail.com) (C.E. Pearson).

<http://dx.doi.org/10.1016/j.dnarep.2015.11.008>

1568-7864/© 2015 Elsevier B.V. All rights reserved.

3.	Structural effects on repeat instability .....	120
3.1.	Expanded repeat sequences form unusual DNA structures .....	120
3.2.	Mechanism of MMR-mediated expansions: slip-out size & junction conformation matters .....	120
3.3.	MutS $\beta$ interaction with repeat tracts: crystal structure of MutS $\beta$ -DNA .....	122
3.4.	Transcription-mediated instability and a role for MMR .....	123
4.	Contractions of expanded repeat sequences .....	124
4.1.	MMR deficiencies can lead to (CAG):(CTG) and (CGG):(CCG) contractions .....	124
4.2.	Polymorphisms in DNA repair genes can regulate levels of (CAG):(CTG) instability .....	124
5.	Conclusions & future goals .....	124
	Conflict of interest .....	124
	Acknowledgments .....	124
	References .....	125

## 1. Introduction

Gene-specific repeat expansions are the cause of 43 genetic diseases, 27 of which are known to be affected by MMR (Table 1). In non-affected individuals, repeat tracts are short and genetically stable, while long, expanding repeats occur in affected individuals. Larger and ongoing expansions reduce the age-of-onset, and drive disease progression and severity.

Many DNA metabolizing pathways have been assessed for their contribution to disease-associated repeat expansions, including DNA replication, base excision repair (BER), double-strand break repair, nucleotide excision repair (NER), and recombination (reviewed in Refs. [1,2]). MMR is the strongest driver of repeat expansions (reviewed in Refs. [2,3]). Despite its role as an anti-mutation system, in certain circumstances, MMR can also drive mutations required for immune system development [3] (reviewed by Zanotti and Gearhart in this Special Issue) or for disease-causing repeat mutations. Some MMR proteins have striking effects on repeat instability (Table 2).

Most studies of repeats in bacteria, yeast, and flies reveal that MMR either has no effect or has a different effect than what occurs in mammals, including affected families [5–7]. This review focuses upon recent advances of how mammalian MMR is involved in disease-associated repeat instability, with limited coverage in other systems.

## 2. Disease-associated repeats tend to expand

### 2.1. MMR promotes instability of (CAG):(CTG) repeats in vivo

The frequency and tissue-specificity of (CAG):(CTG) expansions in transgenic mice closely reflects the pattern in affected humans. Evidence for a role of MMR protein MutS $\beta$  (MSH2-MSH3) on (CAG):(CTG) instability is supported by the requirement of MSH2 for the near 100% repeat expansions in a Huntington's disease (HD) and four myotonic dystrophy type 1 (DM1) transgenic mice [2]. A genetic loss of MSH2 completely ablates CAG expansions [8]. Similarly, a deficient MSH2 ATPase domain, that impairs MMR, also ablates repeat expansions [9]. Thus, MSH2's role in repeat expansions is not limited to binding and stabilization of CAG hairpins [10]. Deficiency of MSH3 in DM1 and HD mice suppresses expansions [11,12], an effect that is MSH3-dose dependent [12]. Conversely, MSH6 is not required in (CAG):(CTG) expansions, discussed further in the Contractions section below.

Both MutL $\alpha$  (MLH1-PMS2) and MutL $\gamma$  (MLH1-MLH3) are involved in driving (CAG):(CTG) expansions *in vivo* [4,13,14]: *Pms2*-null mice show up to a 50% decrease in somatic expansions and increased frequency of large, albeit rare, contractions, while *in vivo* deficiencies of MLH1 or MLH3 completely ablate CAG expansions [4]. The effect of MutL $\alpha$  may be less than that

of MutL $\gamma$ , since deficiency of PMS2 had less impact on instability than deficiency of MLH3. Nonetheless, both subunits of hMutL $\alpha$  (hMLH1 and hPMS2) are required for *in vitro* processing of short, but not long (CAG):(CTG) slip-outs [15]. Thus, the MutL complexes may play similar but not equal roles in (CAG):(CTG) instability.

Cellular models also revealed a requirement of MMR in repeat instability: knocking down MutS $\beta$  in Friedreich's ataxia (FRDA) or DM1 patient-derived cells blocked (GAA):(TTC) and (CAG):(CTG) expansions, respectively [16,17]. Interestingly, knocking down MutS $\beta$  suppressed (CAG):(CTG) expansions, while absence of MutS $\alpha$  leads to expansions exceeding 200 repeats, possibly due to MutS $\beta$  upregulation [18]. Expression of MMR proteins naturally coincides with (CAG):(CTG) expansions, as observed in human embryonic stem cells (hESCs) derived from oocytes and sperm of DM1 and HD parents [19]. In these hESCs, the expression level of MMR proteins coincides with active CTG instability [19]. Upon differentiation of the hESCs, a loss of MMR protein expression was concomitant with stabilization of the repeat [19]. Thus, timing of MMR protein expression is likely important in mediating repeat expansions.

The effects of MMR on (CAG):(CTG) instability in yeast are strikingly different from mice and humans. Yeast predominantly display (CAG):(CTG) contractions, with rare expansions [5,6,19a–c] (contractions at  $10^{-3}$ /generation versus expansions at  $10^{-5}$ /generation). This contrasts humans, who display a near-absolute expansion bias. Expansions in yeast do not appear to be affected by MMR, where a loss of MSH2 or MLH1 does not alter repeat instability [5,6,19a]. Schweitzer & Livingston found that deficiencies of MSH2 or PMS1 (the yeast homolog of PMS2) led to increased levels of contractions and a mild increase of +1 repeat expansions [19b]. Using a yeast model that reports only expansions or only contractions, the Lahue lab did not find any effect for a loss of MSH2 [6,19a]. Interestingly, using the expansion-only model of Lahue, the Surtees lab confirmed that a loss of MSH2 had no effect, however a loss of MSH3 decreased expansions ( $10^{-5}$  to  $10^{-6}$ ) while a loss of MSH6 increased expansions decreased ( $10^{-5}$  to  $10^{-4}$ ), the effect upon contractions was not reported [19c].

### 2.2. Tissue-specificity of repeat instability—a role for DNA repair?

The size of an unstable repeat tract varies between tissues of an affected individual [1–3,20], with the largest expansions arising in the most severely affected tissues. This supports the concept that somatic expansions occur as an individual ages, driving disease progression and severity. The source of this tissue-specific instability has been an area of intense research, but remains poorly understood (reviewed in Refs. [1,3,20]). Genetic ablation of MutS $\beta$ , MLH1, MLH3, and to a lesser degree PMS2, can suppress CAG expansions in all tissues. Genetic deficiencies of other repair proteins can

**Table 1**  
**Disease-associated unstable repeats affected by mismatch repair.**

Repeat-associated disorder	Gene(s)
<b>(CAG)n-(CTG)n - 16 disease loci</b>	
SBMA: spinal and bulbar muscular atrophy	AR
HD: Huntington's disease	HTT/HTTAS
HDL2: Huntington's disease-like 2	JPH3/JPH-AS
Spinocerebellar ataxias: SCA1, SCA2, SCA3/MJD, SCA6, SCA7, SCA8, SCA12, SCA17	ATXN1, ATXN2 or KCNN3, ATXN3/MJD, CACNA1-A, ATXN7/ATXN7-AS, ATXN 8/ATXN8-AS, PPP2R2B, ATXN17 (respectively)
Schizophrenia, Bipolar disorder	KCNN3
Fuch's Endothelial Corneal Dystrophy 2	FECD2
DM1: myotonic dystrophy type 1	DMPK/DMPK-AS
Dentatorubropallidolusian atrophy	ATN1
Breast cancer risk factor	AIB1, also known as NCOA3, SRC-3, ACTR, pCIP, RAC3, and TRAM1
<b>(GAA)n-(TTC)n - 1 disease locus</b>	
FRDA: Friedreich's ataxia	FXN/FAST-1
<b>(CGG)n-(CCG)n - 10 disease loci</b>	
FRAXA: fragile X syndrome/fragile X tremor/ataxia syndrome	FMR1/FMR1-AS
FRAXE,FRAXF: fragile X syndrome	FMR2,FRM3
Various neurological phenotypes or developmental delays/intellectual disabilities	FRA2A, FRA7A, FRA10A, FRA11A, FRA11B, FRA12A, FRA16A

**Table 2**  
**MMR protein functions and effects on (CAG)-(CTG) instability.**

Gene/Protein	Function(s)	Repeat length effect	Mouse germline/somatic instability
MSH2	MutS $\alpha$ /MutS $\beta$ complex component	↑ Repeat length	√√√
MSH3	Forms MutS $\beta$ complex with MSH2	↑ Repeat length	√√√
MSH6	Forms MutS $\alpha$ complex with MSH2	Stabilize/↓ repeat length	-
MLH1	Part of MutL $\alpha$ , MutL $\beta$ and MutL $\gamma$	↑ Repeat length	√√
MLH3	Forms MutL $\gamma$ complex with MLH1; endonuclease	↑ Repeat length	√√
PMS2	Forms MutL $\alpha$ complex with MLH1; DNA endonuclease	↑ (CAG) or (CGG) length ↓ (GAA) length	√
MCM9	Initiation of replication and factor in homologous recombination repair	Stabilize repeat length	√√√
PCNA	Helps load exonuclease 1; recruits pol $\delta$ and ligase1 to fill in gap	Possible stabilization of repeat length	Unclear
RPA	Single-stranded binding protein	Unknown	Unknown
LIG1	Ligates nicked DNA fragments following replication and/or repair	↑ Repeat length	√ (Maternal germline)

affect CAG instability, in some, but not all tissues of transgenic mice [21,22]. Thus, tissue-specific patterns of CAG instability may arise by different mechanisms.

Towards identifying factors responsible for the tissue-specific CAG instability patterns, several groups found that stoichiometric levels of repair proteins are associated with variable levels of CAG instability between the striatum and cortex of HD mice [23,24]. This is supported by tissue-specific gene expression of various repair genes in mice and humans [25]. However, levels of repair protein expression may not be the only factor affecting tissue-specific instability: expression levels of MMR genes have been associated with CAG expansions in some cases [19]. Analysis of 14 different mouse tissue types revealed widely varying levels of MMR proteins between tissues, but no clear correlation with CAG expansion levels [26]. Thus, MMR protein levels may affect repeat instability in some, but not all, tissues. In the HD-susceptible medium spiny striatal neurons, MSH2 can drive both CAG expansions and disease marker expression, which can be blocked by deficiency of MSH2 [27a]. These findings provide support for the role of MSH2 as a promoter of disease by driving CAG expansions. A possible factor that may mediate the tissue-specific CAG instability may be the tissue-specific chromatin packaging of the repeat. Interestingly, the activation of MutS $\alpha$  can be determined by specific chromatin modifications [27b]. Future research must continue to search for factors regulating tissue-specific patterns of CAG expansions.

### 2.3. MMR promotes instability of (CGG)-(CCG) and (GAA)-(TTC) repeats

Might MMR treat other expandable disease-associated repeats in a manner similar to (CAG)-(CTG) repeats? MMR modifies instability of expanded (CGG)-(CCG) repeats, associated with fragile X syndrome (FXS) and fragile X tremor ataxia syndrome (FXTAS). Somatic and germline expansions of unmethylated pre-mutation-length CGG tracts can arise in human and murine tissues, and increase with age [28–30]. In adult patients, the fully-methylated expansions appear somatically stable [28,31,32]. In mice, expansions of these unmethylated repeats depend upon MSH2 and MSH3, and the effects were dose-dependent, as *Msh2*- or *Msh3*-hemizygoty showed intermediate levels of expansions [30]. Notably, an absence of MSH2 or MSH3 increased germline and somatic CGG contractions [33].

Both MutS $\beta$  and MutS $\alpha$  drive (GAA)-(TTC) expansions in patient-derived FRDA cells [16,34,35]. However, MSH2 and MSH3 are not required for intergenerational (germline) GAA repeat expansions [36], which is in stark contrast with their absolute requirement for (CAG)-(CTG) and (CGG)-(CCG) expansions (germline and somatic). MutS $\beta$  and MutS $\alpha$  seemed to protect against germline GAA repeat contractions, similar to (CAG)-(CTG) repeats [36], while PMS2 protected against germline GAA expansions, and promoted contractions [36].

MLH1 and PMS2 have been suggested to mediate transcription of the *FXN* gene by unknown mechanisms [37], which may

suggest the involvement of MMR upon transcription-mediated instability, as well as repair. Mice deficient in MLH1 or PMS2 also exhibit reduced intergenerational and somatic GAA expansions [37]. Curiously, MLH1 and PMS2 appear to have opposing effects on (GAA):(TTC) instability: MLH1 promotes expansions, while PMS2 suppresses large expansions [38]. Interestingly, PMS2 suppressed (GAA):(TTC) expansions in tissues with MutS $\alpha$ -dependent instability [38], indicating that PMS2 and MutS $\alpha$  may overlap in the DNA lesions or structures targeted for repair.

The reason MMR handles (GAA):(TTC) repeat instability so differently from its handling of (CAG):(CTG) or (CGG):(CCG) repeats in germline transmissions, may be due to the mutagenic structures formed by (GAA):(TTC) repeats. Expanded (GAA):(TTC) repeats can form intramolecular triplexes: (GAA):(GAA):(TTC), which may block replication fork progression [39]. Alternately, (GAA):(TTC) instability may involve aberrant processing of R-loop structures, revealed to form on expanded repeats of the *FXN* and *FMR1* genes in patient cells [40]. Tandem copies of the integrated transgene each containing GAA tracts in mice [37] may permit unusual structure formation [41]. These findings reinforce the need to examine unusual structures formed by (CGG):(CCG) and (GAA):(TTC) expanded repeats, the ability for MMR, and other proteins to process these structures, and their combined effects on repeat instability, transcriptional activity, protein expression and disease progression.

### 3. Structural effects on repeat instability

#### 3.1. Expanded repeat sequences form unusual DNA structures

Repeat sequences can form various unusual DNA or RNA:DNA structures (Figs. 1 and 2). Such structures may form during processes that unwind the DNA, including replication, recombination, repair, and transcription. These structures may be aberrantly processed leading to repeat instability. Formation of slipped-DNAs has been observed *in vitro* with (CAG):(CTG) repeats [42,43]. Individual or clustered slip-outs can arise with excesses of either (CAG) or (CTG), and may have variable numbers of slipped repeat units (Fig. 1). *In vivo* evidence of slipped-DNA has been seen during DNA replication in cell models of CAG instability using hairpin-specific zinc-finger nucleases [44]. Slipped (CAG):(CTG) DNAs are present at the DM1 disease locus in patient tissues [45]. Strikingly, allelic levels of slipped-DNA-containing molecules were greater in affected tissues that showed the highest levels of expansions, compared to lower levels of slipped-DNAs in non-affected tissues with smaller expansions. This finding strongly supports an *in vivo* involvement of slipped-DNAs in somatic repeat instability in humans.

Other unusual DNA or RNA:DNA structures may also be critical to MMR-mediated instability. For example, the (CGG):(CCG) repeat of *FXS* has been shown to form slipped-DNAs, hairpins, G-quadruplexes, and Z-DNA [46], while the (GAA):(TTC) repeat of *FRDA* can form hairpins and triple-stranded structures; both of these repeats have recently been shown to be affected by MMR [29,30,47,48]. *In vivo* evidence for these structures of the fragile X disease (FXD) or *FRDA* repeats, or their involvement in repeat instability has, to date, remained elusive. R-loops, or RNA:DNA hybrids (Fig. 2), can form during unidirectional or bidirectional transcription [49] across (CAG):(CTG), (CGG):(GCC), (GAA) but not (TTC), and (G<sub>4</sub>C<sub>2</sub>)-(G<sub>2</sub>C<sub>4</sub>) repeats, and persist following transcription [40,50]. Processing of R-loops at repeats may lead to instability, which could involve the formation and processing of slipped-DNAs (Fig. 2).

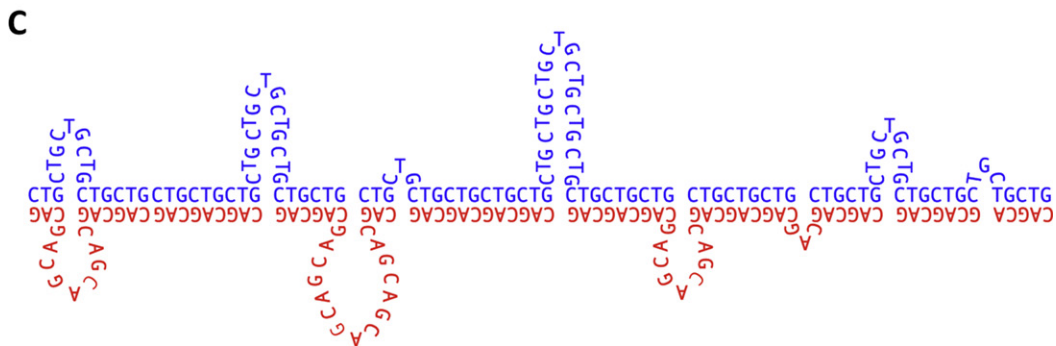
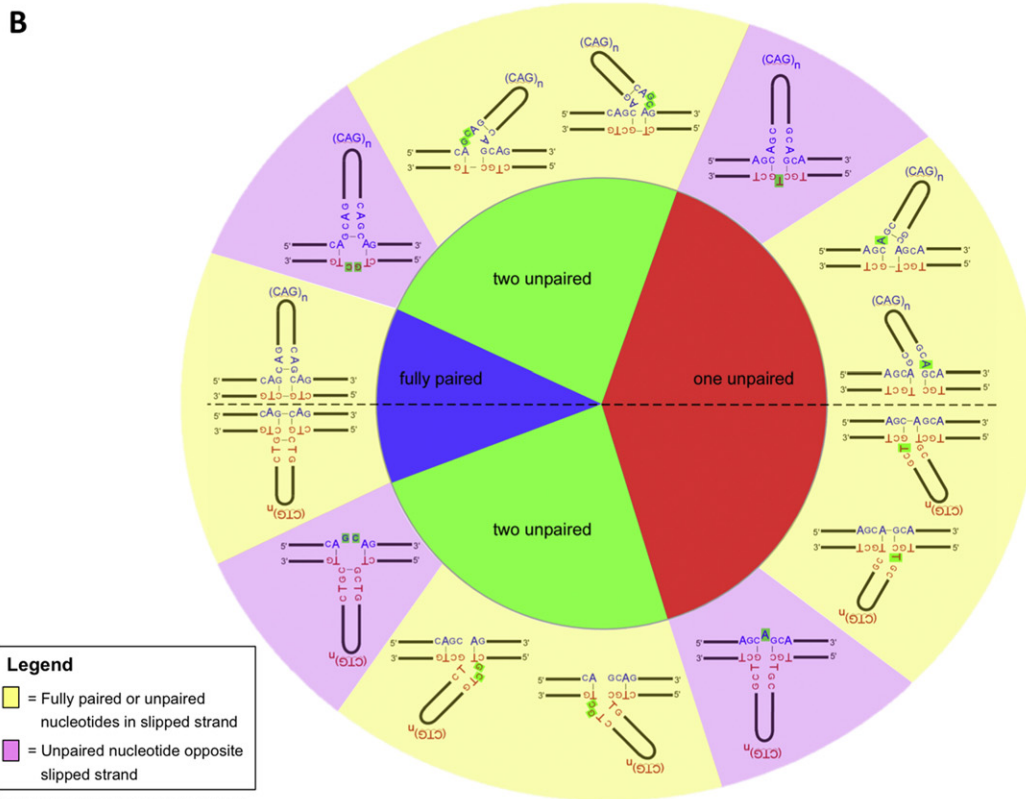
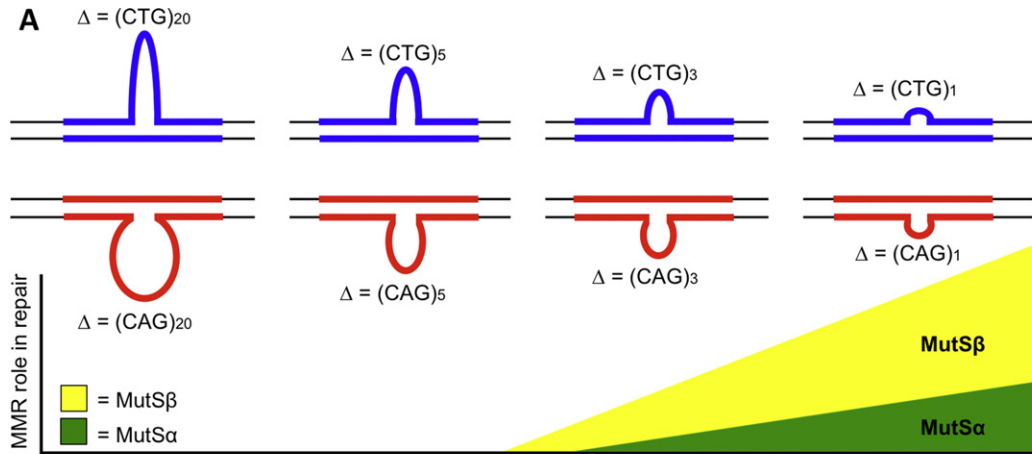
#### 3.2. Mechanism of MMR-mediated expansions: slip-out size & junction conformation matters

The involvement of MMR in processing slipped-(CAG):(CTG) repeats depends upon the size of the slip-out: specifically, shorter slip-outs depend upon MutS $\beta$  (MSH2-MSH3), while longer slip-outs do not (Fig. 1A). MutS $\beta$  is required for repair of isolated, short slip-outs (1–3 excess repeat units) [51], while slip-outs with >3–25 excess repeat units are processed independently of MMR [52,53]. The effect of slip-out size on repair efficiency was confirmed [53], however an involvement of MMR was not seen for these repeat slip-outs, likely because the shortest slip-out size examined had an excess of 5 repeats, which is beyond the length threshold of  $\leq 3$  excess repeat units, identified as a requirement of MutS $\beta$  [51]. The role of slip-out size in determining MMR involvement, initially identified for MutS $\beta$  [51] has been extended to the MutL $\alpha$  (MLH1-PMS2) MMR endonuclease [54]. MutL $\alpha$  can be activated on DNA containing a MutS $\alpha$ - or MutS $\beta$ -bound DNA-lesion, in conjunction with proliferating cell nuclear antigen [54]. Short MutS $\beta$ -bound slip-outs of 2–3 (CAG)<sub>n</sub> or (CTG)<sub>n</sub> repeat units can trigger MutL $\alpha$  activation and subsequent repair, a reaction that is biased for nicked DNA strands [53,54]. However, in the absence of a pre-existing nick, MutL $\alpha$  may incise either of the DNA strands harbouring a short slip-out, leading to repair that occurs without an obvious strand bias—a potential source of repeat instability [54]. Curiously, *in vitro* studies did not reveal an obvious requirement of human MutL $\gamma$  (MLH1-MLH3) [4,15], which contrasts with the apparent absolute requirement of murine MLH3 in HD mice [4]. Together, the requirement of short-slip-outs (isolated or clustered) to involve both MutS $\beta$  and MutL $\alpha$  are consistent with the role of short slip-outs in repeat expansions.

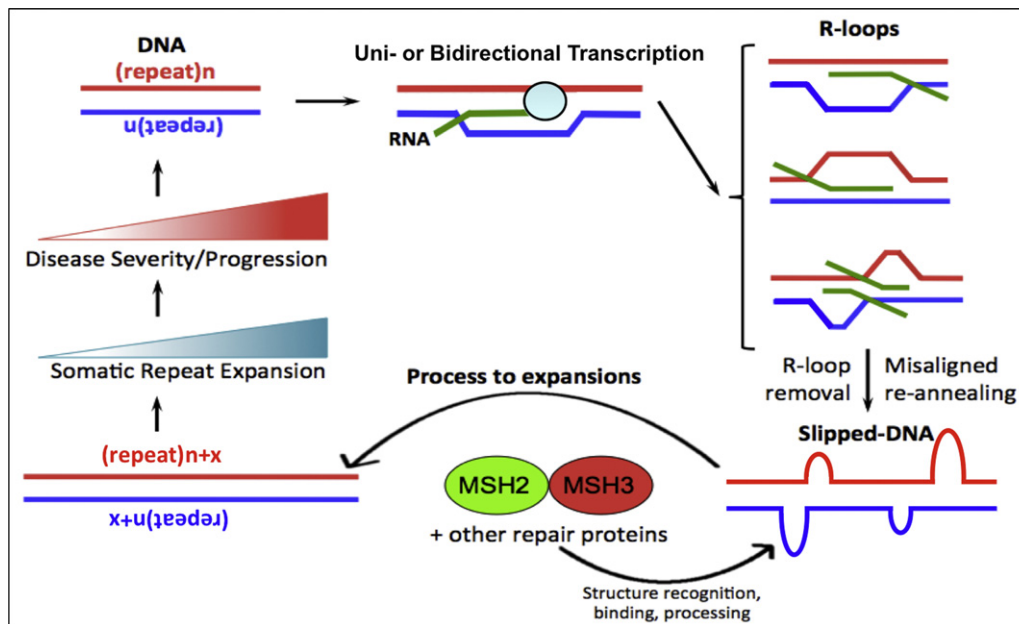
Clusters of short slip-outs with 1–3 excess repeat units along a molecule, require MutS $\beta$  for initiation of repair, but are not efficiently repaired. This may be due to closely clustered slip-outs acting as blockades to repair of adjacent slip-outs (Fig. 1C) [51]. Patient tissues harbor clustered slip-outs [45]; aberrant processing of these may lead to expansions that can occur in the absence of DNA replication.

Slip-out junction conformation can affect repair outcome. Slipped-junctions can have fully base-paired junctions, or junctions with 1–2 unpaired nucleotides opposite the slip-out [55] (Fig. 1B). These junction conformations are in dynamic inter-converting equilibrium with each other. Different junction conformations may confuse the MMR system: a lesion present on only one strand (a fully paired slip-out without unpaired nucleotides on the opposite strand) may be processed differently from a slip-out that also contains 1–2 unpaired nucleotides opposite the slip-out. In the latter case, repair may be directed to the incorrect strand, resulting in MMR-mediated mutagenic repeat expansion [55].

Repeat expansions are likely the result of accumulated short incremental expansions. Until recently, it was unclear whether repeat expansions arise from large jumps of repeats during a single mutation event or by multiple stepwise increases of short expansions. Saltatory large jumps of expansions have been observed in bacteria [56]. However, since human MutS $\beta$  is required to repair short repeat slip-outs suggests that the increment of change during a single mutagenic step might be 1–3 repeat units, similar to those occurring at other simple repeat sequences such as (CA)<sub>n</sub> and (A)<sub>n</sub> repeats [57,58]. Limited evidence from patient tissues supports incremental changes [59–62]. An elegant study in yeast strongly supports that net increases of CAG expansions are the result of small incremental expansions in the presence of MMR [63] and these incremental expansions may be affected in the absence of MMR [19b]. Aberrant processing of slipped-DNAs may involve an orchestration between the proofreading exonuclease activity polymerase  $\delta$  (Pol $\delta$ ) and polymerase  $\beta$  (Pol $\beta$ ), whereby Pol $\beta$  allows



**Fig. 1. Unusual DNA structures implicated in repeat instability and MMR processing.** Mispairing and re-annealing of complementary repeat strands following unwinding of the DNA may lead to the formation of slipped DNA structures. (A) Repeat sequences may snap-back or pair with themselves while transiently single-stranded, leading to the formation of loop-outs (red) or hairpin structures (blue), which are targets for MMR. MutS $\beta$  is implicated to be involved in recognition and processing of these slipped DNAs, with a preference for short (CNG) slip-outs containing  $\leq 3$  extruded repeat units. MutS $\alpha$  also plays a role in repair of short repeat slip-outs, however its contribution to repair of these slipped-DNA structures is less than MutS $\beta$  (B) Slipped-DNAs form interconverting conformations of three-way junctions. Unpaired nucleotides may arise in either of junction arms (yellow), or opposite the slip-out (pink); the resulting conformation may affect binding by MMR proteins and correct repair processing. Junctions containing unpaired nucleotides opposite the slip-out could be mistakenly treated as the lesion, rather than the slip-out. (C) Clustered slip-outs may form on one or both strands of the DNA; adjacent slip-outs are poorly repaired, possibly due to bulky structure interference. Clustered slip-outs have been detected in patient tissues at the DM1 disease locus (Axford *et al.* [45]).



**Fig. 2. Proposed mechanism of repeat expansion mediated by R-loops and MMR.** Expanded repeat-containing DNA is unidirectionally or bidirectionally transcribed. The RNA transcript remains bound to the template DNA, forming a stable RNA:DNA hybrid (R-loop). Unpaired single stranded DNA in single-R-loops may anneal with itself at this point, forming unusual structures. Removal/degradation of the R-loop pre-disposes repetitive DNA to misalignment upon re-annealing, providing the opportunity for further unusual DNA structures, such as slip-outs, to form. These structures are recognized by MMR proteins, in particular MutS $\beta$ , and undergo repair, leading to instability and possible disease-associated repeat expansions.

Pol $\delta$  to escape excision, promoting further expansion [64]. Such activities may lead to short incremental expansions.

Expansions can arise in DNAs devoid of pre-formed structures. Incubating a fully-duplexed (CAG)<sub>22</sub>·(CTG)<sub>22</sub> repeat in a yeast shuttle vector in human cell extracts can lead to expansions detectable in a sensitive yeast screen [65]. Expansions arose at frequencies of 1/1000–12/1000 plasmids, with increases of 4–18 repeats. Expansions did not require either replication or MutS $\beta$ , however frequencies were enhanced in the presence of MutS $\beta$ . While not demonstrated, the authors presumed that expansions arose via mis-repair of DNA damage induced by the human cell extract. Reconstitution *in vitro* of this mutagenic pathway provides potential for mechanistic evaluation of repeat expansions and contractions.

### 3.3. MutS $\beta$ interaction with repeat tracts: crystal structure of MutS $\beta$ -DNA

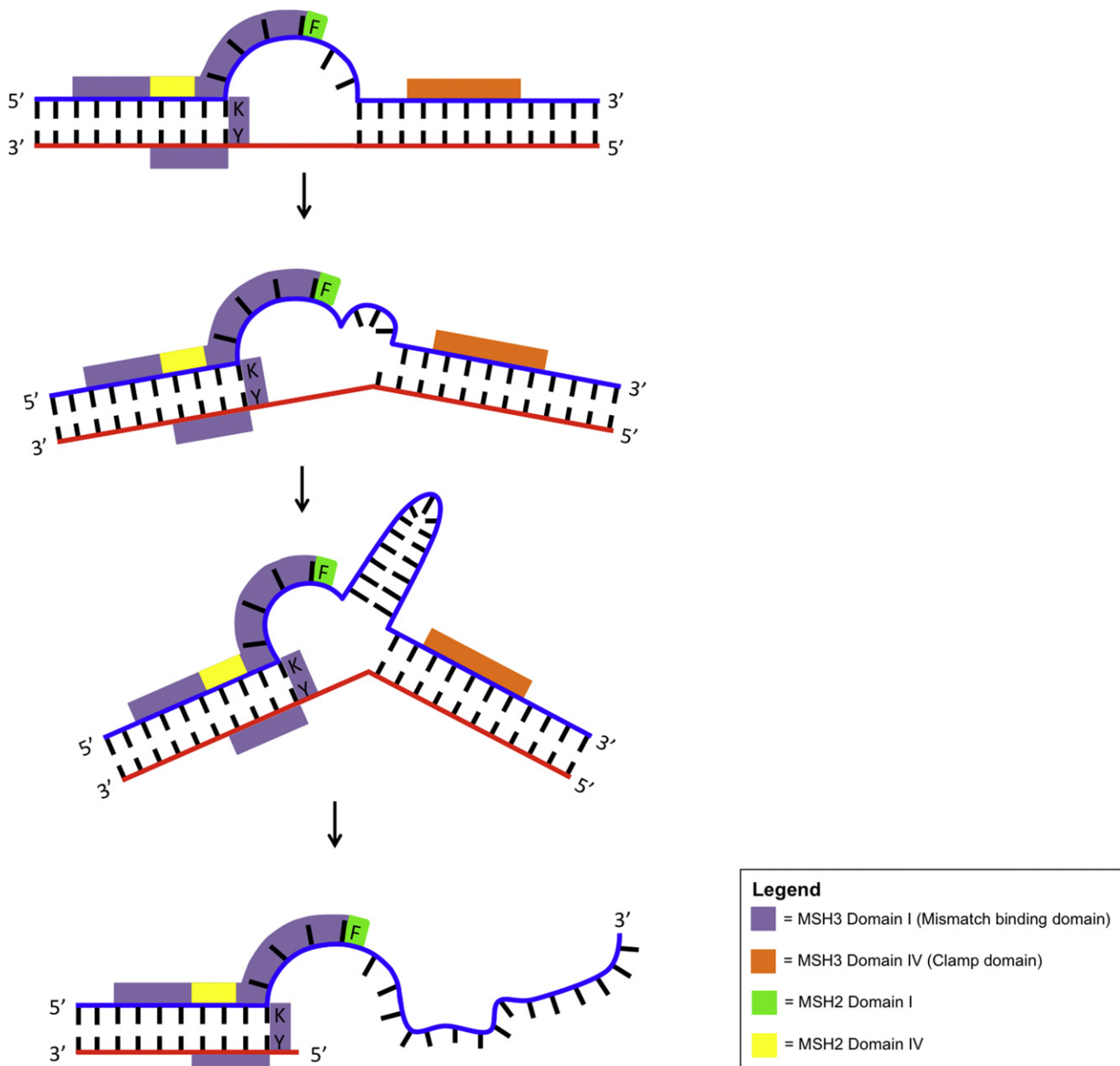
Unraveling how MutS $\beta$  can cause repeat instability can be enhanced by an understanding of its recognition and binding of slipped-DNAs. (For further reading, see articles by Schmidt & Hombauer and Lee *et al.* in this Special Issue.) MutS $\beta$  is involved in at least three distinct processes: repair of insertion-deletion loops, of which (CAG)·(CTG) repeats are a unique set that can be aberrantly repaired to expansions (this review); in yeast, where MutS $\beta$  acts in double-strand break repair by binding 3'DNA-overhangs at the breaks and targeting them for excision, and thirdly, repair by single-strand annealing [66].

MutS $\beta$  has been crystalized in complex with DNAs containing insertion-deletion (CA)<sub>n</sub> loops of 2, 3, 4, and 6 extra nucleotides [67]. The structure reveals insight into MutS $\beta$ 's unique DNA-binding mode, distinct from MutS $\alpha$  and bacterial MutS; how MutS $\beta$  can function in processing both insertion-deletion loops of varying size, as well as of 3'-ends in single-strand annealing. Notably, MutS $\beta$  interacts with both DNA strands 5' of the insertion deletion

loops, but only the loop-containing strand 3' of the loop (Fig. 3). This binding mode suggests that the loop is essentially equivalent to a 3' overhang. The 5' duplex and loop regions are contacted by MSH3 clamp domain (I), followed by stacking of the MSH2 Phenylalanine42 on the 4th base of the loop, while the 3' loop strand in the duplex region is contacted by the MSH3 mismatch-binding domain (IV).

Awareness of the MutS $\beta$ -DNA contacts is likely to be crucial to proper repair or mutation outcome of DNA structures. Binding by MutS $\beta$  is mediated by the MSH3-conserved Tyrosine245-Lysine246 pair, not present in either MutS $\alpha$  or the bacterial MutS protein, implying that this binding mode may define the distinct roles of MutS $\beta$  over MutS $\alpha$ . Tyr245 interacts with nucleotides on the strand opposite the loop at the double stranded/single stranded (ds/ss) DNA junction, while Lys246 interacts with the loop strand on its 5' end. Notably, the critical function of these residues was predicated, as mutations of the homologous residues in the yeast *Msh3* gene, Tyrosine157 and Lysine158, caused a mutator phenotype and microsatellite instability [68,69].

How can MutS $\beta$  bind loops of various sizes? For DNA loops longer than three nucleotides, the Phe42 of MSH2 forms a  $\pi$ -stack with the fourth unpaired base. The direct interaction of domains I of MSH2 and MSH3 are the same for various loop sizes (Fig. 3). Importantly, different loop sizes fit into the MutS $\beta$  binding pocket, whereby the DNA bending angle is tighter for longer loops; in this manner MutS $\beta$  can accommodate longer loops. As predicted [69] and shown by footprinting [70], the MutS $\beta$  complex bends the looped substrate upon binding, where the bending angle increased from 90° for (CA)<sub>1</sub> to 120° for (CA)<sub>3</sub> in the hMSH $\beta$  complex [67]. Based upon the hMutS $\beta$ -(CA)<sub>2</sub>-6 crystal structure and the yMutS $\beta$ -(GT)<sub>4</sub> footprinting from the Alani lab [68,70], it would seem that the extra nucleotides in longer loops or slip-outs beyond the first four nucleotides following the ds/ss junction (demarcated by the MSH2 Phe42), toward the 3' end, would be unbound by MutS $\beta$  and be free and exposed to solvent or nuclease cleavage (MutL $\alpha$ ). Similarly, for



**Fig. 3. Binding of MutSβ components MSH2 and MSH3 to slipped-DNA.** Binding by MutSβ to DNA loops reveals how the complex can bind looped-DNAs of various sizes, and DNAs with 3' overhangs. In the crystal structure of MutSβ, the MSH2 Phenylalanine42 (F), which forms a  $\pi$ -stack with the fourth unpaired base for loops of  $n = 3, 4,$  and 6 extra nucleotides, is separated by precisely 4 nucleotides from the MSH3 Tyrosine245-Lysine246 pair (YK), which stack upon the terminal base-pairs of the duplex just 5' of the loop. See text for details and Gupta et al. [67].

a 3' overhang substrate, the unpaired nucleotides beyond the first 4, following the ds/ss junction would also be free and exposed to solvent or nuclease cleavage (MutL $\alpha$  or XPF-ERCC1, known to interact with MutSβ in ssDNA annealing.) [66].

While the crystal structure of the human MutSβ and footprinting of the yeast MutSβ on (CA)- and (GT)-loops, respectively, present similar binding modes, the same is not true for other substrates. Curiously, footprinting of hMutSβ on a (CA)<sub>4</sub> loop showed complete protection of most of the loop-out and both branch arms [71]. Equally surprising was the almost complete footprint protection of a (CAG)<sub>13</sub> hairpin, with an un-protected 5' loop region [71]. Both of these strikingly different maps suggest that the binding mode of MutSβ may differ for different sequences; even complementary repeat sequences may differ.

### 3.4. Transcription-mediated instability and a role for MMR

Transcription of repeat sequences, which occurs in non-replicating cells that incur expansions, can lead to repeat instability. Bidirectional transcription in a human cell model has been seen to increase instability over unidirectional ([72–76] and citations therein). This instability may arise from RNA:DNA hybrids (R-loops, Fig. 2) formed between the nascent RNA and the template DNA strand. R-loops can form following unidirectional transcription (single-R-loops) or convergent, bidirectional transcription (double-R-loops). Notably, all disease repeat loci studied to date have revealed convergent bidirectional transcription across the repeat tracts. Single R-loops have been observed *in vitro* at the following repeats: CAG, CTG, CGG, CCG, GAA (but not TTC), the GGGGCC, and GGCCCC repeats [50,77]. Double-R-loops have been observed *in vitro* at the (CAG):(CTG) and (CGG):(CCG) [50]. R-

loops have also been detected by anti-RNA:DNA hybrid antibody immunoprecipitations at the FXS and FRDA loci in patient cells [40].

Evidence from bacterial and *in vitro* assays using human cell extracts suggests that aberrant processing of R-loops can lead to repeat instability [73]. The mechanism by which this is perpetuated is not established, however transcription presents an opportunity to form slip-outs, hairpins, quadruplexes and other unusual structures in the displaced non-template strand. Furthermore, transcription leading to R-loop formation allows ssDNA to temporally persist, and upon removal of the R-loop, presents the opportunity for misaligned reannealing of the DNA strands leading to slip-out formation. The resulting unusual DNA structures are targets for aberrant MMR and downstream instability, as discussed above (Fig. 2).

Evidence revealing a direct involvement of MMR proteins in transcription-associated repeat instability is limited, but supported by the finding that knockdown of MSH2 or MSH3, but not MSH6 in human cells decreases CAG and GAA instability following transcription [35,78]. Knockdowns of Cockayne Syndrome protein B (CSB/ERRCC6 - required for transcription-coupled NER) and of MutS $\beta$  reduced CAG contractions [78], while knockdowns of R-loop processing RNases H1 and H2A enhanced contractions [74]. Processing of R-loops may involve formation of slipped-DNAs, whose processing by MMR proteins may lead to instability. The complex interactions between transcription progression, DNA topology, and repair factors in repeat instability, are deserving of future attention.

#### 4. Contractions of expanded repeat sequences

##### 4.1. MMR deficiencies can lead to (CAG)-(CTG) and (CGG)-(CCG) contractions

Contractions of inherited repeat expansions are of clinical interest. Typically, (CAG)-(CTG) repeats tend to expand, but rare families with DM1, HD, spinocerebellar ataxia 1 (SCA1) or spinobulbar muscular atrophy (SBMA) display high levels of contraction of the expanded repeat (reviewed in Ref. [1]). In transgenic (CAG)-(CTG) mice, deficiencies of MSH2 or MSH3, but not MSH6, led to a striking switch from an expansion bias to a contraction bias in transmitted and somatic tissues [79–84]. Not all mouse lines display this switch to a contraction bias, but are stabilized [8]—suggesting a link of MMR with a *cis*-element. MSH6 may protect against paternally-transmitted contractions [11] and its absence in some cell models can lead to increased somatic expansions [18]. Thus, MutS $\alpha$  may be an antimutator of CAG repeats, which in the absence of MSH6, could result in increased available MSH2 for binding to MSH3, forming the expansion-driving MutS $\beta$  complex [18,83]. DNA Ligase I may also be involved in protecting against repeat contractions, as a CTG contraction bias for maternal transmissions was observed for DM1 mice on a defective Ligase I background (46BRLigI<sup>m/m</sup>) [22]. MSH2 also seems to protect against contractions of premutation CGG repeats [30]. Little is known regarding contraction mechanisms of expanded repeat tracts. Understanding the process of repeat contractions is important, since harnessing this process may be clinically beneficial.

##### 4.2. Polymorphisms in DNA repair genes can regulate levels of (CAG)-(CTG) instability

An exciting advance is the recent discovery that polymorphic variants of DNA repair genes can modulate levels of (CAG)-(CTG) repeat expansions. This was first reported in HD mice, where the levels of CAG expansions (transmitted and somatic) depended upon

mouse strain background [12]. Strikingly, the high levels of CAG expansions observed in the C57BL/6J (B6) background, were completely lost when the background was switched to a BALB/cByJ (CBy) background, even though both strains were MMR-proficient. The loss of expansions in the CBy background was as blatant as an MSH2-deficiency on the B6 background. The source of variable CAG instability levels mapped tightly to polymorphic variants of the *Msh3* gene, specifically, MSH3 protein was highly expressed in the B6 but not CBy strain, leading to high levels of CAG expansions or stability, respectively. A genome-wide association study identified a large genomic region harboring the *Mlh1* gene as a possible source of HD mouse strain-specific variations in CAG instability [4]. These polymorphic variations in DNA repair genes may explain the previously observed mouse strain-specific variations of repeat instability [83,85]. In humans, polymorphic variants in DNA repair genes are likely sources of variable instability and disease onset, progression, and severity. Several recent analyses have identified polymorphisms in various DNA repair genes with levels of CAG instability in several diseases [86,87]. The clinical implications that polymorphisms in DNA repair genes may modify levels of CAG instability are far reaching: such polymorphisms may explain the variable levels of disease age-of-onset, disease progression and severity, which may be driven by the highly variable levels of repeat instability between individuals [88]. To this degree, the diagnosis of inheriting a disease-associated repeat expansion, may eventually be prognostically tempered (positively or negatively) with associated DNA repair gene variants.

#### 5. Conclusions & future goals

Mismatch repair proteins MSH2, MSH3, MLH1, MLH3, and to a lesser extent PMS2, are the strongest drivers of (CAG)-(CTG) expansions. Similar, but not identical effects are played by MSH2 and MSH3 for (CGG)-(CCG) and (GAA)-(TTC) repeats. Considering the variety of repeat sequences associated with the greater than forty repeat diseases, it is intriguing to consider whether MMR treats other repeat sequences in a similar manner. The process by which MMR drives repeat expansions likely involves errors during repair of slipped-DNAs or endogenously damaged DNA. While many repair proteins have been assessed for their involvement in repeat instability, gaps in our knowledge remain, such as identifying the initiation of instability, or factors that process large slip-outs, small and large jumps in size, are fodder for future investigations. Identification of specific DNA repair gene polymorphisms that enhance or suppress repeat expansions that may hold prognostic value for affected families. Elucidating the mechanism by which MMR can switch high levels of repeat expansions to high levels of contractions, will prove insightful towards inducing repeat contractions. Enhancement of repeat contraction-inducing processes, possibly by modulation of DNA repair, may present therapeutically beneficial outcomes [2] and is a long-term future goal of understanding the role of DNA repair in disease-associated repeat instability.

#### Conflict of interest

The authors declare that there are no conflicts of interest.

#### Acknowledgments

The Pearson Lab is supported by the National Sciences and Engineering Research Council of Canada (NSERC), the Weston Brain Institute Tribute Communities, the Kazman Family Fund, the Marigold Foundation, and the Muscular Dystrophy Association. MHMS is supported by an Ontario Graduate Scholarships. The



authors would like to thank Dr. Stella Lanni for helpful comments on the manuscript.

## References

- [1] C.E. Pearson, K. Edamura, J.D. Cleary, Repeat instability: mechanisms of dynamic mutations, *Nat. Rev. Genet.* 6 (2005) 729–742, <http://dx.doi.org/10.1038/nrg1689>.
- [2] A. López Castel, J.D. Cleary, C.E. Pearson, Repeat instability as the basis for human diseases and as a potential target for therapy, *Nat. Rev. Mol. Cell Biol.* 11 (2010) 165–170, <http://dx.doi.org/10.1038/nrm2854>.
- [3] M.M. Sleat, G.B. Panigrahi, L.P. Ranum, C.E. Pearson, Mutagenic roles of DNA repair proteins in antibody diversity and disease-associated trinucleotide repeat instability, *DNA Repair* 7 (2008) 1135–1154, <http://dx.doi.org/10.1016/j.dnarep.2008.03.014>.
- [4] R.M. Pinto, E. Dragileva, A. Kirby, A. Lloret, E. Lopez, J. St Claire, et al., Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches, *PLoS Genet.* 9 (2013), <http://dx.doi.org/10.1371/journal.pgen.1003930>.
- [5] C.H. Freudenreich, S.M. Kantrow, V.A. Zakian, Expansion and length-dependent fragility of CTG repeats in yeast, *Science* 279 (1998) 853–856.
- [6] J. Miret, L. Pessoa-Brandão, R. Lahue, Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. U. S. A.* 95 (1998) 12438–12443, <http://dx.doi.org/10.1073/pnas.95.21.12438>.
- [7] S. Jackson, A. Whitworth, J. Greene, R. Libby, A SCA7CAG/CTG repeat expansion is stable in *Drosophila melanogaster* despite modulation of genomic context and gene dosage, *Gene* 347 (2005) 35–41.
- [8] K. Manley, T. Shirley, L. Flaherty, A. Messer, Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice, *Nat. Genet.* 23 (1999) 471–473, <http://dx.doi.org/10.1038/70598>.
- [9] S. Tomé, I. Holt, W. Edelmann, G.E. Morris, A. Munnich, C.E. Pearson, et al., MSH2 ATPase domain mutation affects CTG\* CAG repeat instability in transgenic mice, *PLoS Genet.* 5 (2009) e1000482, <http://dx.doi.org/10.1371/journal.pgen.1000482>.
- [10] C.T. McMurray, Hijacking of the mismatch repair system to cause CAG expansion and cell death in neurodegenerative disease, *DNA Repair* 7 (2008) 1121–1134, <http://dx.doi.org/10.1016/j.dnarep.2008.03.013>.
- [11] E. Dragileva, A. Hendricks, A. Teed, T. Gillis, E.T. Lopez, E.C. Friedberg, et al., Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes, *Neurobiol. Dis.* 33 (2009) 37–47, <http://dx.doi.org/10.1016/j.nbd.2008.09.014>.
- [12] S. Tomé, K. Manley, J.P. Simard, G.W. Clark, M.M. Sleat, M. Swami, et al., MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice, *PLoS Genet.* 9 (2013) e10003280, <http://dx.doi.org/10.1371/journal.pgen.1003280>.
- [13] M. Gomes-Pereira, J.D. Hillely, F. Morales, B. Adam, H.E. James, D.G. Monckton, Disease-associated CAG-CTG triplet repeats expand rapidly in non-dividing mouse cells, but cell cycle arrest is insufficient to drive expansion, *Nucleic Acids Res.* 42 (2014) 7047–7056, <http://dx.doi.org/10.1093/nar/gku285>.
- [14] M. Gomes-Pereira, M. Fortune, L. Ingram, J.P. McAbney, D.G. Monckton, Pms2 is a genetic enhancer of trinucleotide CAG-CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion, *Hum. Mol. Genet.* 13 (2004) 1815–1825, <http://dx.doi.org/10.1093/hmg/ddh186>.
- [15] G.B. Panigrahi, M.M. Sleat, J.P. Simard, C.E. Pearson, Human mismatch repair protein hMutL $\alpha$  is required to repair short slipped-DNAs of trinucleotide repeats, *J. Biol. Chem.* 287 (2012) 41844–41850, <http://dx.doi.org/10.1074/jbc.M112.420398>.
- [16] E. Du, E. Campau, S. Soragni, J.W. Ku, et al., Role of mismatch repair enzymes in GAA-TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells, *J. Biol. Chem.* 287 (2012) 29861–29872, <http://dx.doi.org/10.1074/jbc.M112.391961>.
- [17] J. Du, E. Campau, E. Soragni, C. Jespersen, J.M. Gottesfeld, Length-dependent CTG-CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells, *Hum. Mol. Genet.* 22 (2013) 5276–5287, <http://dx.doi.org/10.1093/hmg/ddt386>.
- [18] R. Nakatani, M. Nakamori, H. Fujimura, H. Mochizuki, M.P. Takahashi, Large expansion of CTG-CAG repeats is exacerbated by MutS $\beta$  in human cells, *Sci. Rep.* 5 (2015) 11020, <http://dx.doi.org/10.1038/srep11020>.
- [19] (a) J. Miret, L. Pessoa-Brandão, Instability of CAG and CTG trinucleotide repeats in *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* 17 (1997) 3382–3387, <http://dx.doi.org/10.1128/MCB.17.6.3382>;  
(b) J.K. Schweitzer, D.M. Livingston, Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast, *Hum. Mol. Genet.* 6 (1997) 349–355;  
(c) A. Kantartzis, G.M. Williams, L. Balakrishnan, R.L. Roberts, J.A. Surtees, R.A. Bambara, Msh2–Msh3 interferes with Okazaki fragment processing to promote trinucleotide repeat expansions, *Cell Rep.* 2 (2012) 216–222, <http://dx.doi.org/10.1016/j.celrep.2012.06.020>.
- [20] J. Cleary, C. Pearson, The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence, *Cytogenet. Genome Res.* 100 (2003) 25–55.
- [21] L. Hubert, Y. Lin, V. Dion, J.H. Wilson, Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1, *Hum. Mol. Genet.* 20 (2011) 4822–4830, <http://dx.doi.org/10.1093/hmg/ddr421>.
- [22] S. Tomé, G.B. Panigrahi, A. López Castel, L. Foirey, D.W. Melton, G. Gourdon, et al., Maternal germline-specific effect of DNA ligase I on CTG/CAG instability, *Hum. Mol. Genet.* 20 (2011) 2131–2143, <http://dx.doi.org/10.1093/hmg/ddr099>.
- [23] A.-V.V. Goula, B.R. Berquist, D.M. Wilson, V.C. Wheeler, Y. Trottier, K. Merienne, Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice, *PLoS Genet.* 5 (2009) e1000749, <http://dx.doi.org/10.1371/journal.pgen.1000749>.
- [24] A.-V.V. Goula, A. Stys, J.P. Chan, Y. Trottier, R. Festenstein, K. Merienne, Transcription elongation and tissue-specific somatic CAG instability, *PLoS Genet.* 8 (2012) e1003051, <http://dx.doi.org/10.1371/journal.pgen.1003051>.
- [25] A.G. Mason, S. Tomé, J.P. Simard, R.T. Libby, T.K. Bammler, R.P. Beyer, et al., Expression levels of DNA replication and repair genes predict regional somatic repeat instability in the brain but are not altered by polyglutamine disease protein expression or age, *Hum. Mol. Genet.* 23 (2014) 1606–1618, <http://dx.doi.org/10.1093/hmg/ddt551>.
- [26] S. Tomé, J.P. Simard, M.M. Sleat, I. Holt, G.E. Morris, K. Wojciechowitz, et al., Tissue-specific mismatch repair protein expression: MSH3 is higher than MSH6 in multiple mouse tissues, *DNA Repair* 12 (2013) 46–52, <http://dx.doi.org/10.1016/j.dnarep.2012.10.006>.
- [27] (a) M. Kovalenko, E. Dragileva, J. St Claire, T. Gillis, J.R. Guide, J. New, et al., Msh2 acts in medium-spiny striatal neurons as an enhancer of CAG instability and mutant huntingtin phenotypes in Huntington's disease knock-in mice, *PLoS One* 7 (2012) e44273, <http://dx.doi.org/10.1371/journal.pone.0044273>;  
(b) F. Li, G. Mao, D. Tong, J. Huang, L. Gu, W. Yang, G.M. Li, The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutS $\alpha$ , *Cell* 153 (2013) 590–600.
- [28] A. Taylor, F. Tassone, P. Dyer, S. Hersch, J. Harris, W. Greenough, et al., Tissue heterogeneity of the *FMRI* mutation in a high-functioning male with fragile X syndrome, *Am. J. Med. Genet.* 84 (1999) 233–239.
- [29] R.A. Lokanga, A. Entezam, D. Kumari, D. Yudkin, M. Qin, C.B. Smith, et al., Somatic expansion in mouse and human carriers of fragile X premutation alleles, *Hum. Mutat.* 34 (2013) 157–166, <http://dx.doi.org/10.1002/humu.22177>.
- [30] R.A. Lokanga, X.-N.N. Zhao, K. Usdin, The mismatch repair protein MSH2 is rate limiting for repeat expansion in a fragile X premutation mouse model, *Hum. Mutat.* 35 (2014) 129–136, <http://dx.doi.org/10.1002/humu.22464>.
- [31] F. Tassone, R. Hagerman, L. Gane, A. Taylor, Strong similarities of the *FMRI* mutation in multiple tissues: postmortem studies of a male with a full mutation and a male carrier of a premutation, *Am. J. Med. Genet.* 84 (1999) 240–244.
- [32] E. Reyniers, J. Martin, P. Cras, E. Van Marck, I. Handig, H. Jorens, et al., Postmortem examination of two fragile X brothers with an *FMRI* full mutation, *Am. J. Med. Genet.* 84 (1999) 245–249.
- [33] X.-N.N. Zhao, D. Kumari, S. Gupta, D. Wu, M. Evanitsky, W. Yang, et al., MutS $\beta$  generates both expansions and contractions in a mouse model of the Fragile X-associated disorders, *Hum. Mol. Genet.* 24 (2015) 7096, <http://dx.doi.org/10.1093/hmg/ddv408>.
- [34] S. Ku, E. Soragni, E. Campau, E.A. Thomas, G. Altun, L.C. Laurent, et al., Friedreich's ataxia induced pluripotent stem cells model intergenerational GAA-TTC triplet repeat instability, *Cell Stem Cell* 7 (2010) 631–637, <http://dx.doi.org/10.1016/j.stem.2010.09.014>.
- [35] A. Halabi, S. Ditch, J. Wang, E. Grabczyk, DNA mismatch repair complex MutS $\beta$  promotes GAA-TTC repeat expansion in human cells, *J. Biol. Chem.* 287 (2012) 29958–29967, <http://dx.doi.org/10.1074/jbc.M112.356758>.
- [36] V. Ezzatizadeh, R.M. Pinto, C. Sandi, M. Sandi, S. Al-Mahdawi, H. Te Riele, et al., The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model, *Neurobiol. Dis.* 46 (2012) 165–171, <http://dx.doi.org/10.1016/j.nbd.2012.01.002>.
- [37] V. Ezzatizadeh, C. Sandi, M. Sandi, S. Anjomani-Virmouni, S. Al-Mahdawi, M.A. Pook, MutL $\alpha$  heterodimers modify the molecular phenotype of Friedreich ataxia, *PLoS One* 9 (2014) e100253, <http://dx.doi.org/10.1371/journal.pone.0100523>.
- [38] R.L. Bourn, I. De Biase, R.M. Pinto, C. Sandi, S. Al-Mahdawi, M.A. Pook, et al., Pms2 suppresses large expansions of the (GAA-TTC) $n$  sequence in neuronal tissues, *PLoS One* 7 (2012) e47085, <http://dx.doi.org/10.1371/journal.pone.0047085>.
- [39] G.S. Chandok, M.P. Patel, S.M. Mirkin, M.M. Krasilnikova, Effects of Friedreich's ataxia GAA repeats on DNA replication in mammalian cells, *Nucleic Acids Res.* 40 (2012) 3964–3974, <http://dx.doi.org/10.1093/nar/gks021>.
- [40] M. Groh, M.M. Lufino, R. Wade-Martins, N. Gromak, R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome, *PLoS Genet.* 10 (2014) e1004318, <http://dx.doi.org/10.1371/journal.pgen.1004318>.
- [41] M.M. Krasilnikova, Complexes between two GAA Repeats within DNA introduced into Cos-1 cells, *Mob. Genet. Elem.* 2 (2012) 267–271, <http://dx.doi.org/10.4161/mge.23194>.
- [42] C. Pearson, Y. Wang, J. Griffith, R. Sinden, Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG) $n$  (CAG) $n$  repeats from the myotonic dystrophy locus, *Nucleic Acids Res.* 26 (1998) 816–823, <http://dx.doi.org/10.1093/nar/26.3.816>.

- [43] C.E. Pearson, M. Tam, Y.-H. Wang, S.E. Montgomery, A.C. Dar, J.D. Cleary, et al., Slipped-strand DNAs formed by long (CAG)<sup>n</sup>(CTG) repeats: slipped-out repeats and slip-out junctions, *Nucleic Acids Res.* 30 (2002) 4534–4547.
- [44] G. Liu, X. Chen, J.J. Bissler, R.R. Sinden, M. Leffak, Replication-dependent instability at (CTG)<sup>n</sup>(CAG) repeat hairpins in human cells, *Nat. Chem. Biol.* 6 (2010) 652–659, <http://dx.doi.org/10.1038/nchembio.416>.
- [45] M.M. Axford, Y.-H.H. Wang, M. Nakamori, M. Zannis-Hadjopoulos, C.A. Thornton, C.E. Pearson, Detection of slipped-DNAs at the trinucleotide repeats of the myotonic dystrophy type I disease locus in patient tissues, *PLoS Genet.* 9 (2013) e1003866, <http://dx.doi.org/10.1371/journal.pgen.1003866>.
- [46] D. Renciuk, J. Kypr, M. Vorlickova, CGG repeats associated with fragile X chromosome form left-handed Z-DNA structure, *Biopolymers* 95 (2011) 174–181, <http://dx.doi.org/10.1002/bip.21555>.
- [47] S. Al-Mahdawi, R.M. Pinto, P. Ruddle, C. Carroll, Z. Webster, M. Pook, GAA repeat instability in Friedreich ataxia YAC transgenic mice, *Genomics* 84 (2004) 301–310, <http://dx.doi.org/10.1016/j.ygeno.2004.04.003>.
- [48] R.M. Clark, I. De Biase, A.P. Malykhina, S. Al-Mahdawi, M. Pook, S.I. Bidichandani, The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model, *Hum. Genet.* 120 (2007) 633–640, <http://dx.doi.org/10.1007/s00439-006-0249-3>.
- [49] R. Batra, K. Charizanis, M.S. Swanson, Partners in crime: bidirectional transcription in unstable microsatellite disease, *Hum. Mol. Genet.* 19 (2010) R77–R82, <http://dx.doi.org/10.1093/hmg/ddq132>.
- [50] K. Reddy, M. Tam, R.P. Bowater, M. Barber, M. Tomlinson, K. Nichol Edamura, et al., Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats, *Nucleic Acids Res.* 39 (2011) 1749–1762, <http://dx.doi.org/10.1093/nar/gkq935>.
- [51] G.B. Panigrahi, M.M. Slean, J.P. Simard, O. Gileadi, C.E. Pearson, Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired, *Proc. Natl. Acad. Sci. U. S. A.* 107 (2010) 12593–12598, <http://dx.doi.org/10.1073/pnas.0909087107>.
- [52] G.B. Panigrahi, R. Lau, S. Montgomery, M.R. Leonard, C.E. Pearson, Slipped (CTG)<sup>n</sup>(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair, *Nat. Struct. Mol. Biol.* 12 (2005) 654–662, <http://dx.doi.org/10.1038/nsmb959>.
- [53] T. Zhang, J. Huang, L. Gu, G.-M.M. Li, In vitro repair of DNA hairpins containing various numbers of CAG/CTG trinucleotide repeats, *DNA Repair* 11 (2012) 201–209, <http://dx.doi.org/10.1016/j.dnarep.2011.10.020>.
- [54] A. Pluciennik, V. Burdett, C. Baitinger, R.R. Iyer, K. Shi, P. Modrich, Extrahelical (CAG)<sup>n</sup>(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLα endonuclease activation, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 12277–12282, <http://dx.doi.org/10.1073/pnas.1311325110>.
- [55] M.M. Slean, K. Reddy, B. Wu, K. Nichol Edamura, M. Kekis, F.H. Nelissen, et al., Interconverting conformations of slipped-DNA junctions formed by trinucleotide repeats affect repair outcome, *52* (2013) 773–785. doi:10.1021/bi301369b>.
- [56] S. Kang, K. Ohshima, A. Jaworski, R. Wells, CTG triplet repeats from the myotonic dystrophy gene are expanded in *Escherichia coli* distal to the replication origin as a single large event, *J. Mol. Biol.* 258 (1996) 543–547, <http://dx.doi.org/10.1006/jmbi.1996.0266>.
- [57] J. Weber, Informativeness of human (dC-dA)<sub>n</sub> (dG-dT)<sub>n</sub> polymorphisms, *Genomics* 7 (1990) 524–530.
- [58] C. Blake, J. Tsao, A. Wu, D. Shibata, Stepwise deletions of polyA sequences in mismatch repair-deficient colorectal cancers, *Am. J. Pathol.* 158 (2001) 1867–1870, [http://dx.doi.org/10.1016/s0002-9440\(10\)64143-0](http://dx.doi.org/10.1016/s0002-9440(10)64143-0).
- [59] Z. Yang, R. Lau, J.L. Marcadier, D. Chitayat, C.E. Pearson, Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells, *Am. J. Hum. Genet.* 73 (2003) 1092–1105, <http://dx.doi.org/10.1086/379523>.
- [60] H. Watanabe, F. Tanaka, M. Doyu, S. Riku, M. Yoshida, Y. Hashizume, et al., Differential somatic CAG repeat instability in variable brain cell lineage in dentatorubral pallidoluysian atrophy (DRPLA): a laser-captured microdissection (LCM)-based analysis, *Hum. Genet.* 107 (2000) 452–457.
- [61] H. Hashida, J. Goto, T. Suzuki, S. Jeong, N. Masuda, T. Ooie, et al., Single cell analysis of CAG repeat in brains of dentatorubral-pallidoluysian atrophy (DRPLA), *J. Neurol. Sci.* 190 (2001) 87–93.
- [62] P.F. Shelbourne, C. Keller-McGandy, W.L. Bi, S.-R.R. Yoon, L. Dubeau, N.J. Veitch, et al., Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain, *Hum. Mol. Genet.* 16 (2007) 1133–1142, <http://dx.doi.org/10.1093/hmg/ddm054>.
- [63] G.M. Williams, J.A. Surtees, MSH3 promotes dynamic behavior of trinucleotide repeat tracts in vivo, *Genetics* 200 (2015) 737–754, <http://dx.doi.org/10.1534/genetics.115.177303>.
- [64] N.L. Chan, J. Guo, T. Zhang, G. Mao, C. Hou, F. Yuan, et al., Coordinated processing of 3' slipped (CAG)<sub>n</sub>(CTG)<sub>n</sub> hairpins by DNA polymerases β and δ preferentially induces repeat expansions, *J. Biol. Chem.* 288 (2013) 15015–15022, <http://dx.doi.org/10.1074/jbc.m113.464370>.
- [65] J.R. Stevens, E.E. Lahue, G.-M.M. Li, R.S. Lahue, Trinucleotide repeat expansions catalyzed by human cell-free extracts, *Cell Res.* 23 (2013) 565–572, <http://dx.doi.org/10.1038/cr.2013.12>.
- [66] A.M. Lyndaker, E. Alani, A tale of tails: insights into the coordination of 3' end processing during homologous recombination, *BioEssays* 31 (2009) 315–321, <http://dx.doi.org/10.1002/bies.200800195>.
- [67] S. Gupta, M. Gellert, W. Yang, Mechanism of mismatch recognition revealed by human MutSβ bound to unpaired DNA loops, *Nat. Struct. Mol. Biol.* 19 (2012) 72–78, <http://dx.doi.org/10.1038/nsmb.2175>.
- [68] S.D. Lee, J.A. Surtees, E. Alani, *Saccharomyces cerevisiae* MSH2-MSH3 and MSH2-MSH6 complexes display distinct requirements for DNA binding domain I in mismatch recognition, *J. Mol. Biol.* 366 (2007) 53–66, <http://dx.doi.org/10.1016/j.jmb.2006.10.099>.
- [69] C.D. Downen, R.D. Putnam, Functional studies and homology modeling of Msh2-Msh3 predict that mispair recognition involves DNA bending and strand separation, *Mol. Cell. Biol.* 30 (2010) 3321–3328, <http://dx.doi.org/10.1128/MCB.01558-09>.
- [70] J.A. Surtees, E. Alani, Mismatch repair factor MSH2-MSH3 binds and alters the conformation of branched DNA structures predicted to form during genetic recombination, *J. Mol. Biol.* 360 (2006) 523–536, <http://dx.doi.org/10.1016/j.jmb.2006.05.032>.
- [71] Z. B.A. Owen, M. Yang, M. Lai, M. Gajek, J.D. Badger Gajek, et al., (CAG)<sub>n</sub>-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition, *Nat. Struct. Mol. Biol.* 12 (2005) 663–670, <http://dx.doi.org/10.1038/nsmb965>.
- [72] M. Nakamori, C.E. Pearson, C.A. Thornton, Bidirectional transcription stimulates expansion and contraction of expanded (CTG)<sup>n</sup>(CAG) repeat, *Hum. Mol. Genet.* 20 (2011) 580–588, <http://dx.doi.org/10.1093/hmg/ddq501>.
- [73] K. Reddy, M.H. Schmidt, J.M. Geist, N.P. Thakkar, G.B. Panigrahi, Y.-H.H. Wang, et al., Processing of double-R-loops in (CAG)<sub>n</sub>(CTG)<sub>n</sub> and C9orf72 (GGGGCC)(GGCCCC) repeats causes instability, *Nucleic Acids Res.* 42 (2014) 10473–10487, <http://dx.doi.org/10.1093/nar/gku658>.
- [74] Y. Lin, J.H. Wilson, Nucleotide excision repair, mismatch repair, and R-loops modulate convergent transcription-induced cell death and repeat instability, *PLoS One* 7 (2012) e46807, <http://dx.doi.org/10.1371/journal.pone.0046807>.
- [75] Y. Lin, J.H. Wilson, Diverse effects of individual mismatch repair components on transcription-induced CAG repeat instability in human cells, *DNA Repair (Amst.)* 8 (2009) 878–885, <http://dx.doi.org/10.1016/j.dnarep.2009.04.024>.
- [76] Y. Lin, L. Hubert, J.H. Wilson, Transcription destabilizes triplet repeats, *Mol. Carcinog.* 48 (2009) 350–361, <http://dx.doi.org/10.1002/mc.20488>.
- [77] A.R. Haeusler, C.J. Donnelly, G. Periz, E.A. Simko, P.G. Shaw, M.-S.S. Kim, et al., C9orf72 nucleotide repeat structures initiate molecular cascades of disease, *Nature* 507 (2014) 195–200, <http://dx.doi.org/10.1038/nature13124>.
- [78] Y. Lin, V. Dion, J.H. Wilson, Transcription promotes contraction of CAG repeat tracts in human cells, *Nat. Struct. Mol. Biol.* 13 (2006) 179–180, <http://dx.doi.org/10.1038/nsmb1042>.
- [79] L. Foirey, L. Dong, C. Savouret, L. Hubert, H. te Riele, C. Junien, et al., Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice, *Hum. Genet.* 119 (2006) 520–526, <http://dx.doi.org/10.1007/s00439-006-0164-7>.
- [80] I. Kovtun, A. Thornhill, C. McMurray, Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations, *Hum. Mol. Genet.* 13 (2004) 3057–3068, <http://dx.doi.org/10.1093/hmg/ddh325>.
- [81] C. Savouret, E. Brisson, J. Essers, R. Kanaar, A. Pastink, H. te Riele, et al., CTG repeat instability and size variation timing in DNA repair-deficient mice, *EMBO J.* 22 (2003) 2264–2273, <http://dx.doi.org/10.1093/emboj/cdg202>.
- [82] C. Savouret, C. Garcia-Cordier, J. Megret, H. te Riele, C. Junien, G. Gourdon, MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice, *Mol. Cell. Biol.* 24 (2004) 629–637.
- [83] W.J. van den Broek, M.R. Nelen, D.G. Wansink, M.M. Coerwinkel, H. te Riele, P.J. Groenen, et al., 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 (2002) 191–198.
- [84] V.C. Wheeler, L.-A.A. Lebel, V. Vrbnac, A. Teed, H. te Riele, M.E. MacDonald, Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum, *Hum. Mol. Genet.* 12 (2003) 273–281.
- [85] A. Lloret, E. Dragileva, A. Teed, J. Espinola, E. Fossale, T. Gillis, et al., Genetic background modifies nuclear mutant huntingtin accumulation and HD CAG repeat instability in Huntington's disease knock-in mice, *Hum. Mol. Genet.* 15 (2006) 2015–2024, <http://dx.doi.org/10.1093/hmg/ddl125>.
- [86] S. Martins, C.E. Pearson, P. Coutinho, S. Provost, A. Amorim, M.-P.P. Dubé, et al., Modifiers of (CAG)<sub>n</sub> instability in Machado-Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes, *Hum. Genet.* 133 (2014) 1311–1318, <http://dx.doi.org/10.1007/s00439-014-1467-8>.
- [87] Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium, Identification of genetic factors that modify clinical onset of Huntington's disease, *Cell* 162 (2015) 516–526, <http://dx.doi.org/10.1016/j.cell.2015.07.003>.
- [88] E.M. Ramos, J. Cerqueira, C. Lemos, J. Pinto-Basto, I. Alonso, J. Sequeiros, Intergenerational instability in Huntington disease: extreme repeat changes among 134 transmissions, *Mov. Disord.* 27 (2012) 583–585, <http://dx.doi.org/10.1002/mds.24065>.