Replication factors fostering the integrity of DNA repeats and the viability of repeat-containing yeast cells

A dissertation submitted by

Chiara Masnovo

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Biology

Tufts University

February 2025

Adviser: Sergei Mirkin

ABSTRACT

In this thesis, we studied how non-catalytic components of the eukaryotic replisome are involved in the maintenance of homopurine-homopyrimidine repeats. Specifically, we focused on the $(GAA)_n$ repeats, which are expanded in the repeat expansion diseases Friedreich's ataxia and SCA27B and the $(GAAA)_n$ repeats, which are expanded in a portion of renal cell carcinomas. In Chapter 1, I provide a comprehensive review of the known mechanisms of $(GAA)_n$ repeat instability, both replication-dependent and replication-independent, as well as other repeat associated phenotypes such as fragility and repeat-induced mutagenesis.

Chapter 2 focuses on the study of the replication protein Mcm10 in the maintenance of repeat stability, and we found that the interaction between Mcm10 and the replicative helicase CMG is essential to stabilize homopurine-homopyrimidine repeats, and deficiencies in Mcm10 cause a synthetic viability defect with long $(GAA)_n$ repeats. We find that in this context, the Rad9 checkpoint promotes cell viability by preserving genomic integrity but comes at the cost of increased repeat expansions, likely mediated by slippery DNA synthesis by Polymerase δ.

In Chapter 3, we studied how a crucial component of any DNA synthesis event, the available deoxyribonucleoside triphosphate pool, regulates expansions and contractions of $(GAA)_n$ repeats using both drug treatments and a genetic approach. We determined that both an overall decrease and increase in the available nucleotide pool promote repeat instability, and specific mutations in the ribonucleotide reductase enzyme can also affect repeat stability.

In Chapter 4, we turn to the *in vitro* study of a different repeat – the $(GAAA)_n$ repeat, whose expansions in the first intron of *UGT2B7* have been associated with clear cell renal cell carcinoma. Using DNA polymerization assays, we determined that the $(GAAA)_n$ repeats cause

an orientation-dependent stall of DNA synthesis in a manner that is consistent with the formation of a triplex DNA structure during polymerization by a B-family polymerase.

As a whole, this thesis broadens our understanding of how replication factors modulate homopurine-homopyrimidine repeat stability and identifies a novel important role for Mcm10 in promoting repeat stability and viability of repeat-containing *S. cerevisiae* cells.

ACKNOWLEDGEMENTS

My advisor, Dr. Sergei Mirkin, has my deepest gratitude. There is no better advisor than someone who so passionately cares about both science and life and how their entwined nature shape our growth and experience as scientists and people. Working in your lab and under your guidance and mentorship has been the best scientific experience of my life.

I would also like to thank my committee members, Dr. Catherine Freudenreich and Dr. Mitch McVey, for their constant feedback and suggestions on my project. I have learned so much from each of you. Thank you also to Dr. Anja Bielinsky for acting on my external thesis committee and as my Mcm10 north star.

I am grateful to all of my past and future mentors, my mentees during my time at Tufts and all lab members, present and past! Having you all as my colleagues has made working in the lab a pleasure. Thank you for all of your help, discussions and support.

I want to thank the entire 2019 Tufts Biology Ph.D. cohort. We went from strangers to hanging out together in the lounge all the time in a matter of days and have supported each other for the past five years. I am very proud of where we all are now.

Having good, kind, and fun friends is important any time in life, but especially during graduate school: Victoria, Nastya, Shoshi, Adam, Kara, Kelly, and more. You all mean so much to me. A special shout out to Tyler for walking to and from work with me most days, ensuring I work reasonable hours and introducing me to the state of Maine and many great book series.

My family is an integral part of anything I do and I am forever grateful that they decided to trust me with every decision I have made and have supported me from across the ocean. My

mother Paola and father Giorgio and sister Maria have always allowed me to set my own path and have inspired me to be kind but determined. I would also like to thank my grandparents Savina and Luigi for always making sure I had a slice of bread with Nutella after I finished my homework when I was a kid, the positive reinforcement clearly worked. My other grandparents, Linda and Carlo, are not with us anymore but always in my heart. To the incredible number of uncles and aunts and cousins I grew up with – being surrounded by so many people with different interests and personalities is both very entertaining and intellectually stimulating, thank you for never making the dinner table a boring place to be at. I also want to thank parents in law Giovanni Battista and Gabriella – your large collection of biology books is a large reason why I am here now.

The last two years have been made immensely more enjoyable and fun by two beautiful gray cats: Lisp and Scheme. Going home to them, holding them after a hard day in the lab or dancing around with them during happy moments is worth all the sneezes.

Finally, I would like to express all of my gratitude to my husband Enrico. It's been almost 14 years of experiences together, and this was definitely one where your support, love, wittiness and kindness were appreciated more than ever. Thank you for believing in me more than anyone else. I am looking forward to seeing what comes next for us.

ABBREVIATIONS

5-FOA 5-fluoroorotic acid

A Adenine

BIR Break-induced replication

CGR Complex Genome Rearrangement

CFU Colony Forming Units

DDC DNA Damage Checkpoint

DMSO Dimethylsulfoxide

dNTP Deoxyribonucleoside Triphosphate

DRC DNA Replication Checkpoint

DSB Double Strand Breaks

DUE DNA Unwinding Element

FPC Fork Protection Complex

FRDA Friedreich's ataxia

hPu/hPy Homopurine/homopyrimidine DNA run

HU Hydroxyurea

LNA Locked Nucleic Acid

MCM Minichromosome Maintenance

MMR Mismatch repair

PCNA Proliferating Cell Nuclear Antigen

RED Repeat Expansion Disease

RFC Replication Factor C

RIM Repeat-induced Mutagenesis

RNR Ribonucleotide Reductase

rRE Recurrent Repeat Expansion

SCA Spinocerebellar Ataxia

ssDNA Single-stranded DNA

TNR Trinucleotide Repeat

TRC Transcription-Replication Collisions

TS Template Switching

U Uracil

YPD Yeast Extract Peptone Dextrose Media

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 Replication dependent and independent mechanisms of GAA repeat instability

Chiara Masnovo[1], Ayesha F. Lobo[1], Sergei M. Mirkin[1*]

[1] Department of Biology, Tufts University, Medford, MA, 02155, U.S.A.

[*]To whom correspondence may be addressed: sergei.mirkin@tufts.edu

Published as: **Masnovo, Chiara**, Ayesha F. Lobo, and Sergei M. Mirkin. "Replication dependent and independent mechanisms of GAA repeat instability." *DNA repair* 118 (2022): 103385.

ABSTRACT

Trinucleotide repeat instability is a driver of human disease. Large expansions of $(GAA)_n$ repeats in the first intron of the FXN gene are the cause Friedreich's ataxia (FRDA), a progressive degenerative disorder which cannot yet be prevented or treated. $(GAA)_n$ repeat instability arises during both replication-dependent processes, such as cell division and intergenerational transmission, as well as in terminally differentiated somatic tissues. Here, we provide a brief historical overview on the discovery of $(GAA)_n$ repeat expansions and their association to FRDA, followed by recent advances in the identification of triplex H-DNA formation and replication fork stalling. The main body of this review focuses on the last decade of progress in understanding the mechanism of $(GAA)_n$ repeat instability during DNA replication and/or DNA repair. We propose that the discovery of additional mechanisms of $(GAA)_n$ repeat instability can be achieved via both comparative approaches to other repeat expansion diseases and genome-wide association studies. Finally, we discuss the advances towards FRDA prevention or amelioration that specifically target $(GAA)_n$ repeat expansion.

INTRODUCTION

DNA microsatellites, 1-to-9 base-pair long tandemly duplicated sequences, comprise up to 3% of the human genome [1–3]. Expansions of a subset of microsatellites are associated with over 50 repeat expansion diseases (REDs), and the number is ever-growing [4]. Disease-associated repeats are characterized by length variability (expansions and contractions) and are known to induce fragility and repeat-induced mutagenesis, resulting in genomic instability (for an extensive review, see [5]). In the first years after their discovery in 1991 [6], most repeat expansion diseases were identified as autosomal dominant disorders associated with expansions of $(CNG)_n$ repeats, resulting in a toxic gain of function at the protein level. $(CNG)_n$ expansion diseases include Huntington's disease (HD), spinocerebellar ataxias (SCAs), myotonic dystrophy (MD) and Fragile X syndrome (FXS).

All expandable $(CNG)_n$ repeats can form imperfect hairpins stabilized by CG base pairs or slipped strand DNA structures that result from the formation of two such hairpins in complementary DNA strands [7]. Thus, strand slippage and hairpin formation during DNA replication was initially proposed to be at the center of trinucleotide repeat (TNR) instability [8]. Therefore, the discovery of the genetic basis of Friedreich's ataxia (FRDA) in 1996 came as a surprise [9]. FRDA is an autosomal recessive disease caused by the expansion of a $(GAA)_n$ repeat, which, unlike $(CGN)_n$ repeats, cannot form a hairpin structure [9]. Overall, $(GAA)_n$ runs are amongst the most expansion-prone trinucleotide repeats in the human genome, the majority of which originated from 3′ poly(A) tracts of various *Alu* elements upon AAA to GAA transition [10,11]. In the case of FRDA, however, the $(GAA)_n$ repeat originated from the $A_n(TAC)A_n$ sequence at the center of the *Alu* Sq element located in the first intron of the *FXN* gene [10].

FRDA is the most common form of hereditary ataxia in humans [12]. $(GAA)_{1-33}$ repeats are in the normal range, $(GAA)_{34-65}$ repeats are pre-mutational, and longer repeats are pathogenic. The length of the $(GAA)_n$ repeat positively correlates with the age of disease onset, its severity and progression [13–17]. $(GAA)_n$ expansions ultimately lead to chromatin changes and *FXN* gene silencing, resulting in a drastic reduction in the levels of the mitochondrial protein frataxin [18–22]. Reduced frataxin levels lead to increased oxidative stress, accumulation of iron species in the mitochondria and subsequent cell death, primarily affecting neuronal tissues [18,23–25]. The loss of gene function upon repeat expansions accounts for the recessive mode of inheritance. FRDA is associated with cerebellar and sensory ataxia, diabetes mellitus, and cardiomyopathies, leading to early death (for a clinical review, see [26]). There is currently no effective cure or treatment for FRDA [27].

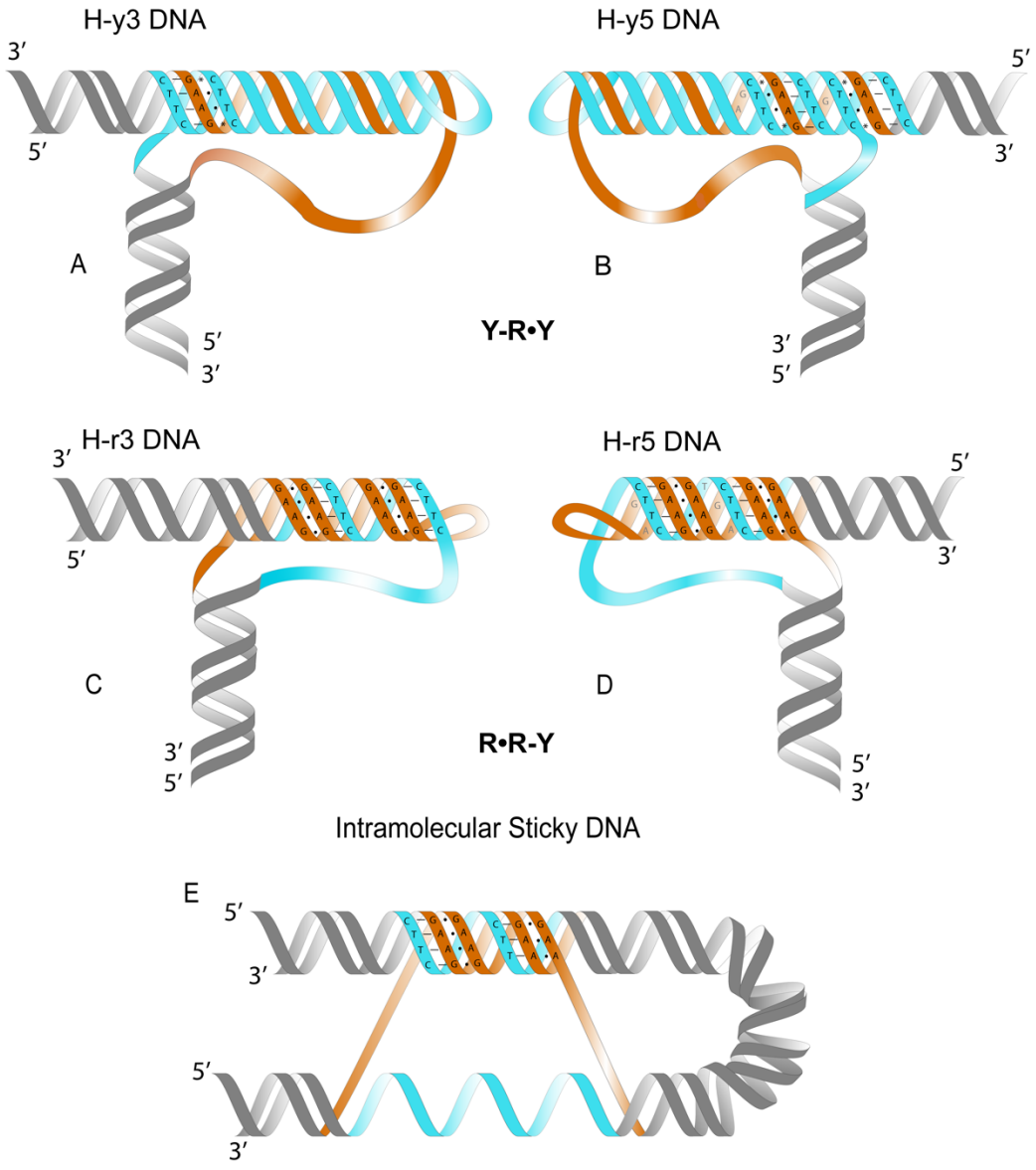

Figure 1-1: Various types of triplex DNA structures formed by long $(GAA)_n$ repeats.

Each H-DNA conformation, YRY or RRY, can exists in two isoforms, depending on whether the 3' or the 5' of a strand is donated to the triplex. Homopurine strands are in orange, homopyrimidine strands are in blue. Dashes indicate Watson-Crick base-pairing, circles indicate Hoogsteen or reverse Hoogsteen base-pairing, asterisks indicate protonated cytosines.

GAA REPEATS FORM TRIPLEX DNA STRUCTURES DURING TRANSCRIPTION AND REPLICATION

Being a homopurine-homopyrimidine (hPu/hPy) mirror repeat, $(GAA)_n$ runs can assume the non-B DNA structure termed DNA triplex or H-DNA. H-DNA is formed when a DNA strand corresponding to one half of the repeat folds back forming a triplex with the duplex half of the repeat via Hoogsteen base pairing, while its complementary strand remains single-stranded **(Figure 1-1 a-d)** [28,29]. Like other alternative DNA structures, H-DNA is thermodynamically unfavorable in linear double-stranded DNA, but becomes favorable in negatively supercoiled DNA, which is topologically equivalent to unwound DNA, as it relieves torsional stress [30,31]. Consequently, H-DNA is not a steady-state presence in genomic DNA but can rather form at different stages of the cell-cycle during specific genetic processes such as replication and transcription, hence it is a dynamic DNA structure.

H-DNA can exist in several conformations [32]. In an H-y triplex (YRY), the p**y**rimidine strand contributes to triplex formation by Hoogsteen base pairing **(Figure 1-1 a-b)**. Normally, H-y formation is favored at low pH since it requires cytosine protonation, but because of their high AT-content, $(GAA)_n$ repeats can readily form H-y triplexes under physiological conditions [31,33]. In an H-r triplex (RRY), the pu**r**ine strand contributes to triplex formation through reverse Hoogsteen base pairing **(Figure 1-1 c-d)**. H-r triplexes, including those formed by the $(GAA)_n$ repeat, were observed at neutral pH in the presence of divalent cations [34,35]. Finally, two distant $(GAA)_n$ repeats within the same supercoiled DNA molecule can form a structure termed sticky DNA **(Figure 1-1 e)** [34,36]. In this case, a homopurine strand from one of the repeats forms an H-r triplex with another repeat, also in the presence of divalent cations [34,37].

While all these structures can be formed under specific conditions *in vitro*, it remains unclear which one is the most common at the FRDA locus and how each contributes to repeat instability and disease. It is generally challenging to detect dynamic DNA structures at endogenous genomic loci *in vivo*, given that they may only be formed transiently. It is particularly difficult in mammalian cells, in which chemical probing has proven to be highly cumbersome, partially due to the extreme genome size (reviewed in [38]). Triplex-specific antibodies were shown to bind *in situ* to multiple sites in human chromosomes, some of which contained $(GAA)_n$ repeats [39–41]. Note however, that the resolution power of this technique is not at the nucleotide level.

As discussed above, negative DNA supercoiling is the main driver for H-DNA formation *in vitro*. In mammalian nuclei, transcription is the main source of negative DNA supercoiling ([42] and references therein). Kouzine et al. used permanganate treatment, which oxidizes thymine residues in single-stranded DNA (ssDNA), in combination with S1 nuclease cleavage and high-throughput sequencing to detect non-B DNA structures in the genome of mouse B-cells [43]. This study identified approximately 17,000 sites of H-DNA out of ~728,000 predicted H-DNA motifs, the prevalence of which positively correlated with transcription levels **(Figure 1-2 a)** [43]. More recently, an S1-seq based method, which relies on mapping of S1-cleavage sites in permeabilized cells [44], identified about 144,000 H-DNA forming structures, preferably H-y5 triplexes **(Figure 1-2 b)** [45]. Notably, most H-DNA motifs were observed at relatively short homopurine-homopyrimidine repeats, and the authors believe that they could have formed *ex vivo* during sample preparation owing to the acidic pH of the S1-nuclease reaction buffer. In FRDA patient cells, a similar S1-END-seq approach revealed H-DNA at the expanded $(GAA)_n$

repeats specifically when the *FXN* locus was transcribed[46] **(Figure 1-2 c)**. It is concluded, therefore, that in this case, the triplex is formed *in vivo* during transcription of the $(GAA)_n$ repeat.

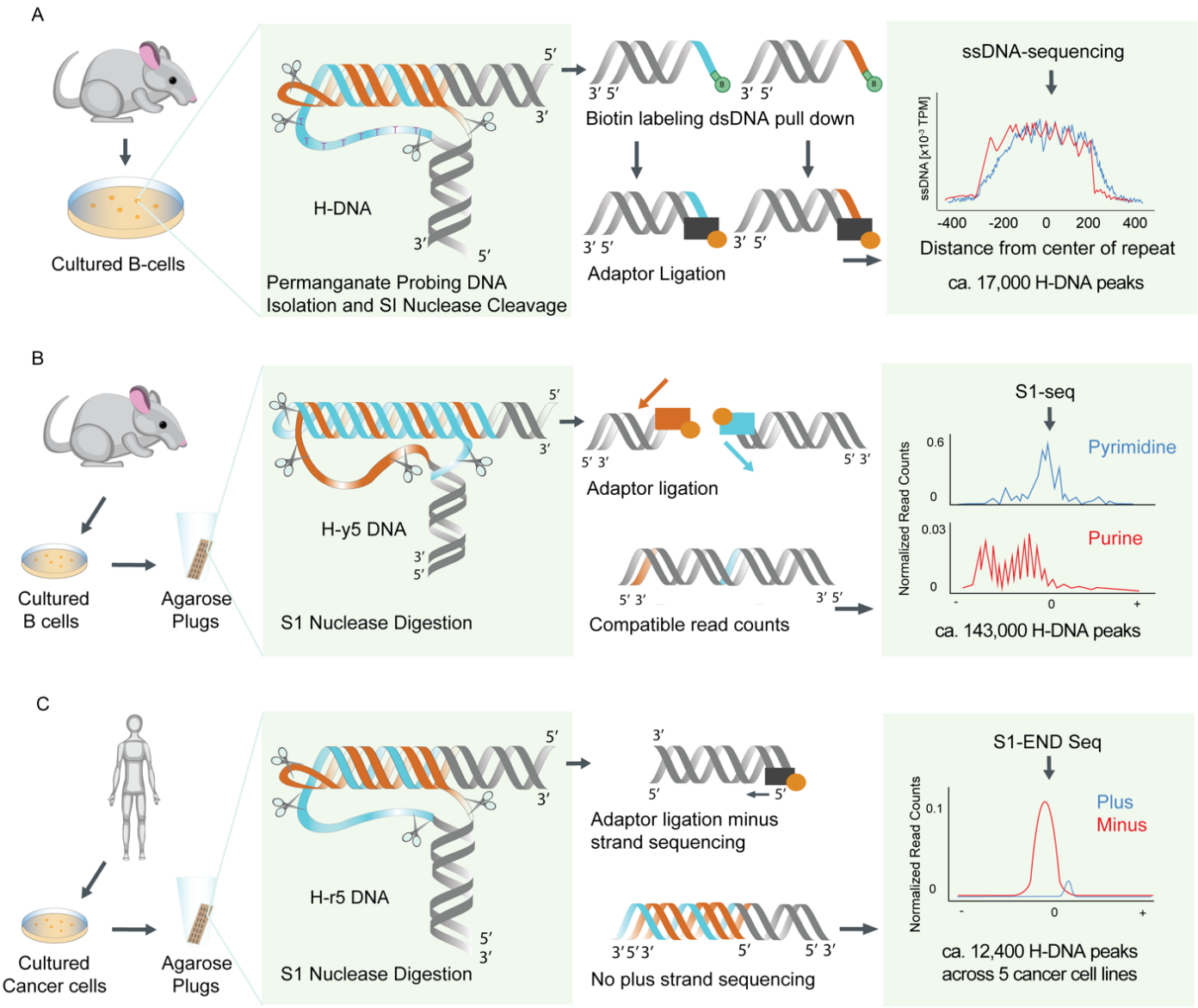

Figure 1-2: Methods to detect in vivo formation of H-DNA in higher eukaryotes.

(a) Genome-wide H-DNA formation in cultured mouse B-cells (adapted from [43]). **(b)** Genome-wide H-DNA formation in primary mouse cells via S1-seq (adapted from [45]). **(c)** Genome-wide H-DNA formation in human cancer cell lines via S1-END Seq (adapted from [46]).

Besides transcriptional supercoiling, formation of H-DNA can be promoted by DNA strand unwinding during DNA replication. During DNA polymerization *in vitro*, a DNA strand from a partially unwound repeat can fold into a triplex structure, effectively blocking further DNA polymerase progression. Consequently, triplex forming motifs were called "suicidal sequences" for DNA polymerization [47]. A similar mechanism accounts for the blockage of DNA polymerization by expanded $(GAA)_n$ repeats [31,48]. Recently, analysis of DNA synthesis through a reconstituted eukaryotic replication fork revealed a weak but reproducible stalling only when a long $(GAA)_n$ repeat was located on the leading strand template [49]. Long $(GAA)_n$ repeats were also shown to stall replication *in vivo* both in yeast and mammalian cells [50–54]. In yeast, $(GAA)_n$-mediated fork stalling is orientation-dependent, as it was only observed when the $(GAA)_n$ run was on the lagging strand template [53,55]. This was interpreted as the formation of an H-r triplex in front of the fork, followed by strand unwinding by the replicative CMG (Cdc45/Mcm2-7/GINS) helicase [53]. In human cells, $(GAA)_n$ repeats cause fork stalling in SV40-based episomes in an orientation-independent manner [50,51,56]. This inconsistency between the two systems could be explained by the fact that in an SV40 replisome, the T-antigen helicase is used instead of the endogenous CMG helicase, and synthesis is performed by Pol δ on both the leading and the lagging strand [57]. Fork stalling at $(GAA)_n$ repeats result in fork reversal [51,56],which can lead to subsequent fragility and instability [55,56,58].

It is important, therefore, to understand whether triplexes can cause the stalling of regular replication forks at endogenous chromosomal locations. The use of the S1-END-seq approach described above [44] identified two types of non-B DNA structures in human cells: DNA cruciforms formed by $(AT)_n$ repeats and H-DNA formed by long hPu/hPy mirror repeats[46]. H-DNA was detected in multiple cancer cell types, and a considerable portion of sequences was

shared between genomes, indicating the presence of highly conserved H-DNA regions **(Figure 1-2 c)**. Most of the detected triplexes were formed by $(GAAA)_n$, $(GAA)_n$ and $(GGAA)_n$ repeats and corresponded to the H-r5 conformation. To address the concern that the acidic pH of S1-nuclease treatment could promote triplex formation ex vivo, the authors replaced S1-nuclease with P1-nuclease, which cleaves ssDNA at neutral pH. This new approach, called P1-END-seq, revealed a similar number and distribution of triplex peaks in cancer cell lines. Another important argument supporting H-DNA formation *in vivo* is that the triplex peaks spike in S-phase, which correlates with orientation-dependent replication fork stalling, similar to that observed in yeast.

Replication fork stalling at expanded $(GAA)_n$ repeats at the *FXN* locus was directly shown using single-molecule analysis of replicated DNA (SMARD) isolated from FRDA patient cells [52]. Stalling was observed in both orientations, being particularly pronounced when the $(TTC)_n$ run is on the lagging strand template. This polarity is opposite to that observed in yeast and in human cells via S1-END-seq. While the reason for this difference remains unclear, one possibility is that head-on collision of *FXN* transcription with replication going in the opposite direction adds to the strength of the stall. Importantly, GAA-specific polyamides that disrupt triplex DNA rescue replication fork stalling, implying that H-DNA causes fork stalling at the repeat [46,52].

Further characterization of non-B DNA structures at the single nucleotide level has the potential to provide insight into not only location and frequency of H-DNA, but also requirements for its formation, mutagenic potential and association with cell or tissue type. This information will be important to determine whether H-DNA is associated with genetic diseases such as cancer besides FRDA, as proposed by recent computational analyses [59,60]. So far,

individual sequencing methods show detection biases toward specific H-DNA isomers, and different types of sample processing might introduce artificial H-DNA formation to a certain extent. Newer methods based on the use of small molecules as modifiers are being developed to allow more sensitive and accurate detection of alternative DNA structures [61,62]. Where, then, are the commonalities? Firstly, all studies agree that even though both replication and transcription can promote H-DNA formation, the major contributor varies based on the specific repeat and its genetic environment, such as chromatin context and specific location. In addition, interruptions in the repeat decrease its ability to form H-DNA and stall the replication fork [45,46], in accordance with previous *in vitro* studies for long interrupted $(GAA)_n$ repeats [36]. Importantly, interrupted $(GAA)_n$ repeats are rarely found in FRDA patients, and when present lead to delayed disease onset and milder phenotypes, strongly suggesting that the ability of the repeat to form a triplex is essential for disease pathogenesis [13,63]. Finally, the prevalence of H-DNA forming sequence warrants extensive studies of their potential biological functions, which likely remain widely underestimated.

GENOME INSTABILITY MEDIATED BY $(GAA)_N$ REPEATS

As is true for other repeat expansion diseases, the longer the $(GAA)_n$ repeat is, the more prone to length instability it becomes [64,65]. Instability is observed both during intergenerational transmission and in post-mitotic somatic cells [66–70]. Expansions predominantly happen during intergenerational transmission and cell division. In somatic cells, contractions are the most prevalent form of $(GAA)_n$ instability, although expansions were observed in affected tissues, including the heart, pancreas and neuronal tissues [66,71]. Somatic mosaicism – the presence of a

variety of $(GAA)_n$ lengths in the same patient – was observed in multiple patient tissues in an age-dependent manner, with the length of the largest allele determining the scale of the observed mosaicism [66,71–74].

Since $(GAA)_n$ repeat expansions cause a human hereditary disease, research has focused on the study of their expansion mechanisms. Somewhat less appreciated is the fact that $(GAA)_n$ repeats can cause additional local and global genome rearrangements. The original examples came from studies in yeast. First, it was directly demonstrated that expanded $(GAA)_n$ repeats are fragile, resulting in double-strand breaks [55]. Second, $(GAA)_n$ repeats appear to cause mutagenesis at a distance, in a process that we called repeat-induced mutagenesis (RIM) [75]. Expanded $(GAA)_n$ repeats at the *FXN* locus also increase mutagenesis in the area surrounding the repeat, likely through double-strand break (DSB) repair processes [75–79]. Note that other triplex-forming sequences were also associated with increased break-induced mutagenesis in mammalian cells [77,80,81].

Altogether, various types of $(GAA)_n$ repeat-mediated instability contribute to the accumulation of mutations at and around the expanded locus, as well as rearrangements in other genomic regions. These events can modulate the age and onset of FRDA and/or lead to the emergence of other pathogenic mutations, which can in turn modulate $(GAA)_n$ length stability. We will cover the mechanisms of $(GAA)_n$ mediated genome instability during replication and in non-dividing cells in the next sections.

Fragility

As is common for other disease-related repeats, $(GAA)_n$ repeats were shown to be fragile in multiple model systems, causing DSBs and genome rearrangements. In yeast, $(GAA)_n$ fragility

was shown to be dependent on its orientation relative to the replication origin, being higher when the $(GAA)_n$ is on the lagging-strand template – the same orientation that causes replication stalling [55]. In this case, fragility was dependent on the mismatch repair (MMR) machinery. We hypothesize that yeast MMR cleaves the single-stranded loops of H-DNA, erroneously perceiving them as mismatched loop-outs. It would be of great interest to substantiate this idea in biochemical studies. Indirect evidence also indicates that there is increased fragility at the endogenous $(GAA)_n$ repeat in the *FXN* locus in FRDA patient cells [58].

Expansions

Expanded $(GAA)_n$ repeats were shown to affect replication fork progression in every experimental system studied to date, likely owing to their triplex-forming potential. In yeast, $(GAA)_n$ repeats start expanding at the carrier length of $(GAA)_{52}$, and the rate of expansion increases exponentially with the repeat's length [65]. This begs the question, is there a link between replication through the repeat and its instability? An unambiguous affirmative answer came from a yeast experimental system. A genome-wide screen identified genes that modulate $(GAA)_n$ instability, including its fragility and propensity for expansions [82]. The screen had hits in three main categories: replication-associated genes, transcription initiation genes, and two components of the CST (Cdc13-Stn1-Ten1) complex, which regulates telomere maintenance. Further studies expanded on the role of each category in $(GAA)_n$ repeat instability.

First, an intact and processive core replisome was shown to counteract instability [82,83], as mutations in Pol ε and δ, as well as in the fork stabilization complex (Tof1-Csm3-Mrc1), promote large-scale $(GAA)_n$ repeat expansions [83]. Mutations in subunits of the CMG helicase also increase mid-scale expansions of short $(GAA)_{25}$ repeats, through a mechanism consistent with template switching (TS) and break-induced replication (BIR) [84,85]. During lagging strand

synthesis, each Okazaki fragment needs to be processed by 5' flap endonucleases prior to ligation. The major flap endonucleases in yeast are Rad27 and Dna2, which cleave short and long flaps, respectively. Mutation of either flap endonuclease dramatically increases $(GAA)_n$ repeat expansions, possibly due to an imbalance in the total amount of ssDNA in the cells which can result in increased formation of triplexes on the flaps, ultimately resulting in $(GAA)_n$ repeat instability **(Figure 1-3 a)** [65,83].

A strong replication stall could lead to dramatic consequences such as template switching or replication fork reversal [51]. The goal of both of these processes is to bypass a template "lesion", but can result in $(GAA)_n$ repeat instability [86]. Nevertheless, whereas fork stalling is orientation-dependent [52,53,65,69], large-scale $(GAA)_n$ repeat instability seems largely orientation-independent, even though there is a slight bias for increased instability when the $(GAA)_n$ run is on the lagging strand template in all of the studied systems [64,65]. This bias has been related to the asymmetrical nature of DNA replication and led to the proposal of the "ori-switch" model, in which origin activity and orientation relative to the position of the repeat influences predisposition for either expansions or contractions [87]. What other processes, in addition to fork stalling, are then causing instability during replication? Template-switching during DNA synthesis could also occur independently of stalling, especially at the site of long repetitive templates in which TS could happen at a higher rate **(Figure 1-3 a)**.

In addition, factors other than the ones involved directly in the replication fork can contribute to $(GAA)_n$ expansions during replication, as illustrated by the identification of the CST complex in the genetic screen described above [82]. Recently, mutations in the CST complex were shown to lead to large-scale $(GAA)_n$ repeat expansions through a mechanism involving the Rad9-dependent G2/M DNA damage checkpoint activation and post-replicative repair (PPR) of

gaps and nicks [88]. Even though $(GAA)_n$ repeats, unlike expanded $(CAG)_n$ repeats, do not activate the checkpoint by themselves, the checkpoint response likely becomes relevant in the context of other sources of replicative stress [89].

An episomal experimental system was previously established in our laboratory to study the link between replication and expansions of $(GAA)_{100}$ repeats in human cells[56]. Using siRNA to deplete specific proteins, this study shows that large-scale repeat expansions are promoted by proteins involved in replication fork reversal (SHPRH/SMARCAL1/HLTF) and counteracted by proteins that are involved in the restoration of reversed forks (RAD52/RECQL1/WRN). Further, the DDX11 helicase, which was shown to untangle triplex DNA *in vitro* [90], also counteracts repeat expansions. Altogether, this data indicates that expansions occur while the replication fork attempts to bypass a triplex formed by the $(GAA)_n$ repeat **(Figure 1-3 c)**.

Contractions

Repeat contractions are another type of $(GAA)_n$ repeat instability which has been intimately tied with ongoing DNA replication. $(GAA)_n$ repeats become more prone to contractions the longer they are, and contractions occur progressively in somatic cells of FRDA patients, as well as during intergenerational transmission [66,67,70]. The question of whether contractions and expansions occur through a shared mechanism remained a mystery for a long time. A recent study in yeast set up a genetic assay to study large-scale contractions (> 20 repeats) of long $(GAA)_{124}$ tracts [64]. First, contractions were associated with the ability of the repeat to form H-DNA, directly tying contraction events to triplex formation during DNA synthesis. Second, mutations of lagging strand synthesis polymerases (Pol α and Pol δ) and flap-processing nucleases (Rad27 and Dna2), which result in the accumulation of long ssDNA tracts, promoted large contractions in an orientation-independent manner. Third, the ssDNA-binding replication

protein A (RPA) strongly counteracted triplex formation between the nascent lagging strand and its template, preventing contractions **(Figure 1-3 a)** [64]. At the same time, all major DSB-repair pathways did not influence $(GAA)_n$ repeat contractions. This led us to conclude that large-scale contractions occur during triplex bypass during lagging strand synthesis.

Importantly, the average contraction size of the starting $(GAA)_{124}$ was of ca. 60 repeats in this system, and likely to be the result of a one-step process. This corresponds to a contraction of a $(GAA)_{124}$ repeat in the mutational range down to pre-mutational or normal repeat sizes, which is very encouraging from a potential therapeutic standpoint. The challenge remains to identify proteins which exclusively promote contractions or prevent expansions, without affecting the other side of instability or have genome-wide mutagenic effects.

Altogether, the expansion and contraction data point to a link between replication through $(GAA)_n$ repeats and their instability. Both fork stalling at the repeat and instability become apparent at $(GAA)_n$ repeat lengths corresponding to carrier sizes in patients and become more pronounced at disease length. That being said, in the experimental systems studied so far, there is no obvious direct correlation between the strength of fork stalling and repeat instability causing disease. This ambiguity warrants further studies of the mechanisms of replication-dependent repeat instability.

Complex genome rearrangements

Two classes of repeat-mediated genome rearrangements were revealed in a yeast experimental system. First, repair of DSBs within expanded $(GAA)_n$ repeats occasionally leads to the formation of large deletions which encompass the repeat and its adjacent regions [65]. Second, genetic assays combined with Nanopore sequencing found that $(GAA)_n$ repeats cause complex

genome rearrangements (CGRs) of the yeast genome that are yet another byproduct of DSB repair [91]. These CGRs resulted from a mixture of reciprocal and non-reciprocal gene-conversion events, which involve both intra- and inter-chromosomal interactions [91].

We also want to emphasize that with the advent of long-read sequencing and dedicated computational tools, the sequencing of expanded repeats appears more practical, and provides the opportunity to widely survey both the general and affected populations to assemble a comprehensive picture of $(GAA)_n$ repeat sizes and distributions [91–98]. Recently, Nanopore sequencing was used to detect replication fork stalling associated with structure formation during sequencing [99,100]. The combination of long-read sequencing with GWAS, such as the ones conducted on $(CAG)_n$ repeats and Huntington's disease [101–103], could provide invaluable information on genetic modifiers of $(GAA)_n$ repeat stability and FRDA onset and severity, revealing new mechanisms of instability and guiding the development of novel therapeutic avenues.

Repeat-induced mutagenesis

As mentioned above, expanded $(GAA)_n$ repeats in the *FXN* locus increase mutagenesis in surrounding genomic regions in FRDA patients. The mechanisms of this mutagenesis were thus far studied only in a yeast experimental system. Repeat-induced mutagenesis (RIM) was detected up to 10 kb upstream and downstream of long $(GAA)_n$ repeats [79]. Conditional mutations in Pol Ɛ dramatically elevated the rate of RIM, implicating DNA replication in the process [83]. It is not yet clear what causes the mutagenesis. In two studies, RIM depended on translesion synthesis (TLS) by DNA Polymerase ζ [79,104]. In another study, translesion synthesis was only involved in RIM if Pol δ activity was compromised [83]. Two mechanisms are being considered: repair of post-replication gaps that involves Pol ζ [79,104], and a BIR-like pathway [83].

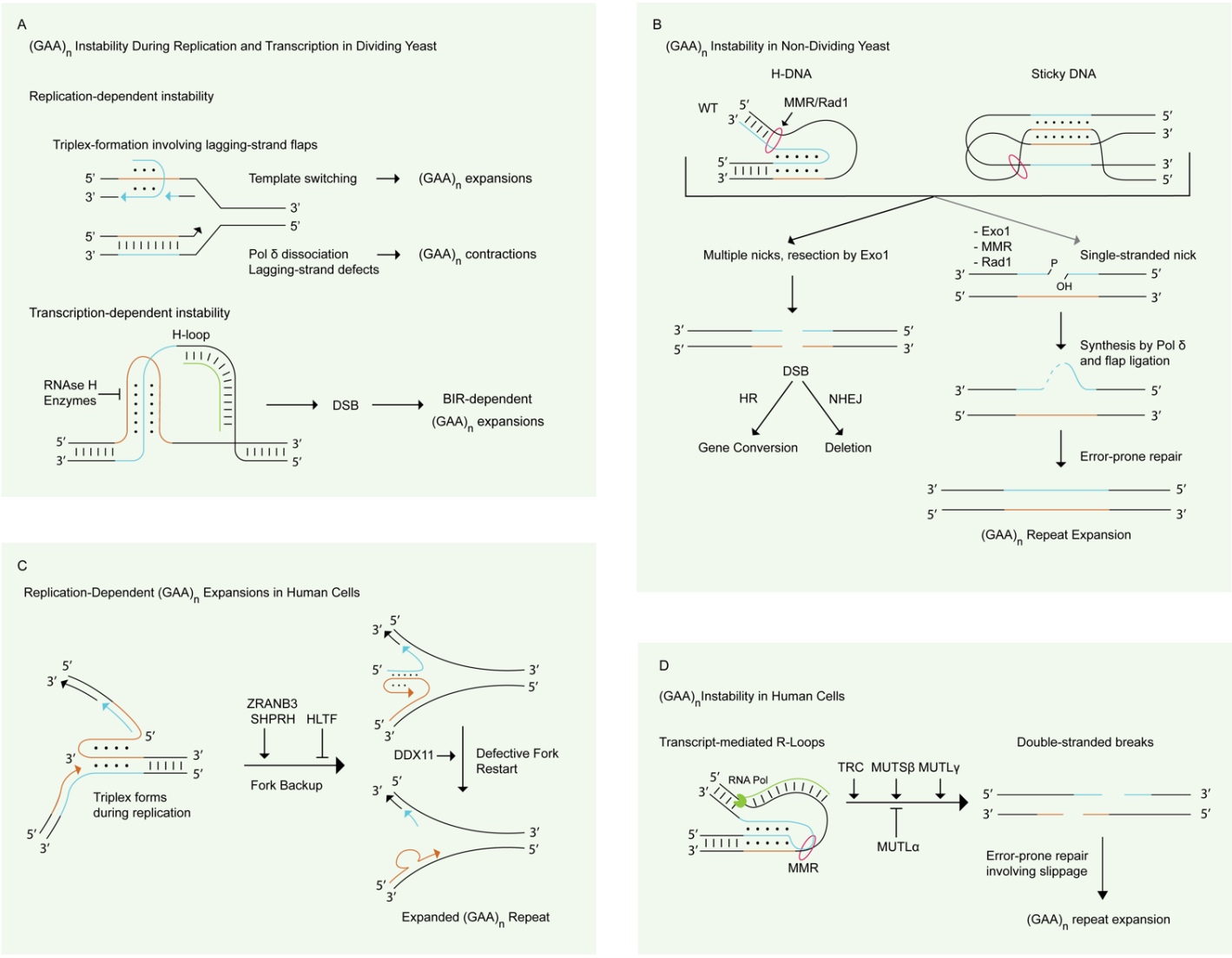

Figure 1-3: Models of $(GAA)_n$ repeat instability.

(A) $(GAA)_n$ repeat instability in dividing yeast mainly occurs via replication-dependent mechanisms, such as template switching (TS) or flap ligation, leading to expansions (see 4.2) [65,83]. Pol δ dissociation during lagging strand synthesis causes contractions (see 4.3) [64]. During transcription, formation of a triplex-stabilizing R-loop (H-loop) can trigger break-induced recombination leading to repeat expansions [105]. (B) $(GAA)_n$ repeat instability in non-dividing yeast is characterized by two different types of events, depending on whether the MMR machinery is functional. MMR drives incisions in H-DNA and sticky DNA structures, which are then converted into DSBs. DSB repair by HR leads to gene conversions events, while repair by NHEJ results in deletions. In MMR-deficient strains, nick repair leads to expansions. (C) Replication dependent $(GAA)_n$ repeat expansions in human cells is initiated by triplex formation ahead of the fork leading to its regression. Repeats can expand upon strand slippage during the restoration of the regressed

fork [56]. (D) $(GAA)_n$ repeat instability in non-dividing and somatic human cells is promoted by the formation of R-loops during transcription and/or upon transcription-replication collisions (TRCs), which are then recognized by MMR and converted into DSBs. Other sources of DSBs can initiate this process as well. Error-prone repair results in $(GAA)_n$ repeat expansions [52,106–108].

ROLE OF TRANSCRIPTION AND R-LOOPS IN $(GAA)_N$ REPEAT INSTABILITY

Expanded $(GAA)_n$ repeats pose an obstacle to transcription in an orientation-dependent manner [54,109,110]. During transcription, a sense $r(GAA)_n$ strand or an antisense $r(UUC)_n$ strand can participate in the formation of stable DNA:RNA duplexes, or R-loops, consisting of two DNA and one RNA strands [111,112]. Interestingly, recent analyses have shown that R-loop formation might be promoted at repetitive elements across eukaryotic genomes [113,114]. Enrichment of unscheduled R-loops can result in genomic instability and has been shown to modulate stability of other TNRs (reviewed in [115,116]). R-loop formation during transcription of expanded $(GAA)_n$ repeats has been proposed as a pathogenic mechanism contributing to transcriptional silencing at the *FXN* locus [117]. The formation of DNA:RNA hybrids at expanded $(GAA)_n$ repeats is thermodynamically favorable since the sense RNA is homopurine [118,119]. Interestingly, transcription has been shown to increase $(GAA)_n$ repeat instability in both human and yeast experimental systems [82,120–122]. In yeast, the transcription-dependent increase in instability was further exacerbated in the absence of the RNase H enzymes, which counteract the accumulation of R-loops [105]. Furthermore, this increased instability was caused by BIR. Altogether, these data point to the role of R-loops and transcription-replication collisions (TRC) in $(GAA)_n$ repeat instability **(Figure 1-3 a)**.

It is generally believed that R-loops can cause genome instability when replication and transcription collide head-on. It was surprising, therefore, that R-loop-dependent instability of

$(GAA)_n$ repeats did not depend on relative orientation of replication and transcription [105]. To explain this difference, it was hypothesized that H-DNA transiently formed upstream of elongating RNA polymerase interacts with the repetitive RNA transcript forming the so-called H-loop [5]. This H-loop is much more stable than either H-DNA or R-loop alone. Therefore, DNA replication could be blocked notwithstanding its directionality. It is yet to be determined whether transcription can promote $(GAA)_n$ repeat instability independently of DNA replication. The answer to this question will allow us to determine whether transcription-mediated instability might be a universally shared mechanism in both dividing and non-dividing cells. It will be important to study whether other factors which contribute to R-loop balance and/or transcription-coupled repair (TCR) affect $(GAA)_n$ repeat stability. One promising candidate is the Senataxin helicase (Sen1 in *S. cerevisiae*), which associates with the replication fork to resolve R-loops and prevents their accumulation during DNA replication and DNA damage repair [123–127].

REPLICATION-INDEPENDENT PATHWAYS OF $(GAA)_N$ REPEAT INSTABILITY

The tissues which are most severely affected in repeat expansion diseases are usually terminally differentiated and do not undergo cellular division. In such tissues, DNA replication-dependent processes cannot be a major source of repeat instability. Nevertheless, $(GAA)_n$ repeats progressively expand in neuronal tissues, such as the cerebellum and dorsal root ganglia (DRG), as well as in cardiac muscles, in both human and mouse models [66,128–132]. A longitudinal study in FRDA patients confirmed lifetime-long addition of $(GAA)_n$ repeats in multiple non-dividing tissues and shows that longer starting repeat sizes lead to a greater magnitude in expansions over time [71]. DNA damage and subsequent repair, involving tracts of DNA synthesis through the

repetitive tract, occur in post-mitotic tissues such as neuronal tissues at every stage of the cell cycle [133,134]. This is relevant for $(GAA)_n$ repeat instability since DNA damage repair can be affected by the presence of the triplex H-DNA structure, which may be perceived as a lesion by DNA repair machineries. It is also foreseeable that repeat instability could arise during cell-cycle reactivation in postmitotic neurons, which occasionally involves aberrant DNA synthesis [135–138]. In the next section, we focus on the DNA repair mechanism that has received most of the recent attention regarding $(GAA)_n$ repeat instability – mismatch repair.

Mismatch repair in $(GAA)_n$ repeat instability

Mismatch repair (MMR) is responsible for the correction of base mismatches and small insertions and deletions (indels), which can arise during DNA replication, recombination, and repair. In MMR, mismatches and small loops (1-3 bp) are recognized by MutSα (Msh2-Msh6), and larger loop-outs by MutSβ (Msh2-Msh3). Subsequently, the MutL complexes, MutLα (Mlh1-Pms2) and MutLγ (Mlh1-Mlh3), excise the mismatch, followed by DNA strand resection and fill-in synthesis. In this capacity, MMR is essential for the maintenance of genome stability. Among other things, defects in MMR lead to microsatellite instability – a characteristic feature of various cancers (reviewed in [139]). But not all microsatellites are equal in the eyes of the MMR machinery. Counterintuitively, functional MMR overall promotes instability of the microsatellites responsible for repeat expansion diseases, including $(CAG)_n$, $(CGG)_n$ and $(GAA)_n$ [140,141]. The individual effects of the MMR components on repeat instability depend on both the model organism and whether somatic or intergenerational instability is under investigation **(Table 1)**. In humanized FRDA mice, MMR affects repeat instability, although fine molecular mechanisms are somewhat controversial. That is, MUTSβ prevents intergenerational contractions, MUTSα precludes both expansions and contractions, while MUTLα counteracts

expansions but promotes contractions [142,143]. In somatic cells, specifically in neurological tissues, MUTLα suppresses expansions, but MUTSα promotes expansions [106].

In contrast to mice, MUTSβ promotes small-scale expansions in cultured human cells [107]. This depends on the activity of MUTLγ, and MUTLα protected against expansions, as observed in mouse models **(Figure 1-3 d)** [108]. In FRDA patient-derived induced-pluripotent stem cells (iPSCs), MUTSβ promotes large-scale $(GAA)_n$ expansions [144]. Whether allele variants of the individual proteins are actually modifiers of repeat stability in humans remains to be ascertained [145–148]. Notably, genetic analysis of $(GAA)_n$ repeat instability in non-dividing somatic cells cannot be easily done in patient-derived tissues, as they present a static picture or are passaged derived cells, which underwent further replication cycles.

To address this problem, Neil et al. developed a novel experimental system to study instability of a long $(GAA)_{100}$ repeat in chronologically aging quiescent (G0) cells in *S. cerevisiae* [149]. Three categories of mutagenic events were observed: repeat expansions, NHEJ-dependent deletions, and HR-mediated gene conversions. Whereas the main mutational event in dividing yeast cells is repeat expansions, the balance rapidly shifted to large deletions frequently including the whole $(GAA)_n$ repeat as the cells entered quiescence **(Figure 1-3 b)** Deletions were triggered by DSBs at the repeat generated during MMR as discussed in 4.1.. These DSBs are then repaired by Exo1-mediated resection and NHEJ. Consequently, a functional MMR machinery suppresses expansions of $(GAA)_n$ repeats specifically in non-dividing yeast cells (Table 1) [55,65,149]. Inactivation of MutSβ and MutLα and, above all, Exo1 increased both the frequency and the size of $(GAA)_n$ repeat expansions in quiescent cells. Finally, expansions involved processive DNA synthesis by Pol δ, likely occurring during error-prone repair of nicks accumulated in the damage-susceptible single-stranded parts of H-DNA [150]. Whereas large-scale

expansions occur in one step during replication [83], expansions in quiescent cells could result from multiple smaller-scale expansions [149].

In sum, MMR has been established as a major pathway modulating $(GAA)_n$ repeat stability, albeit with dramatic differences when it comes to the role of individual MMR components between different model systems. Note, also, that the effects of the MMR machinery on $(GAA)_n$ repeat instability are strikingly different from their role in $(CAG)_n$ repeat instability, as is comprehensively displayed in Table 1. In short, the MMR machinery binds but cannot excise long stable hairpins, further stabilizing those hairpins. Consequently, MMR promotes $(CAG)_n$ expansions in mice and humans [151–155]. The association between MMR and $(GAA)_n$ repeat instability clearly does not fit the same scheme. Furthermore, the effects of MMR on $(GAA)_n$ stability are clearly different in dividing and non-dividing cells, as evidenced by different effects of MutLα and MutSβ described above and in **Table 1**. The latter difference could result from different expression levels of MMR proteins as well as a different relationship between replication, transcription, and DNA repair.

Table 1: Role of MMR proteins during length instability of $(GAA)_n$ and $(CAG)_n$ repeats

Complex	MMR protein	GAA somatic and non-dividing	GAA intergenerational and replication-models	CAG somatic and non-dividing	CAG intergenerational and replication-models
MutLα MutLγ	MLH1	Promotes expansions Mouse model [143] No effect on expansions Promotes deletions Non-dividing yeast cells [149]	Promotes expansions Mouse model and human cells [108,143]	Promotes expansions HD mouse model [156] Human HD population [157]	Prevents expansions *S. cerevisiae* [158]
MutLα	PMS2 (Pms1 in	Prevents expansions Mouse model	Prevents expansions Promotes contractions	Promotes expansions	Prevents expansions

	S. cerevisiae)	Non-dividing yeast cells [106,149] Promotes deletions Non-dividing yeast cells [149]	Mouse model [142] Prevents expansions Human cells [108]	Prevents deletions Mouse model [159]	*S. cerevisiae* [158]
MutLγ	MLH3	Prevents expansions Human cells [108]	Prevents expansions Human cells [108]	Promotes expansions HD mouse models and patient derived cells [156,160]	No effect *S. cerevisiae* [158]
MutSβ MutSα	MSH2	Promotes expansions Mouse model Human cells [106,107]	Prevents contractions Mouse model [142] Promotes expansions iPSC GAA expansion model [144,161]	Promotes expansions Mouse models Human cell model [126,162,163]	Promotes expansions *S. cerevisiae* Mouse model [164,165] Prevents contractions *S. cerevisiae* [158]
MutSβ	MSH3	Prevents expansions Promotes deletions Non-dividing yeast cells [149] Promotes expansions Human cells and FRDA fibroblasts [107]	No effect on expansions G0 yeast cells [149] Prevents contractions Mouse model [142]	Promotes expansions Human cell model [126]	Promotes expansions HD mouse model DM1 mouse Human cell lines *S. cerevisiae* [164,166–170]
MutSα	MSH6	Promotes expansions Mouse model [106] No effect on expansions and promotes deletions Non-dividing yeast cells [149]	Prevents expansions and contractions Mouse model iPSC model [142,161] No effect on expansions Human cells [107]	Prevents expansions Human cell model [126]	Prevents contractions HD mouse model [166] Prevents expansions *S. cerevisiae* [164]

Base excision repair in $(GAA)_n$ repeat instability

Base excision repair (BER) is a DNA damage repair pathway which processes lesions initiated by oxidative damage, alkylation and base deamination. In short, the modified base is removed by a DNA glycosylase and the resulting abasic site is cleaved by the AP endonuclease creating a nick in a DNA strand. DNA polymerase β carries out repair DNA synthesis, which is followed by flap removal by flap-endonucleases and ligation. BER has been proposed to be a major pathway leading to small-scale, age-dependent somatic expansions of $(CAG)_n$ repeats [171,172]. Trinucleotide repeat stability is influenced by BER based on the site of DNA modifications within the repeat, the presence of additional proteins, and its balance with MMR processes [173].

Lai et al. showed that alkylation of $(GAA)_n$ repeats by the chemotherapeutic temozolomide results in BER-mediated $(GAA)_n$ repeat contractions both *in vitro* and in lymphoblasts of FRDA patients [174]. A structure-prone repetitive flap may compromise coordination between Pol β and flap cleavage by the flap endonuclease FEN1. BER was also proposed as a mechanism of repair of abasic sites within R-loops formed at $(CAG)_n$ repeat [158]. Laverde et al. conducted a biochemical study of BER activity on R-loop substrates containing a $(GAA)_{20}$ repeat. The BER enzyme AP endonuclease 1 (APE1) was found to incise the abasic site in the $(GAA)_n$ R-loop, creating a double flap intermediate that hinders synthesis by Pol β while stimulating 5'-flap cleavage by FEN1. Cleavage by FEN1 promotes R-loop resolution and contractions of about half the repeat length [175]. Therefore, the data so far indicate that processing of lesions in $(GAA)_n$ repeats by BER primarily results in repeat contractions.

What remains to be determined is the direct contribution of individual oxidizing agents and lesions on $(GAA)_n$ repeat stability. The position of the lesion relative to the repeat could also be involved in determining whether repair will result in expansions or contractions, as is the case

for $(CAG)_n$ repeats [176]. Oxidative damage is particularly relevant in the context of FRDA, as it is a prominent form of DNA damage in aging neurons and the frataxin protein itself is involved in the processing of oxidative damage [177,178]. Mitochondrial dysfunction, accumulation of reactive oxygen species (ROS) and subsequent cell death has been proposed as a driving cause of FRDA [179]. ROS accumulation promotes elevated levels of oxidative damage in the cell. If oxidative damage repair modulates $(GAA)_n$ repeat instability in somatic cells, it could be one of the main drivers of age-dependent somatic instability.

$(GAA)_n$ repeats are hotspots of homologous recombination

$(GAA)_n$ repeats were shown to promote homologous recombination in bacterial and yeast experimental systems [55,180,181], a feature they share with other trinucleotide repeats [182]. However, there are sensitive differences between the two systems. In bacteria, the recombinogenic potential of $(GAA)_n$ repeats decreased with their length, which was attributed to the formation of sticky DNA by longer repeats (see Section 2). In yeast, in contrast, the repeat's recombinogenic potential increased with its length. Furthermore, repeat-mediated recombination in bacteria, but not in yeast, occasionally led to length instability of the repeat itself. Finally, the data from the tetrad analysis in yeast indicated that repeat-mediated recombination occurred during the G1 phase of the cell cycle – it was replication-independent [180]. We want to emphasize that unlike $(CAG)_n$ repeats, $(GAA)_n$ repeat instability is not modulated by homologous recombination factors in yeast experimental systems, except for when the RNase H enzymes had been deleted, eliciting a BIR response [64,82,105,182].

DOES FAN1 PLAY A ROLE IN $(GAA)_N$ REPEAT INSTABILITY?

Cells contain a variety of structure-specific nucleases, which process flaps and other structures during DNA replication, repair, and recombination. Whereas each nuclease optimally processes a specific substrate, it is possible they can mis-recognize unusual DNA structures, influencing their stability. Two such nucleases are FEN1 (Rad27 in *S. cerevisiae*) and the Fanconi-associated nuclease FAN1. FEN1 processes 5' flaps of Okazaki fragments during lagging strand synthesis and DNA damage repair [183]. FAN1 is an interstrand cross-link (ICL) repair protein which exhibits 5'-to-3' exonuclease activity as well as endonuclease activity, participates in HR and processing of stalled forks [184]. FEN1 and FAN1 belong to the same class of enzymes and have been shown to have overlapping substrates, suggesting they might have similar or redundant roles in the regulation of repeat instability.

In yeast, Rad27 (FEN1) prevents instability of both $(GAA)_n$ and $(CAG)_n$ repeats [64,185–188], and multiple models propose that it does so through its flap equilibration abilities, which likely counteracts the formation of non-B DNA structures. In contrast, mammalian FEN1 does not seem to fully share this important role, as its depletion does not affect $(GAA)_n$ instability and has contrasting effects on $(CAG)_n$ stability, and has not been identified as a disease regulator [189–193].

What could explain this striking difference between organisms? It is possible that flap processing during Okazaki fragment synthesis might not be a major contributor to somatic instability overall. Alternatively, flap processing performed by other nucleases might be more important in human cells. Recently, FAN1 has emerged as a prominent candidate. Genome-wide association studies (GWAS) of Huntington's disease have identified FAN1 as a strong genetic modifier of disease onset, with FAN1 mutations being associated with earlier onset [103,194]. In addition, FAN1 prevents expansions of $(CGG)_n$ repeats in Fragile X syndrome mouse models

and rare missense variants in FAN1 were associated with expanded $(CGG)_n$ repeats present in individuals with autism spectrum disorders [195–197].

Goold et al. propose that FAN1 acts through a nuclease-independent pathway to stabilize $(CAG)_n$ repeats and prevent expansions, possibly by recruiting DNA damage repair proteins to the repeat, promoting conservative repair [198]. Candidates for key FAN1 interactors are its physical interactors in the MutL family [199,200]. FAN1 and MLH1 have been recently shown to have opposite but interdependent effects on $(CAG)_n$ repeat instability. FAN1 sequesters MLH1 through a *SPYF* motif, leaving it unable to promote $(CAG)_n$ expansions through its canonical MMR function [201–204] (Table 1). In these studies, the nuclease activity of FAN1 was needed to prevent expansions. Variants of FAN1 with reduced nuclease activity have been found in patients with particularly early HD onset and highlight the endo- and exonuclease activity of FAN1 in protecting against expansions [195,205]. Therefore, it has become clear that FAN1 is a major regulator of $(CAG)_n$ repeat stability [206].

Since MLH1 activity has been shown to regulate $(GAA)_n$ repeat instability in yeast and mouse models as well as in cultured human cells, it is foreseeable that the FAN1-MLH1 interaction might influence the balance of $(GAA)_n$ repeat stability possibly by the processing of 5' flaps generated by strand displacement during DNA synthesis. The latter role would more closely mirror the effects of Rad27 on $(GAA)_n$ repeat instability, as Rad27 does not interact with the MMR machinery in *S. cerevisiae* in otherwise unperturbed conditions [207]. Thus, we believe it is of great interest to study the role of FAN1 nuclease in $(GAA)_n$ repeat instability.

THERAPEUTIC AVENUES TARGETING $(GAA)_N$ REPEAT STABILITY

Gene editing and replacement therapies to modulate $(GAA)_n$ repeat size and stability

Friedreich's ataxia is a predominantly monogenic disease. Therefore, removal of expanded $(GAA)_n$ repeats at the *FXN* locus via gene editing is a potentially promising approach to modulate and even prevent disease. Excision of the $(GAA)_n$ repeat and some of the flanking sequence with zinc-finger nucleases in FRDA derived cells can partially rescue defects in *FXN* expression and ameliorates pathological phenotypes in iPSC-derived neuronal cells [208]. More recently, CRISPR technology has been applied to the study of repeat expansion diseases [209–211]. Removal of expanded $(GAA)_n$ repeats in an FRDA mouse cell line containing one expanded $(GAA)_{190}$ allele promoted partial transcriptional rescue of *FXN* gene expression, with an associated increase in protein levels [211]. The same result, though, was not observed in a cell line with two expanded alleles [211]. We envision that both length of the expanded allele and the relative size of the other allele can influence the success of CRISPR targeting.

Genome editing techniques to restore frataxin levels can be combined with cell replacement therapies to overcome an additional hurdle in therapy. Before the advent of CRISPR, same-species (allogeneic) transplant of healthy cells had been explored as a method to treat FRDA symptoms in mice [212]. The major hurdles of allogeneic transplantation are immunosuppression and graft rejection. The possibility of using the patient's own cells, modifying their genome, and reintroducing them into the patient (autologous graft) is therefore much more attractive, as it circumvents the mentioned issues. Rocca et al. removed the expanded repeat from human FRDA fibroblasts and hematopoietic stem and progenitor cells (HPSC), reaching a substantial increase in frataxin protein levels and rescue of mitochondrial defects. HPSCs then underwent hematopoiesis but displayed reduced cell proliferation rates [210].

Can CRISPR succeed also in the context of a functional brain? FRDA-derived iPSCs and embryonic stem cells differentiated into neuronal derivatives are able to withstand grafting into rodent brains, survive and even mature into dorsal root ganglia (DRG) – the primary tissue affected by neurodegeneration in FRDA [213,214]. Consequently, they constitute a model in which the utility of gene editing can be more reliably tested. FRDA-derived iPSCs have also been used to develop an *in vitro* 3D DRG organoid (DRGO) model [215]. Removal of the expanded $(GAA)_n$ repeat by CRISPR partially restored the frataxin protein levels and rescued FRDA-associated phenotypes, and the level of rescue was greater when shorter $(GAA)_n$ repeats were deleted. On the other hand, an almost complete deletion of the first *FXN* intron restored frataxin expression levels to approximately wildtype levels. The repressing chromatin markers associated with *FXN* transcription silencing in FRDA were permanently removed when the whole intron was removed, indicating that long $(GAA)_n$ repeats propagate chromatin silencing through its upstream and downstream regions [215]. Thus, removal of the expanded $(GAA)_n$ repeat alone might not be sufficient and will only work in association with the concomitant loss of repressive chromatin marks in its surroundings.

What are some of the caveats? First, it seems that the extent of the *FXN* region to be removed will have to be tailored to the starting $(GAA)_n$ repeat length of the individual patient, and different guides might be needed in each specific case. Secondly, off-target effects need to be carefully studied and minimized, and this will have to be tested for each target site. Third, whether the same approach can be applied to intergenerational instability remains to be determined. Finally, long-term studies are needed to test whether once the repeat has been shortened it remains stable over a long period of time, or whether it eventually becomes unstable again.

An alternative approach is gene addition, in which transgenic wild-type *FXN* is reintroduced into cells using viral vectors to rescue frataxin expression (reviewed in [216]). Since gene addition was successfully used for treating recessive genetic diseases in humans [217,218], this approach has been broadly investigated for FRDA, using various mouse models, patient-derived fibroblasts and in non-human primates ([216] and references therein). The main caveats in the FRDA case, however, was toxicity upon delivery and failure to rescue frataxin expression in neurological tissues.

Oligonucleotide-based approaches

The availability of repeat-stabilizing agents, such as small molecules that stabilize H-DNA or promote contractions, is being investigated as a complementary approach for treating repeat expansion diseases [52,219]. Since the protein coding sequence of frataxin remains unaltered in FRDA, upregulation of transcription and protein levels is a viable therapeutic avenue. Recently, the Napierala group has pioneered an oligonucleotide-based approach in which frataxin mRNA levels were stabilized resulting in increased frataxin protein levels in both FRDA fibroblasts and iPSC-derived neuronal progenitor cell lines [220]. Targeting the 5' and 3' untranslated regions of the *FXN* in combination led to a modest increase in mRNA half-life protein levels without altering the chromatin status of the *FXN* gene [220]. Since FRDA patients only have 5% to 35% of the control frataxin levels, it remains to be determined whether such an increase has the potential to alleviate frataxin-deficiency associated phenotypes [221,222].

Oligonucleotides can also be used to directly try and prevent $(GAA)_n$ repeat expansions. The first support for this idea comes from the use of locked nucleic acids (LNA), which have been shown to be able to interfere with triplex DNA formation [223]. LNA-DNA mixmers, which are not toxic for human cells, nearly completely prevented large-scale expansions of $(GAA)_n$

repeats in human cells [224]. This approach directly inhibits repeat expansions and is therefore very promising, albeit their effectiveness in FRDA patients remains to be determined. One of the major challenges regarding patient treatment with these oligonucleotides, as well as other promising small molecules [225–227], is the inefficiency of drug delivery to the central nervous system. While delivery to other affected tissues such as heart or pancreas can be achieved with viral vectors, the current ways of delivering drugs to the central nervous system are very invasive [228]. We hope that a better understanding of the mechanism of $(GAA)_n$ repeat instability and *FXN* expression during human development would lead to defining the most effective spatiotemporal windows for long-lasting treatment.

AKNOWLEDGEMENTS

We would like to thank Victoria Brown, Tyler Maclay and Julia Hisey for helpful discussion and proofreading the manuscript. We are also grateful to the anonymous reviewers for their insightful comments and suggestions during the publishing process. This work was supported by National Institute of General Medical Sciences grant R35GM130322 to S.M.M.

CRediT AUTHOR STATEMENT

Chiara Masnovo: Conceptualization, Writing- Original draft preparation, Writing- Reviewing and Editing **Ayesha F. Lobo:** Visualization **Sergei M. Mirkin:** Conceptualization, Writing- Reviewing and Editing.

CHAPTER 2 Stabilization of expandable DNA repeats by the replication factor Mcm10 promotes cell viability

Chiara Masnovo[1], Zohar Paleiov[2], Daniel Dovrat[2], Laurel K. Baxter[1], Sofia Movafaghi[1], Amir Aharoni[2], Sergei M. Mirkin[1]

[1] Department of Biology, Tufts University, Medford, MA, USA 02155

[2] Department of Life Sciences and the National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Be'er Sheva, Israel 8410501

ABSTRACT

Trinucleotide repeats, including Friedreich's ataxia $(GAA)_n$ repeats, become pathogenic upon expansions during DNA replication and repair. Here, we show that deficiency of the essential replisome component Mcm10 dramatically elevates $(GAA)_n$ repeat instability in a budding yeast model by loss of proper CMG helicase interaction. Supporting this conclusion, live-cell microscopy experiments reveal increased replication fork stalling at the repeat in *mcm10-1* cells. Unexpectedly, the viability of strains containing a single $(GAA)_{100}$ repeat at an essential chromosomal location strongly depends on Mcm10 function and cellular RPA levels. This coincides with Rad9 checkpoint activation, which promotes cell viability, but initiates repeat expansions via DNA synthesis by polymerase δ. When repair is inefficient, such as in the case of RPA depletion, breakage of under-replicated repetitive DNA can occur during G2/M, leading to loss of essential genes and cell death. We hypothesize that the CMG-Mcm10 interaction promotes replication through hard-to-replicate regions, assuring genome stability and cell survival.

INTRODUCTION

Expandable DNA repeats are at the heart of over 50 diseases, spanning from neurodegenerative disorders to cancer [4,229]. Expansions of $(GAA)_n$ repeats are known to cause two diseases: Friedreich's ataxia (FRDA) upon biallelic repeat expansions in the 1st intron of the Frataxin (*FXN*) gene [9] and spinocerebellar ataxia 27B (SCA27B) caused by repeat expansions in the 1st intron of the Fibroblast Growth Factor 14 (*FGF14*) gene [230,231]. $(GAA)_n$ repeats are a subgroup of homopurine-homopyrimidine mirror repeats that can fold into an alternative DNA secondary structure called H-DNA – an intramolecular DNA triplex, which was shown to hinder both DNA replication and transcription [28,29,31,110,232].

Studies in *S. cerevisiae* and human cells collectively showed that (i) expanded $(GAA)_n$ repeats stall replication fork progression [50,52,53,56,233], (ii) mutations in replication-associated genes including replicative DNA polymerases promote $(GAA)_n$ repeat instability [64,65,83,234], and (iii) processes that deal with stalled replication fork repair and restart, such as template switching and restoration of reversed replication forks, modulate $(GAA)_n$ repeat stability and trinucleotide repeat stability generally [56,65,86,235]. Nevertheless, much about the interaction between $(GAA)_n$ repeats, their structure, and the replication fork remains to be elucidated.

Natural replication impediments, including DNA structures, can cause physical uncoupling of leading strand synthesis progression from CMG unwinding and lagging strand synthesis [236–242]. As a result, single-stranded DNA (ssDNA) is exposed and coated by the ssDNA binding protein RPA, triggering the activation of the intra S-phase checkpoint, which ultimately leads to fork restoration and safeguards genome integrity [237,238,243–246]. Therefore, physical and functional coordination of the replication fork could be central to repeat length maintenance. In

addition, components that might be dispensable for replication elongation during unperturbed replication might become more important when replicating through repetitive sequences such as long $(GAA)_n$ repeats.

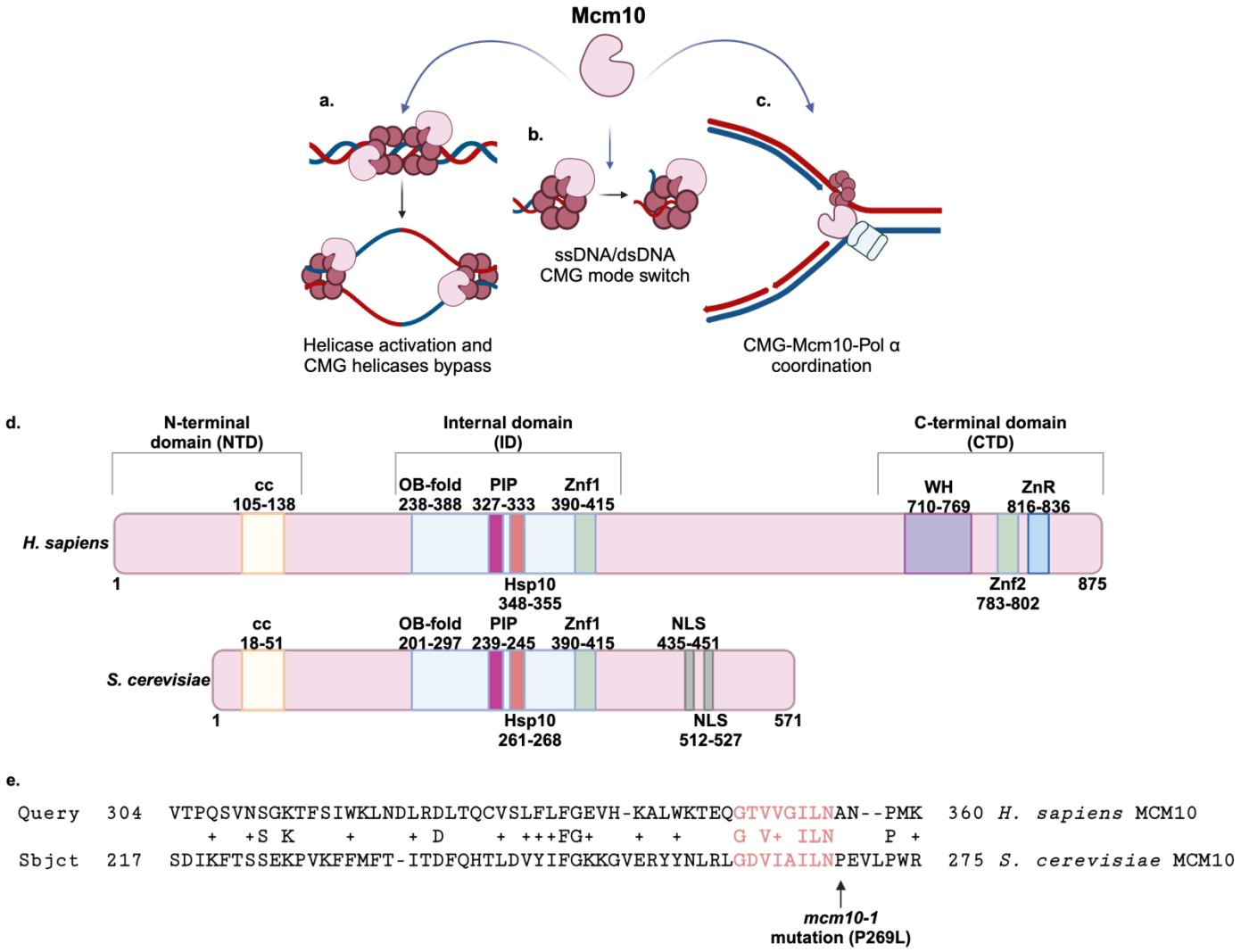

Figure 2-1: Roles of Mcm10 during DNA replication.

(a) Mcm10 promotes CMG helicase isomerization and bypass during replication initiation. **(b)** Mcm10 allows for a CMG mode switch between encircling ssDNA and dsDNA, which is important for replication restart. **(c)** Mcm10 interacts with both the CMG helicase and Pol α during replication elongation, stabilizing Pol α and coordinating leading and lagging strand synthesis. **(d)** Schematic of the domains of human and budding yeast Mcm10 proteins. **(e)** Partial amino-acid BLAST alignment of human and budding yeast Mcm10. Residyes in orange indicate the Hsp10-like domain where the *mcm10-1* mutation is located. Created with BioRender.com

The integrity of the replication fork during elongation and coordination between leading and lagging strand synthesis are promoted by accessory replication fork proteins – including Ctf4^{AND-1} and Mcm10 – which have roles in both replication initiation and elongation [247–253]. Both Ctf4 and Mcm10 interact with the CMG helicase as well as with Pol α-primase [251,254–263], and Mcm10 also contacts the PCNA clamp [264]. Ctf4 and Mcm10 interact with each other both in mammalian cells and in yeast, and Mcm10 stabilizes Ctf4 on chromatin [257,265]. While Ctf4 is non-essential in yeast and only becomes crucial for replication in the context of low Pol α levels [252], Mcm10 is essential for replication in all organisms in which it is present. It is important to note that direct interactions between the CMG helicase and Pol α-primase have recently been shown to recruit the complex to the lagging-strand template and promote priming without the need for Ctf4 or Mcm10 [266,267].

Mcm10 comprises an N-terminal coiled-coil domain important for oligomerization, an internal domain that includes an OB-fold with a PIP box and an Hsp10-like domain, which are highly conserved from yeast to humans **(Figure 2-1 d)**. Metazoans have an additional C-terminal domain mediates further interactions with DNA and proteins (reviewed in [250]). During replication initiation, Mcm10 contributes to the activation of the assembled CMG helicase and origin unwinding by facilitating the bypass of the two CMG hexamers [268–270] **(Figure 2-1 a)**. The ssDNA/dsDNA gate function of Mcm10 has also been implicated in promoting bypass of lagging strand blocks in a manner mediated by its interactions with MCM [271–273] **(Figure 2-1 b)**, but whether Mcm10 has a more prominent role in elongation as part of the replisome through its interactions with Pol α and under which conditions remains to be determined **(Figure 2-1 c)**. Deficiencies in Mcm10 lead to impaired replication initiation, slower replication, increased ssDNA exposure, DNA damage and checkpoint activation [256,274–277]. Furthermore, *MCM10*

haploinsufficiency leads to telomere erosion and micronuclei formation in iPSC cells, indicating that it has an important role in preventing genome instability [278].

In this chapter, we focused on the role of the replication factor Mcm10 in the stability of expanded $(GAA)_n$ repeats in a yeast model system. We found that the Mcm10 protein strongly counteracts both repeat expansions and contractions. Strikingly, the viability of yeast strains containing unique expanded $(GAA)_n$ repeats at an essential portion of a chromosome arm is substantially decreased in Mcm10-deficient strains. Cell survival in this case is ensured by the Rad9 checkpoint activity, which facilitates DNA repair synthesis by DNA polymerase δ while simultaneously promoting expansions.

RESULTS

Mcm10 deficiency elevates $(GAA)_n$ repeat instability due to impaired interactions with the CMG helicase

To study the role of Mcm10 on $(GAA)_n$ repeat instability we used an experimental system previously established in the lab [65]. In this system, a $(GAA)_{100}$ repeat is located within the intron of an artificially split *URA3* gene on chromosome III adjacent to the *ARS306* origin. The repeat is flanked by non-repetitive sequences, for a total intron length of 974 bp. In *S. cerevisiae,* only introns shorter than ~1 kb can be spliced efficiently [279]. Thus, repeat expansions that bring the total intron length over this threshold result in the inactivation of the *URA3* gene, making the yeast cells resistant to 5-fluoroorotic acid (5-FOA). Other events, such as mutations and various recombinational events can also result in *URA3* loss [83,280] (**Figure 2-2 a)**. Therefore, repeat expansions were confirmed by PCR using repeat-flanking primers (Source Data file). In the

contraction assay, a longer $(GAA)_{124}$ tract is inserted in the intron, bringing the total intron length to 1108 bp. Contractions of more than 20 repeat units reactivate the *URA3* gene making the yeast cells URA^{+} [64] **(Figure 2-2 b)**. In both the expansion and contraction cassettes, the $(GAA)_n$ repeats serve as the lagging strand template.

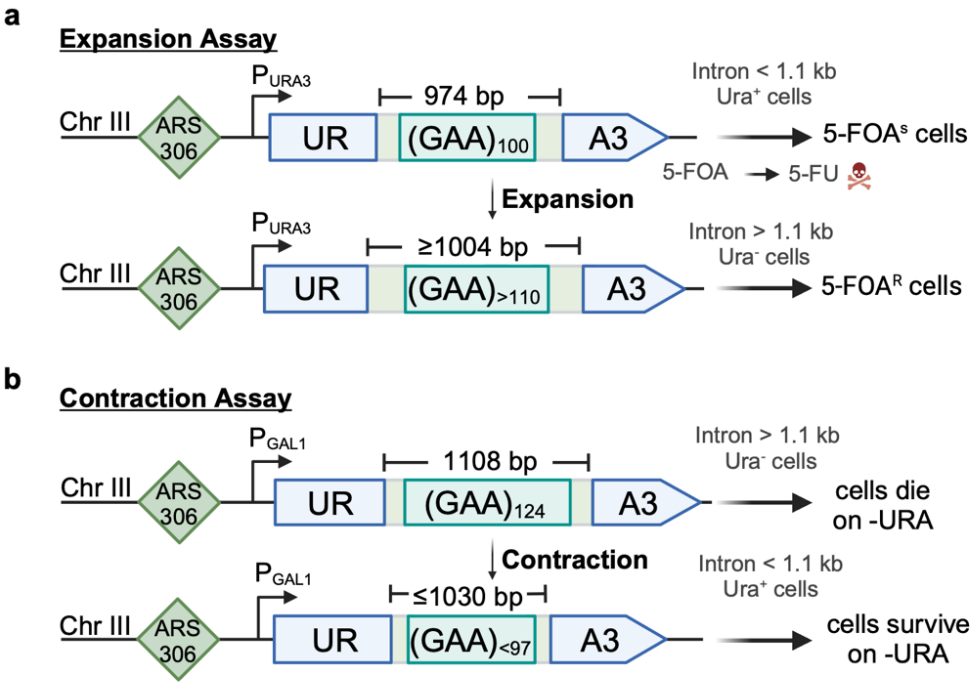

Figure 2-2: Schematic of the instability assays used in this Chapter and Chapter 3.

(a) Genetic assay system to measure repeat expansion rates. **(b)** Genetic assay system to measure repeat contraction rates. Created with BioRender.com

To determine how Mcm10 affects repeat instability in these systems, we introduced a previously characterized *mcm10-1* mutation. This *P269L* substitution lies in the structurally and functionally conserved Hsp10-like domain – a part of the larger internal domain of Mcm10 responsible for its interactions with DNA, Pol α and PCNA **(Figure 2-1 e)** [281]. This mutation

results in a temperature-sensitive (*ts*) phenotype, owing to both Mcm10 and Pol α degradation at restrictive temperatures [254]. We found that the *mcm10-1* mutation led to a 33-fold increase in expansion rate at the semi-permissive temperature (27°C), and a 6-fold increase even at the permissive temperature (23°C) **(Figure 2-3 a)**. We also observed a 10-fold increase in repeat contractions at the semi-permissive temperature **(Figure 2-3 b)**. We conclude that Mcm10 is an important replication factor in preventing the instability of long $(GAA)_n$ repeats, especially their expansion.

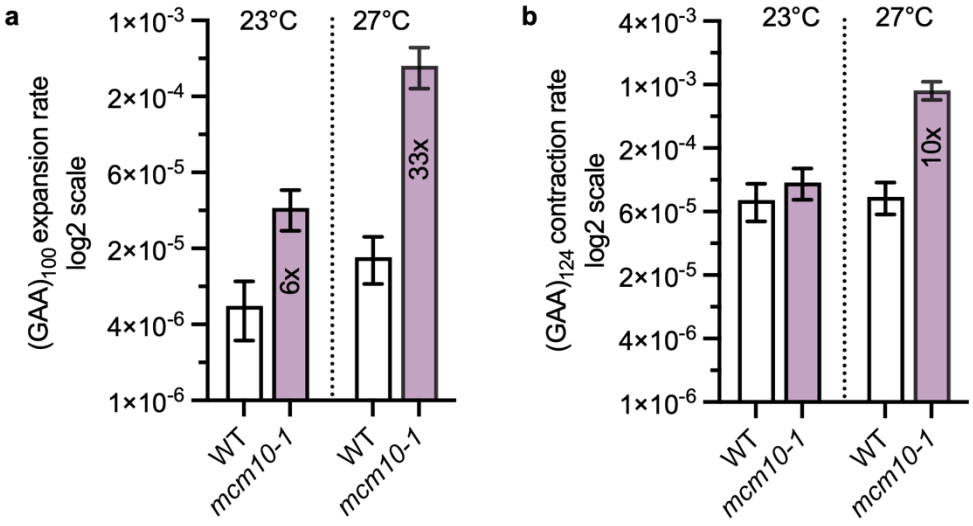

Figure 2-3: Effects of the *mcm10-1* mutation on (GAA)n repeat instability.

(a) Expansion rates in the *mcm10-1* mutant at the permissive (23°C) and semi-permissive (27°C) temperatures. **(b)** Contraction rates for the *mcm10-1* mutant at the permissive (23°C) and semi-permissive (27°C) temperatures. Plotted values indicate the corrected rate calculated with FluCalc (https://flucalc.ase.tufts.edu/) [280] and the error bars represent 95% confidence intervals. Numbers within bars indicate fold increase over the respective wild-type value. Expansion rates were determined by PCR of at least 96 FOA^R colonies derived from two biological replicates. An event was considered an expansion when at least 10 repeats were added as detected by PCR. All expansion and contraction data for this chapter can be found in **Table 7** and **Table 8**.

Mcm10 has been shown to interact with the CMG helicase, thereby promoting both replication initiation and elongation [262,268,271,272]. It specifically interacts with a conserved motif

in Mcm2, as well as with other MCM subunits of the CMG helicase [272,282]. To study the role of Mcm10-CMG interactions in repeat instability, we looked at the effects of a previously identified dominant suppressor mutation in the Mcm2 subunit of the CMG helicase (*mcm2G400D*), which was shown to rescue the temperature-sensitive phenotype of the *mcm10-1* mutant, minimizes ssDNA exposure and restores Pol α stability, particularly rescuing the elongation defects observed in the *mcm10-1* mutant [274]. This mutation is located in the allosteric control loop of the Mcm2 subunit, which is important to couple activities between CMG subunits [283], thus, on its own, it results in a decrease in the unwinding activity of the helicase [274]. We found that the *mcm2G400D* mutation alone did not affect repeat instability **(Figure 2-4 a and b)**, indicating that a decrease in the helicase unwinding rate alone is insufficient to trigger repeat instability. At the same time, we observed a near complete rescue of the elevated expansion and contraction rates in the *mcm10-1 mcm2G400D* double mutant compared to the *mcm10-1* mutant alone **(Figure 2-4 a and b)**. At the permissive temperature (23°C), expansions are only partially rescued **(Figure 2-4 c)**, whereas contractions are completely rescues in this case as well **(Figure 2-4 d)**. Together, these data suggest that Mcm10 prevents repeat instability through its interaction with the CMG helicase during replication through the repeat.

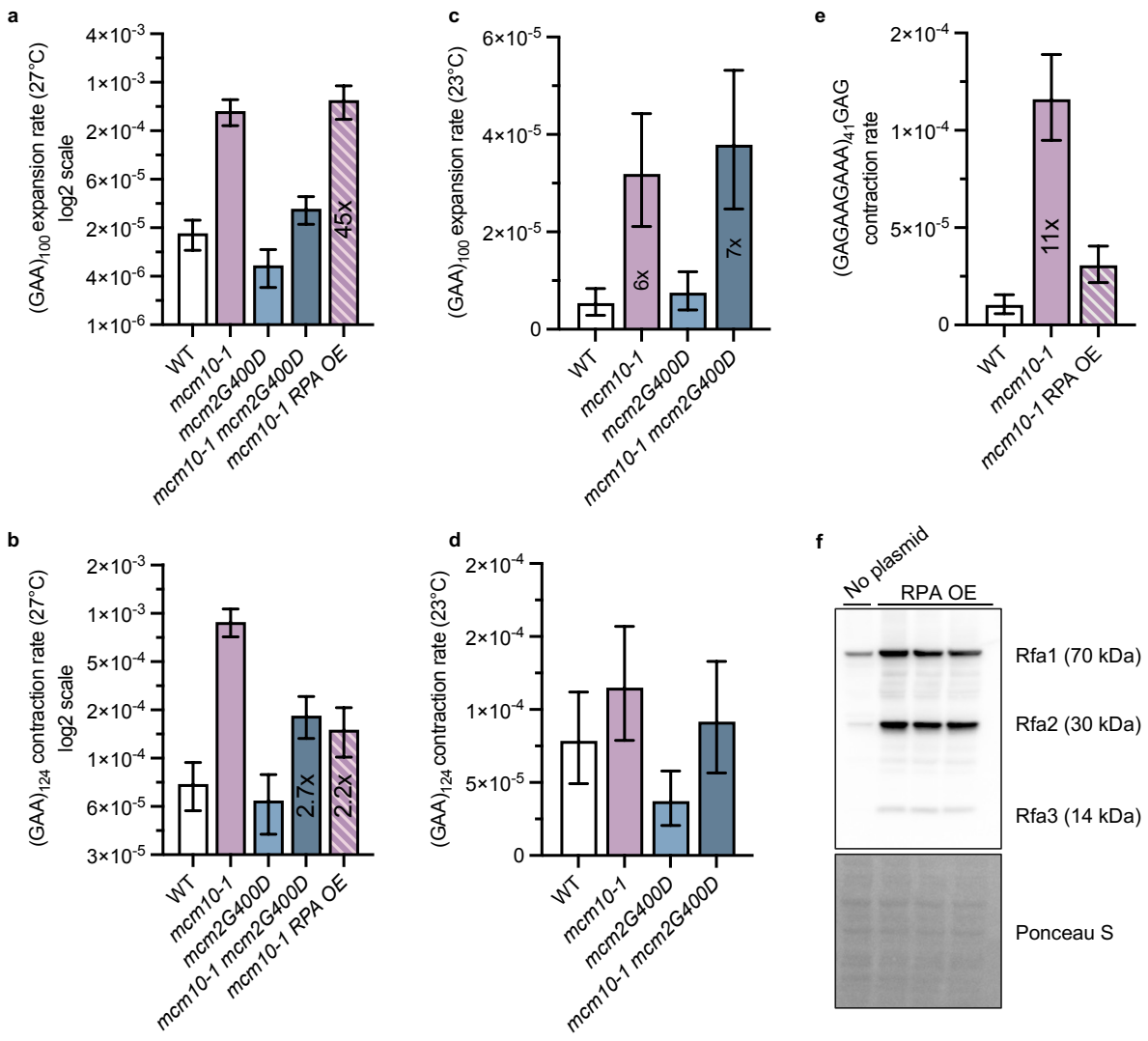

Figure 2-4: A suppressor mutation in the Mcm2 subunit of the CMG helicase rescues instability of mcm10-1.

(a) Expansion rates in the *mcm10-1* and *mcm2-G400D* mutants at the semi-permissive temperature (27°C). *RPA OE* indicates RPA overexpression via 2μ multicopy plasmid containing all three RPA genes: *RFA1*, *RFA2* and *RFA3*. **(b)** Contraction rates in the *mcm10-1* and *mcm2-G400D* mutants at 27°C. **(c)** Expansion rates of the *mcm2-G400D* mutants at the permissive temperature (23°C). **(d)** Contraction rates of the *mcm2-G400D* mutants at 23°C. **(e)** Contraction rates of $(GAGAAGAAA)_{41}$GAG repeat in the *mcm10-1* and *mcm10-1* RPA overexpression conditions at 27°C. **(f)** Western blot analysis of various strains either containing the 2μ multicopy plasmid containing all three RPA genes (*RFA1*, *RFA2* and *RFA3*), or without plasmid. The RPA antibody (AS07-214) detects all three RPA subunits: Rfa1p (70 kDa), Rfa2p (30 kDa) and Rfa3p (14 kDa). Ponceau S staining was used as a control for total protein loading. For a-e the plotted values indicate the corrected rate and the error bars represent 95% confidence intervals. Numbers within bars indicate fold increase over the respective wild-type value.

$(GAA)_n$ repeat instability in *mcm10-1* is not caused by lower levels of DNA polymerase α-primase

Mcm10 function has been proposed to be important for the stability of the Pol α-primase complex both in yeast and human cells [255,256]. We investigated whether the phenotypes of the *mcm10-1* mutants resulted from Pol α-primase complex degradation. The Pol1 subunit of Pol α-primase was tagged with a 3x Flag-tag and protein levels were measured by western blotting. We indeed observed lower levels of Pol1 in *mcm10-1* mutants compared to wild-type at the restrictive temperature (37°C), but not at the semi-permissive temperature of 30°C, which is higher than the temperature we conducted the instability assays at **(Figure 2-5 a)**. In addition, Pol α-deficient cells were previously shown to have larger expansions, likely resulting from longer Okazaki fragments [83]. We do not observe a meaningful difference in the median number of repeats added in the case of the *mcm10-1* mutant (60 repeat units) when compared to the wild-type (62 repeat units) **(Figure 2-5 b)**, albeit two sample Kolmogorov–Smirnov test shows a small but significant difference in the shape of repeat distribution compared to WT.

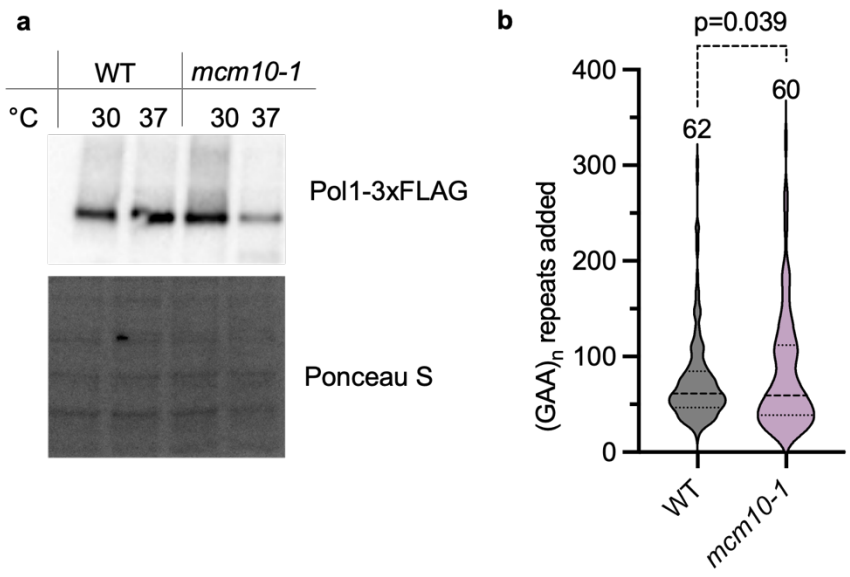

Figure 2-5: Effects of the mcm10-1 on Pol α levels in the tested strains.

(a) Western blot of Pol1 (Cdc17) tagged with a 3xFLAG tag in WT and *mcm10-1* strains at the semi-permissive (30°C) and at the restrictive (37°C) temperatures. Ponceau S staining was used as a control for total protein loading. **(b)** Distribution of the number of added repeats units in the WT and *mcm10-1* from fluctuation assay experiments in **(Figure 2-3 a)**. Dashed lines represent the median number of repeats added and quartiles, numbers above bars are the median. Pairwise comparisons of distributions were conducted using the nonparametric goodness-of-fit Kolmogorov-Smirnov test. p=0.039.

Finally, we tested a different Mcm10 mutant, *mcm10-G261D*, which was previously shown to lead to a decrease in Pol1 levels while maintaining Mcm10 levels [284]. In this mutant, we did not observe an increase in instability over wild-type levels for either type of instability **(Figure 2-6 a and b)**. We note, however, that we did not determine the Pol1 protein levels in our hands in the case of this mutant, which would be needed to make a stronger conclusion based on this set of experiments.

Altogether, the results presented in this section indicate that lower Pol α levels are not responsible for increased instability in the *mcm10-1* context.

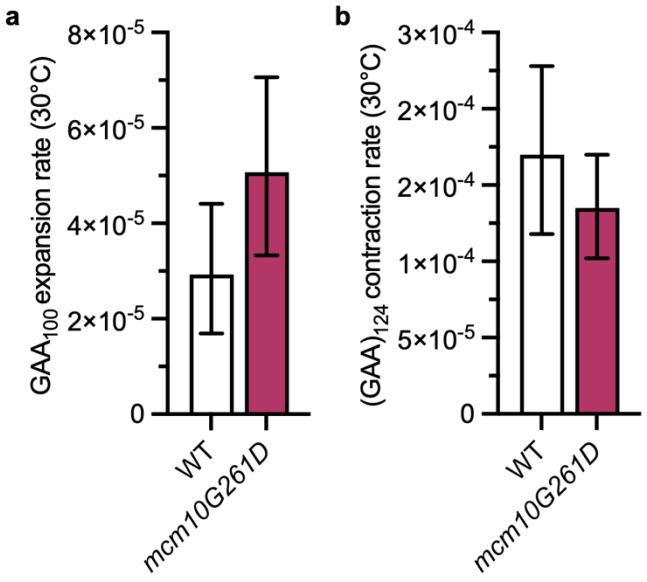

Figure 2-6: $(GAA)_n$ repeat expansion (a) and contraction (b) rates in the *mcm10-G261D* mutant. Plotted values indicate the corrected rate and the error bars represent 95% confidence intervals.

Mcm10 facilitates replication elongation through $(GAA)_n$ repeats

Mcm10 is both an initiation and an elongation factor, and multiple studies showed that it is needed for efficient and complete DNA replication [285]. Since expanded $(GAA)_n$ repeats were previously shown to cause replication fork stalling [52,53,56], and our current data suggest that Mcm10 is needed for stable repeat maintenance, we decided to investigate replication fork progression through the $(GAA)_{100}$ repeats in the *mcm10-1* context.

To this end, we adopted a method for live-cell imaging of replication fork progression **(Figure 2-7 a)** [286]. In this system, a non-repetitive 128xlacO and 128xtetO arrays are placed ~3 kb and ~37 kb downstream from the *ARS413* replication origin, respectively. These arrays are bound by their cognate LacI-Envy and TetR-tdTomato, resulting in a green and red focus, respectively. Replication of each array coincides with an increase in intensity of the respective

focus. The time difference between replication of the lacO and tetO arrays is used to monitor fork progression through the repeat sequences in live cells. Inserting the repeat in the sequence between the two arrays allowed us to monitor fork progression through the repeats and determine the effect of *mcm10-1* with $(GAA)_{100}$ serving as the lagging-strand template **(Figure 2-7 b)**.

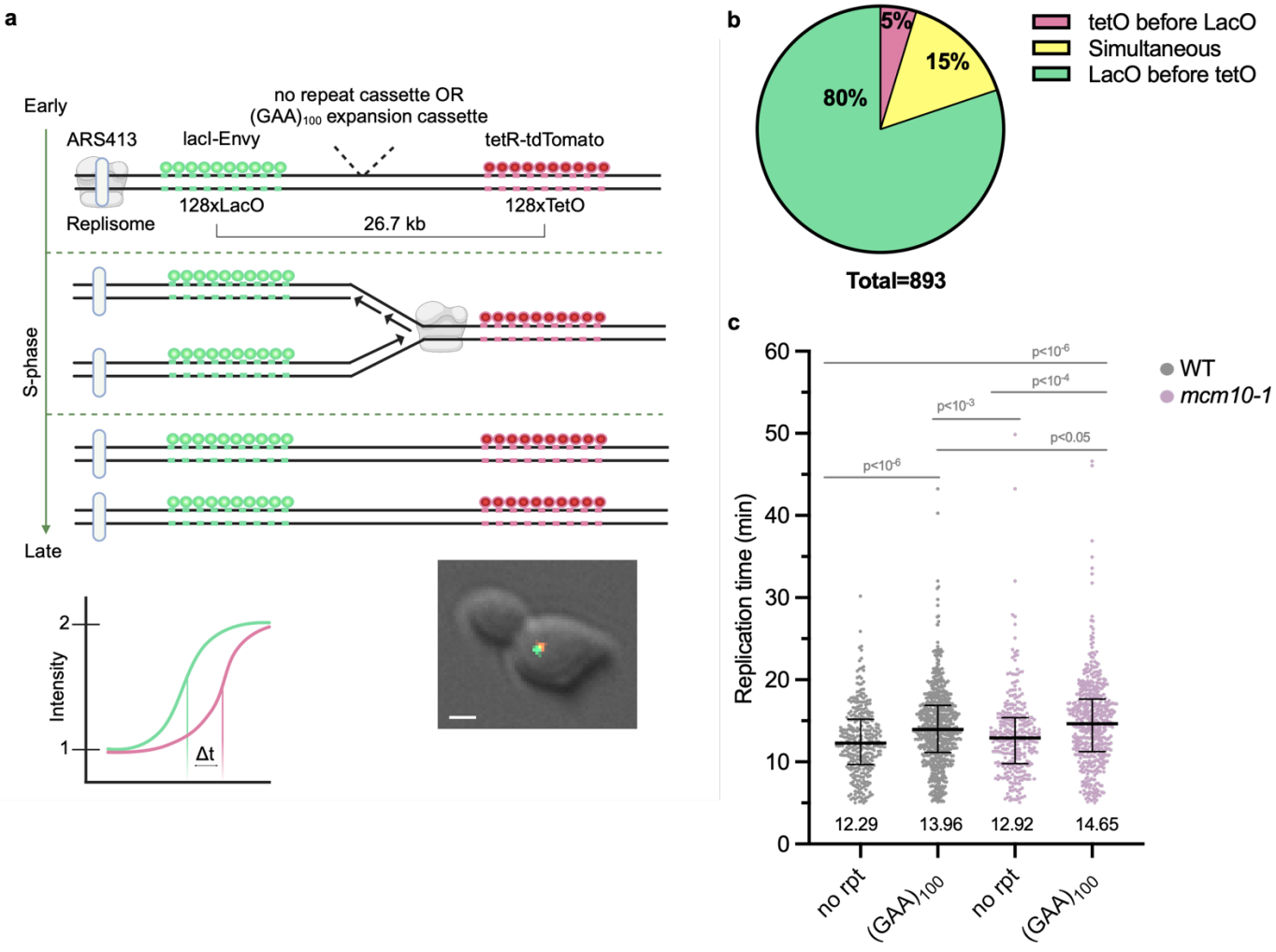

Figure 2-7: Live-cell microscopy of replication fork progression through $(GAA)_{100}$ repeats in WT and *mcm10-1* cells.

(a) Schematic of the live-microscopy assay to measure replication fork progression (created with BioRender.com), including a representative image from the microscopy measurements. Scale bar represents 2 μm. **(b)** Replication fork directionality chart for the wild-type live-cell microscopy strain containing the expansion cassette with the $(GAA)_{91}$ repeat (strain CM454). Cells were counted as "lacO before tetO" (left-to-right fork) if the Δ time between the doubling in intensity

of the lacO and tetO was ≥ + 5 min, as "simultaneous" if the Δ time was between -5 and + 5 min and as "tetO before lacO" (right-to-left fork) if the Δ time was ≤ - 5 min. Data is represented in % of cells in each category. **(c)** Time required to replicate a region of 28.4 kb, in WT and *mcm10-1* strains containing the expansion cassette with either no repeat or the $(GAA)_{91}$ repeat at 30°C. The plotted numbers represent the median value and bars represent the interquartile range. Statistical analysis of replication rates was performed using Monte Carlo resampling with 1,000,000 iterations. p values are indicated for significant comparisons.

We compared strains carrying the expansion cassette with the $(GAA)_{100}$ repeat to a those carrying the same cassette but with a filler sequence in place of the repeat (no repeat control). Wild-type and *mcm10-1* derivatives of those strains were analyzed at the semi-permissive temperature of 30°C. In accordance with our previous data [65], there was repeat-mediated slowing of replication in the wild-type strains **(Figure 2-7 c)**. There was a small, non-significant slowdown caused by the *mcm10-1* mutation in the no repeat control, but the difference between WT and *mcm10-1* became significant in the case of the $(GAA)_{100}$ repeat **(Figure 2-7 c)**. This comparison indicates that Mcm10 promotes replication elongation, especially when replicating through the expanded $(GAA)_n$ repeats.

Single-stranded DNA at the replication fork primarily promotes repeat contractions

Exposure of ssDNA during replication promotes the formation of non-B DNA structures and overexpression of the single-stranded DNA-binding RPA complex counteracts contractions of long $(GAA)_n$ repeats [64]. The *mcm10-1* mutation was shown to cause an increase in RPA foci at the non-permissive temperature, pointing to the accumulation of ssDNA [274]. We therefore hypothesized that an increased single-strandedness of the $(GAA)_n$ repeat in the *mcm10-1* mutant could result in its instability. To test this hypothesis, we overexpressed all three RPA subunits on a multicopy 2μ plasmid [64] and verified their overexpression via western blot **(Figure 2-4 f)**. In

line with our hypothesis, RPA overexpression rescued the elevated contraction rate observed in the *mcm10-1* mutant **(Figure 2-4 b)**. In contrast, the expansion rate in the *mcm10-1* mutant was not rescued at all by RPA overexpression **(Figure 2-4 a)**.

The $(GAGAAGAAA)_{41}GAG$ repeat is a homopurine-homopyrimidine repeat of the same length and GC-content as the $(GAA)_{124}$ repeat, but lacks the mirror symmetry required for the formation of a stable triplex structure[64]. In accordance with our previous data [64], its contraction rate is about 10-fold lower than the one of the $(GAA)_{124}$ repeat **(Figure 2-4 e)**. However, the *mcm10-1* mutation led to a similar fold increase in contraction rate for the $(GAGAAGAAA)_{41}GAG$ repeat as for the $(GAA)_{124}$ repeat and this effect is similarly rescued by RPA overexpression. Altogether, these results show that an excess of uncoated ssDNA at the repeat, rather than strong triplex formation, accounts for the elevated contraction rate in the *mcm10-1* mutant. An increase in overall ssDNA exposure in the *mcm10-1* mutant has been observed previously[274], and is likely due to the mis-coordination between the helicase and the lagging-strand synthesis machinery.

To explore this further, we first triggered replication fork uncoupling of the CMG helicase and polymerase ε using hydroxyurea (HU). HU depletes the deoxyribonucleotide pool available during replication resulting in fork uncoupling, replication stress and increased ssDNA exposure [287]. Treatment with 100 mM HU moderately increased $(GAA)_{100}$ expansions (3.4-fold) **(Figure 2-8 a)** and had a more substantial increase on repeat contractions (8.3-fold) **(Figure 2-8 b)**. We then looked at the role of the Ctf4 protein trimer, which was proposed to have a role in replication fork coordination similar to Mcm10 by coordinating CMG and Polymerase α [257,260,265]. Deletion of the *CTF4* gene had no effect on repeat expansions and only had a modest 3.6-fold increase in repeat contraction rate **(Figure 2-8 c and d)**.

Finally, we also mutated the Rfc1 subunit of the RFC clamp loader complex. Loading of the PCNA clamp by the RFC complex promotes processivity of Polymerase δ (reviewed in [288]). The cold-sensitive *rfc1-1* mutation (*D513N*) resides in the putative nucleotide binding domain of the protein and causes defective kinetics in PCNA loading and unloading, which should primarily affect lagging strand synthesis processes [289–291]. At the permissive temperature (30°C), *rfc1-1* had no effect on $(GAA)_{100}$ expansions but caused a 15-fold increase in repeat contractions **(Figure 2-8 e and f)**. Under semi-permissive conditions (27°C), both expansion and contraction rates were elevated, with the effect on contractions being further exacerbated **(Figure 2-8 e and f)**. Since PCNA needs to be continuously reloaded during lagging strand synthesis, we conclude that expansions, unlike contractions, do not primarily occur during lagging strand synthesis and are not strongly affected by its processivity. Altogether, these results indicate that the observed phenotypes are intrinsic to Mcm10 and are not shared with other replication fork coupling modalities, which seem to be predominantly preventing repeat contractions. This indicates that in the *mcm10-1* mutant, the temporary increase of ssDNA at the replication fork (especially on the lagging-strand template) is not the driving force responsible for the expansion events. This could mean that 1) there is persistence of ssDNA accumulation in *mcm10-1* and/or 2) the ssDNA gaps mostly affect the leading strand template and are processed through a different mechanism.

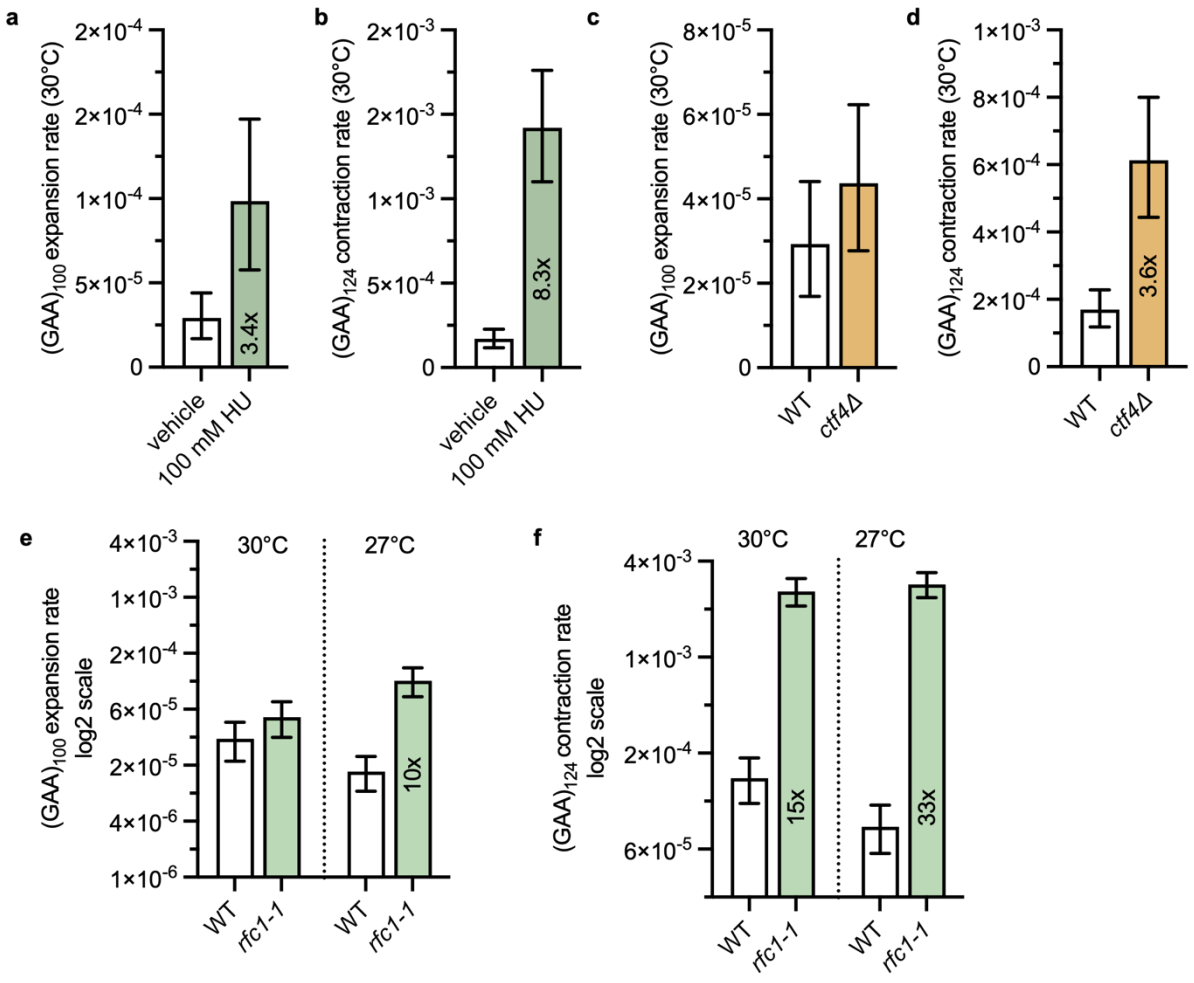

Figure 2-8: Effects of hydroxyurea and mutations affecting fork coordination on $(GAA)_n$ repeat instability.

(a) Rates of expansion in vehicle (H_2O) versus 100 mM HU treatment conditions (30°C). **(b)** Rates of contraction in vehicle (H_2O) versus 100 mM HU treatment conditions (30°C). **(c)** Rate of expansion in *ctfΔ* strains at 30°C. **(d)** Rate of contraction in in *ctfΔ* strains at 30°C. **(e)** Rates of expansion for the *rfc1-1* mutation at the permissive (30°C) and semi-permissive (27°C). **(f)** Rates of contraction for the *rfc1-1* mutation at the permissive (30°C) and semi-permissive (27°C). For a-f, the plotted values indicate the corrected rate and the error bars represent 95% confidence intervals. Numbers within bars indicate fold increase over the respective wild-type value.

Mcm10 deficiency causes repeat length- and position-dependent viability defects

During our instability assays with the *mcm10-1* mutants, we observed a viability defect more pronounced than previously described in the literature [274,292]. We wondered if the presence of the

long repeat could be responsible for this phenotype. We conducted serial dilutions and observed that at the semi-permissive temperature of 30°C, strains containing both the *mcm10-1* mutation and the $(GAA)_{100}$ repeat on the lagging strand template on our chromosome III location had a viability defect when compared to the *mcm10-1* strain without the repeats, which only showed a delay in growth **(Figure 2-9 a)**. In a more quantitative assay, we compared colony-forming units (CFUs) at 30°C and 23°C and saw that a carrier length of $(GAA)_{40}$ repeats already caused a significant decrease in viability compared to the no-repeat strain. This effect was further exacerbated at the disease-causing lengths of $(GAA)_{100}$ and $(GAA)_{124}$, leading to a striking ~80% loss of viability (**Figure 2-9 c)**. This effect was also observed, albeit in a milder form, when the $(TTC)_n$ run serves as the lagging strand template (**Figure 2-9 d)**. To establish if the viability defects are caused by the triplex-forming potential of the $(GAA)_n$ repeat, we analyzed the viability of *mcm10-1* strains carrying the $(GAGAAGAAA)_{41}GAG$ repeat, and we observed a viability decrease comparable to that of the $(GAA)_{124}$ repeat (**Figure 2-9 c)**. As was observed in the case of repeat instability, the *mcm10-1 mcm2G400D* double mutant fully rescued the viability defects for both types of homopurine-homopyrimidine repeats. Minimizing the presence of uncoated ssDNA at the repeat by overexpressing the RPA complex also led to a viability rescue for both repeats (**Figure 2-9 c)**. The formation of ssDNA at the repeat in *mcm10-1*, which is due to fork miscoordination, therefore affects repeat viability independently of formation of a strong DNA triplex.

Notably, our instability cassettes are historically located within the essential arm of chromosome III, where loss of telomeric-proximal DNA would lead to loss of cell viability. We then moved our expansion cassette adjacent to the *ARS507* replication origin in a non-essential location on a chromosome V arm, making the loss of the telomere-proximal chromosome arm

possible [122,293]. In this case, the viability of the *mcm10-1* strains bearing the $(GAA)_{100}$ repeat was the same as the no-repeat control in the same location **(Figure 2-9 b)**, albeit the rate of cell growth was slower. Altogether, these results imply that cells containing expanded homopurine-homopyrimidine repeats rely on Mcm10 function for survival only when the repeats are located on an essential chromosome arm.

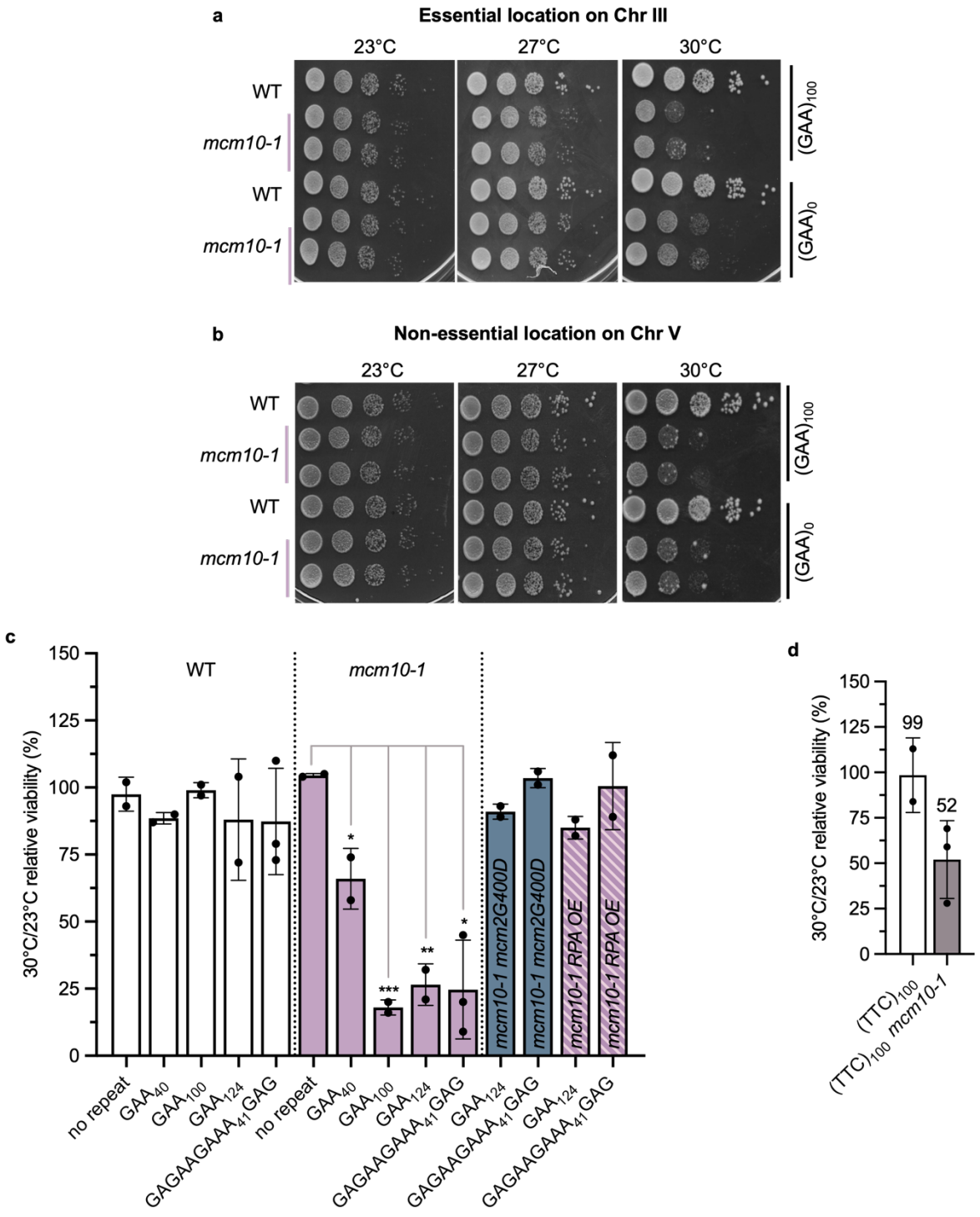

Figure 2-9: The *mcm10-1* mutation causes repeat-length dependent viability defects due to the loss of an essential chromosome arm.

(a) Spot tests of strains containing either a no repeat cassette or the expansion $(GAA)_{100}$ repeat cassette located at an essential region on chromosome III *(ARS306)* in wild-type and *mcm10-1*

backgrounds at the permissive (23°C), semi-permissive (27°C) and semi-restrictive (30°C) temperatures. Cells were grown to an OD_{600}=1 in YPDUA medium at the permissive temperature and then spotted as a 1:10 serial dilution on YPDUA plates and grown for 3 days. **(b)** Spot tests of strains containing either a no repeat cassette or the expansion $(GAA)_{100}$ repeat cassette located at a non-essential region on chromosome V (*ARS507*). Assay conducted as in (a). **(c)** Viability assay of strains containing no repeats or repeats of various lengths and compositions. The same number of cells was plated in duplicates at the semi-restrictive temperature (30°C) and at the permissive temperature (23°C) and grown for 3 days. Each datapoint represents the percent relative survival as determined by colony forming unit (CFU) counts. Each assay was conducted using at least two biological replicates. Plotted value indicated the mean and the bars indicate standard deviation. Statistical analysis was performed using unpaired t-test, *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$. **(d)** Viability assays for strains containing the expansion cassette represented in **(Figure 2-2 b)** with an inverted repeat, in which $(TTC)_{100}$ is the lagging-strand template. Assay conducted as in (c).

Viability defects in *mcm10-1* $(GAA)_{100}$ cells are exacerbated in DNA repair and Polymerase δ mutants

We conducted candidate gene analysis to decipher which cellular pathways contribute to the observed viability defect in the *mcm10-1* strains carrying the $(GAA)_{100}$ repeats. Rad51 and Rad52 are both involved in the homologous recombination and template switching (TS) processes [294]. We observed a minor exacerbation of the viability defect in *mcm10-1 rad51Δ* and a much stronger effect in the *mcm10-1 rad52Δ* double mutant **(Figure 2-10)**, with the latter being already pronounced even at permissive temperatures **Figure 6-1**. Template switching is also mediated by Rad5, a polyubiquitin ligase and DNA-dependent ATPase. The *mcm10-1 rad5Δ* double mutant displayed an exacerbated growth defect, particularly at the semi-permissive temperature of 27°C (**Figure 2-10 and Figure 6-1),** even though it did not further reduce viability in the colony formation assay **(Figure 6-2)**. Preventing DSBs and gap resection by knocking out the Exo1 nuclease in the *mcm10-1* background led to a partial rescue in growth in

the spot test, even though it did not enhance viability as measured by colony forming units **(Figure 2-10 and Figure 6-2)**.

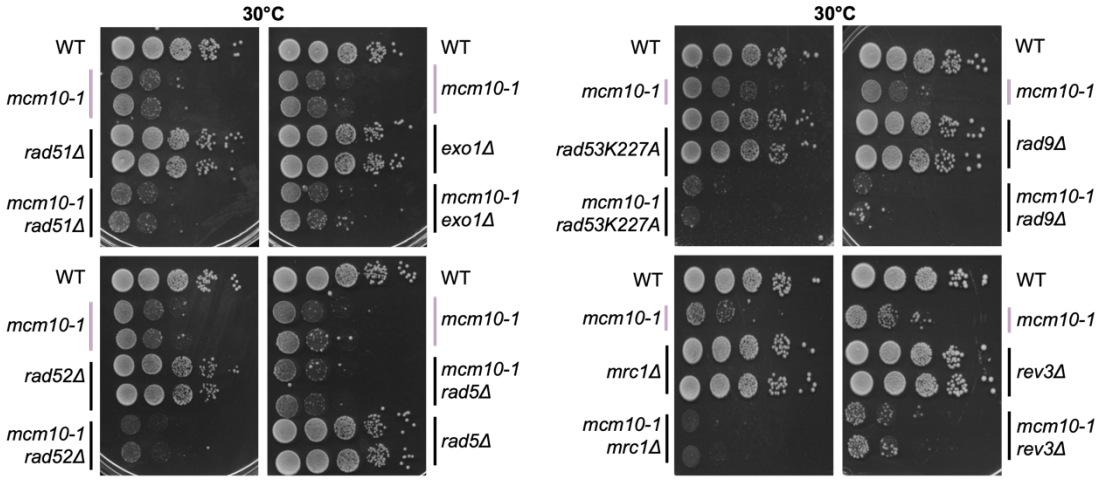

Figure 2-10: Spot tests for various double mutants between DNA repair and checkpoint genes and *mcm10-1*.

Cells were grown to an OD_{600}=1 in YPDUA medium at the permissive temperature (23°C) and then spotted as a 1:10 serial dilution on YPDUA plates and 2 days at 30°C.

Which DNA polymerase could be responsible for processing replication defects in the *mcm10-1* mutants with $(GAA)_n$ repeats, and is thus facilitating cell survival? Knocking out the translesion synthesis polymerase Rev3 only had a minor effect on strain viability **(Figure 2-10)**. We then turned to determining the role of DNA polymerase δ, which is implicated in replication stress survival (reviewed in [295]). First, we tested the effect of the *pol3-Y708A* mutation in the catalytic subunit of Pol δ[296]. The *mcm10-1 pol3-Y708A* double mutant displayed poor growth at all temperatures, including the permissive temperature of 23°C **(Figure 2-11 a)**. Second, we placed Pol32, which is the processivity subunit of Pol δ and also works as part of the translesion polymerase ζ, under an inducible galactose promoter, creating *pGAL1-3xHA-POL32* strains [297]. Differences in protein levels between conditions were confirmed via western blotting **(Figure 2-11 c)**. Under repressive conditions (glucose), we observed synthetic lethality with the *mcm10-1*

mutation, which was not observed under galactose induction **(Figure 2-11 b)**. These data demonstrate that the lack of Pol32 protein renders *mcm10-1* strains carrying $(GAA)_{100}$ repeats practically inviable. We conclude that various DNA repair and post-replicative repair processes are needed to process the defects that arise in the *mcm10-1* mutants and maintain cell viability, but Pol δ function preserves genomic integrity of *mcm10-1* cells. Additionally, the pol3-Y708A mutation led to a partial rescue of repeat expansions in *mcm10-1*, indicating that synthesis by Pol δ is ultimately responsible for most $(GAA)_n$ expansion events.

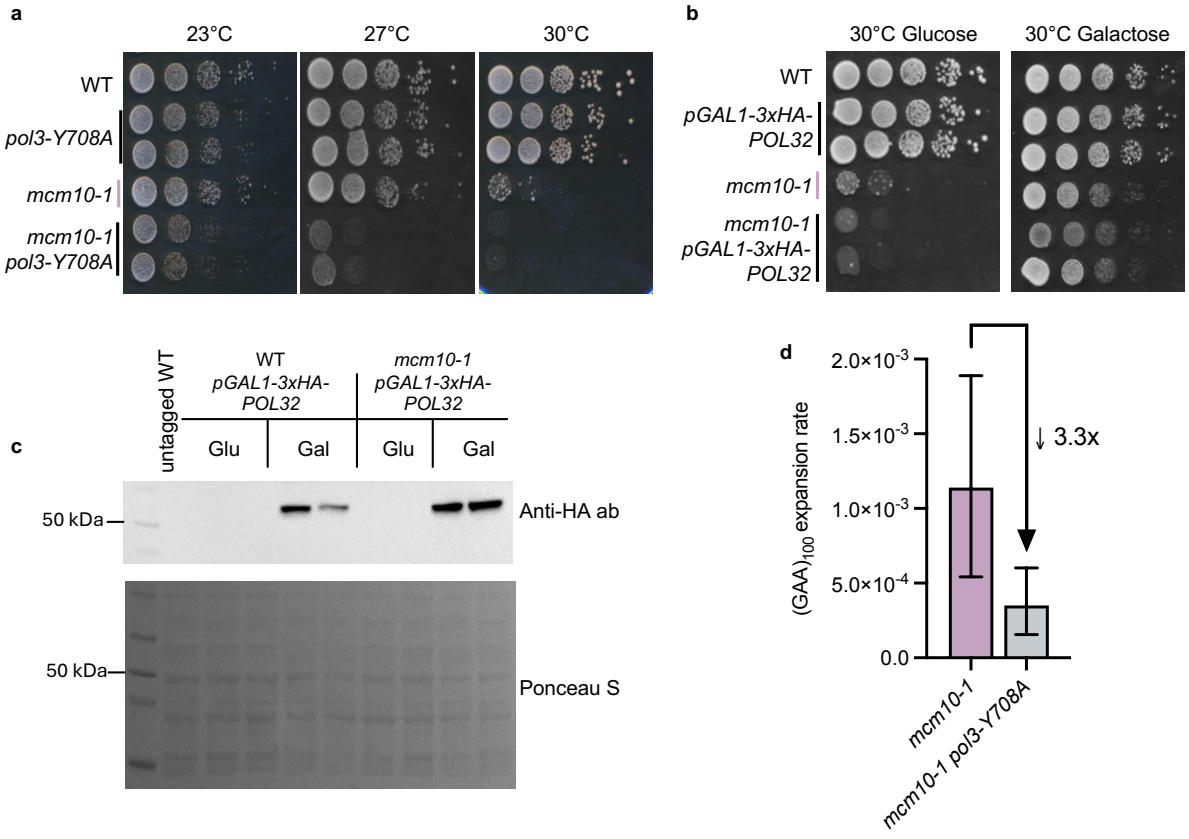

Figure 2-11: Effects of polymerase δ mutations on the viability of *mcm10-1* cells.

(a) Spot tests of strains containing the $(GAA)_{100}$ repeat with the *mcm10-1* and the *pol3-Y708A* mutant. **(b)** Spot tests for strains in which the endogenous POL32 promoter was substituted for inducible the *pGAL1* promoter to regulate POL32 expression level. For the *pGAL1-3xHA-POL32*

strains, cells were grown overnight in YPUA + 2% Raffinose and diluted and switched to media containing glucose or galactose until OD_{600}=1 and then spotted as a 1:10 serial dilution on YPDUA or YPGalUA plates and grown for 2-3 days at 30°C. **(c)** Western blot of the *pGAL1-3xHA-POL32* strains to detect Pol32 protein levels under repressive (glucose) and inducting (galactose) conditions at 23°C. *pGAL1-3xHA-POL32* was detected using Anti-HA Tag antibody 05-904 (Sigma Aldrich, 1 μg/ml dilution in 5% skim milk in TBS-T). **(d)** Expansion rates for the *mcm10-1 pol3-Y708A* double mutant. Plotted values indicate the corrected rate and the error bars represent 95% confidence intervals.

Mrc1 and Rad9 promote viability and expansions through different mechanisms

While we observed that many of the DNA damage tolerance and repair pathways are essential to promote the viability of *mcm10-1* cells, this does not necessarily mean that their absence would also exacerbate repeat expansion rates. The rate of repeat expansions remained unchanged in the *mcm10-1 rad51Δ* and *mcm10-1 rad52Δ* double mutant as compared to the *mcm10-1* mutant alone, while the *mcm10-1 rad5Δ* mutant led to a modest additional 3-fold increase in the repeat expansion rate **(Figure 2-12 a)**. Deletion of either the Exo1 exonuclease or the Rev3 TLS polymerase also did not affect the expansion rate of the *mcm10-1* mutant **(Figure 2-12 a)**.

We reasoned that a common consequence of excessive ssDNA exposure, as is seen in the *mcm10-1* mutant [274], is the activation of checkpoint pathways via the recruitment of various mediators and kinases, culminating with the activation of the effector kinase $Rad53^{CHK2}$ [298]. The DNA replication checkpoint (DRC) recognizes ssDNA at stalled replication forks mainly using the $Mec1^{ATR}$-$Mrc1^{Claspin}$ axis. The intra-S phase DNA damage checkpoint (DDC) is also triggered by persistent replication stress and accumulation of post-replicative ssDNA gaps is mediated by $Rad9^{TP53BP1}$ [299–302]. In addition, the checkpoint is directly activated in response to double-strand breaks via $Tel1^{ATM}$- $Rad9^{53BP1}$ (reviewed in [303–305]). Activation of the checkpoint results in fork stabilization, slower replication and recruitment of additional factors for repair and restart (reviewed in [306], [307]).

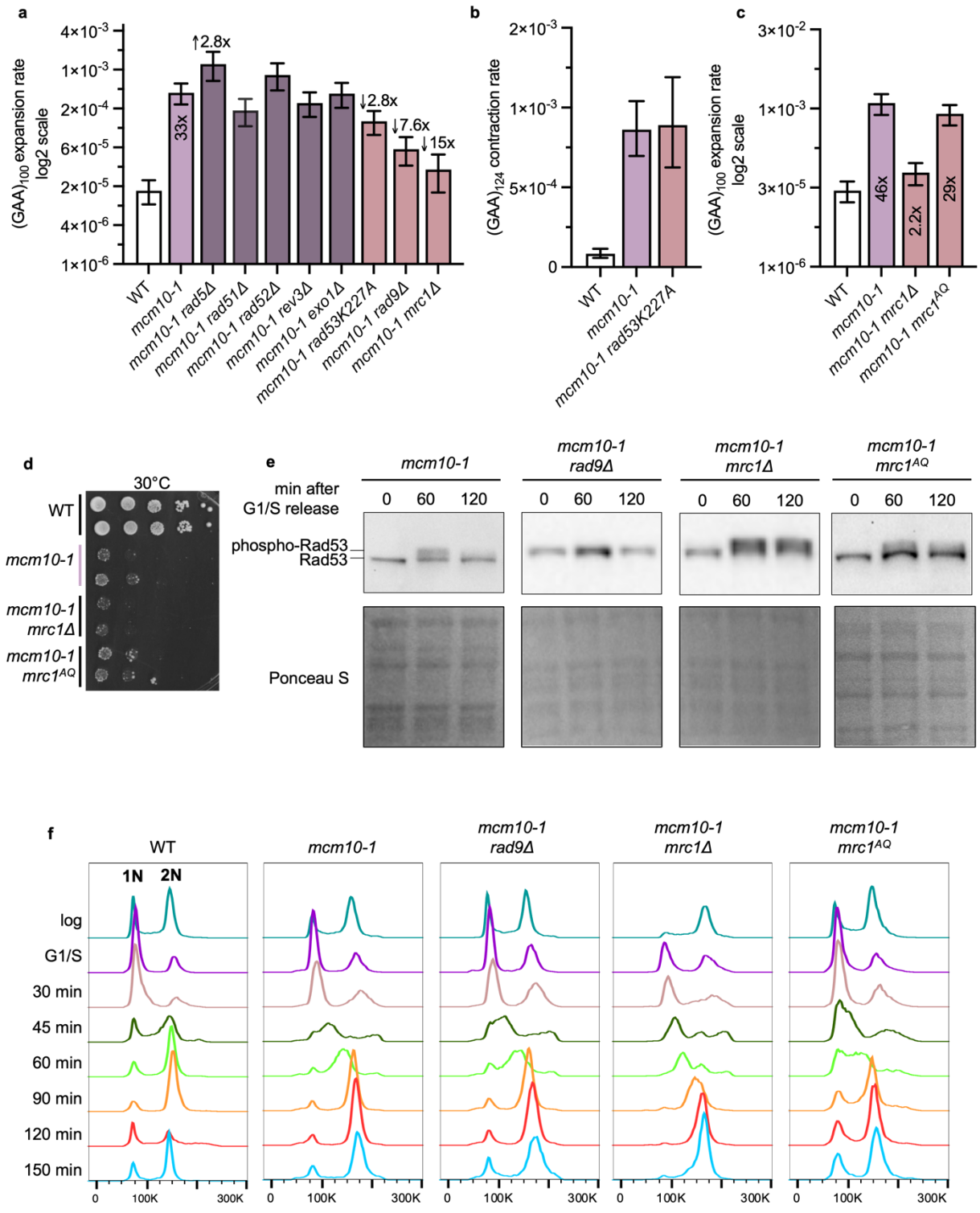

Figure 2-12: The Rad9 checkpoint and the Mrc1 replication function promote expansions in *mcm10-1*.

(a) Expansion rates in various *mcm10-1* double mutants with genes involved in response to replication stress and DNA repair at the semi-permissive temperature (27°C). Numbers within bars represent the fold increase relative to the corresponding wild-type value. Numbers above bars represent fold increase over the single *mcm10-1* mutant. **(b)** Contraction rates for the *mcm10-1 rad53K227A* double mutant deficient in checkpoint activation at 27°C. **(c)** Expansion rates at 27°C using the separation of function $mrc1^{AQ}$ mutant in *mcm10-1*, which is deficient in the checkpoint function of Mrc1 and proficient in its replication function. The *mcm10-1* $mrc1^{AQ}$ strain contains the pAO139 plasmid. All other strains on this graph are the same as in Fig. 6a but supplemented with the empty vector pRS415 for direct comparison. Non-selective stage was conducted on -LEU media to maintain the plasmids. **(d)** Spot tests for the strains used in Fig. 6c. Cells were grown to an OD_{600}=1 in -LEU at the permissive temperature (23°C), then spotted as a 1:10 serial dilution on -LEU plates and grown for 2 days at 30°C. **(e)** Western blots of Rad53 to detect phosphorylation status in *mcm10-1* and various double mutants. Cells were grown at 23°C, arrested in G1/S using α-factor at 30°C and then released into the cell cycle. Samples were collected at the indicated time points and the protein detected as described in the Methods. Ponceau S staining was used as a loading control. **(f)** Cell cycle analysis of *mcm10-1* and various double mutants. Cells were grown at 23°C, arrested in G1/S using α-factor at 30°C and then released into the cell cycle. Cells were collected at the indicated timepoints and analyzed by flow cytometry using the SYTOX™ green DNA stain. 1N indicates G1 cells, 2N indicates G2/M cells after completion of S-phase. For a-c, the plotted values indicate the corrected rate and the error bars represent 95% confidence intervals. Numbers within bars indicate fold increase over the respective wild-type value. Numbers with arrows above bars indicates the fold change from the single *mcm10-1* mutant.

We studied the effect of key checkpoint regulators on the viability and repeat instability of *mcm10-1* strains containing our experimental cassettes. The *rad53K227A* mutation affects the kinase activity of the Rad53 protein without compromising its role in maintaining normal dNTP levels via Dun1 phosphorylation [308]. Both the *rad53K227A* mutation and *RAD9* deletion caused a profound additional viability defect in *mcm10-1*, which was evident even at the permissive temperature **(Figure 2-10 and Figure 6-1)**. We then tested how checkpoint defects affect the stability of the $(GAA)_n$ repeat. Both the *rad53K227A* and *rad9Δ* mutations significantly rescued hyper-expansion phenotype of the *mcm10-1* mutant **(Figure 2-12 a)**. In contrast, the *mcm10-1 rad53K227A* mutant did not rescue $(GAA)_{124}$ contractions, further indicating that different mechanisms account for expansions and contractions in Mcm10-deficient cells **(Figure 2-12 b)**. Checkpoint activation is commonly reflected by changes in the cell cycle profile. We conducted

cell cycle profiling of WT and *mcm10-1* strains containing the $(GAA)_{100}$ expansion cassette using flow cytometry of DNA content. We observed that while the cell cycle profile of WT and *mcm10-1* strains is virtually identical at the permissive temperature of 23°C, at 30°C there is a delay in S-phase progression in the *mcm10-1* mutant, with the bulk of replication occurring 60 min after release instead of 30-45 min after release **(Figure 2-12 f and Figure 6-3)**, as was previously observed[274]. To directly relate cell-cycle stage with checkpoint activation, we conducted western blots using an antibody against the checkpoint effector kinase Rad53, which is phosphorylated upon checkpoint activation. The checkpoint was indeed activated in the *mcm10-1* mutant during late S-phase (60 min), as can be observed by Rad53 phosphorylation, but no checkpoint activation could be detected during G2/M phase (120 min) **(Figure 2-12 f)**. No checkpoint activation was detected in the *mcm10-1 rad9Δ* strains during late S-phase **(Figure 2-12 e and f)**. In summary, we observed that the Rad9 checkpoint senses the defects in *mcm10-1* likely during late replication and post-replicatively and initiates a response that ensures cell viability. During this response, however, $(GAA)_{100}$ repeat expansions can occur.

Mrc1 functions as an intra S-phase checkpoint mediator of replication stress but also works together with $Tof1^{TIM}$ and $Csm3^{TIPIN}$ to form the so called Fork Protection Complex (FPC), which regulates CMG helicase activity, limiting replication fork uncoupling and ssDNA exposure (reviewed in [245]). In the *mcm10-1 mrc1Δ* double mutant, we observed a massive rescue in repeat expansions **(Figure 2-12 a and c)** as well as a synthetic viability defect with *mcm10-1* **(Figure 2-10 and Figure 2-12 d**). We then used a separation of function mutant to discern which function of Mrc1 is responsible for these phenotypes. In the $Mrc1^{AQ}$ mutant, the replication function of Mrc1 is intact while its checkpoint function is compromised [309]. We introduced a plasmid containing the $mrc1^{AQ}$ allele in our *mcm10-1 mrc1Δ* strains and observed no rescue in

$(GAA)_{100}$ repeat expansions compared to the empty vector control **(Figure 2-12 c)**. Thus, Mrc1's replication function promotes repeat expansions in the *mcm10-1* mutant.

Furthermore, $Mrc1^{AQ}$ expression rescued the viability defect back to the *mcm10-1* single mutant levels **(Figure 2-12 d)**. In the *mcm10-1 mrc1Δ* double mutant, we observed both a delay in cell cycle progression **(Figure 2-12 f)** and checkpoint activation during both S-phase and G2/M **(Figure 2-12 e)**, as was previously shown for the *mrc1Δ* single mutant[309]. In contrast, the *mcm10-1 mrc1^{AQ}* mutant displays checkpoint activation only in late S-phase, as did the *mcm10-1* single mutant **(Figure 2-12 e)**, but does not fully rescue the S-phase delay observed in the *mcm10-1 mrc1Δ* context **(Figure 2-12 f)**. Therefore, we conclude that the replication function of Mrc1, but not its checkpoint function, is required for the viability of the *mcm10-1* strains carrying $(GAA)_n$ repeats.

The *mcm10-1* mutation increases fragility at the $(GAA)_{100}$ repeat

Expanded $(GAA)_n$ repeats induce double-stranded breaks (DSBs) in an orientation- and length-dependent manner [55,65]. In addition, Mcm10 deficiencies can generally lead to increased DSB levels [256,274,310]. We examined whether the Mcm10 protein plays a role in counteracting fragility at the repeat by conducting an arm loss assay. A $(GAA)_{100}$ repeat or a control sequence were placed within a split *URA3* gene on the non-essential arm of chromosome V, in a manner centromere-proximal to the *CAN1* gene and with the $(GAA)_{100}$ on the lagging strand template **(Figure 2-13 a)**. This is the orientation that has previously been associated with higher fragility levels at the repeat [55]. The presence of the repeat led to an 3.8-fold increase over the no repeat control and was consistent with previously published results [122]. The *mcm10-1* mutation led to an increase of over one order of magnitude at the semi-permissive temperature for both the no-repeat control and the $(GAA)_{100}$ cassettes **(Figure 2-13 b)**. As was the case for expansions,

fragility was not rescued by RPA overexpression in *mcm10-1* strains, but was rescued when the CMG helicase was additionally mutated (*mcm10-1 mcm2G400D*) **(Figure 2-13 b)**. We conclude that the observed increase in fragility observed in the *mcm10-1* mutant is repeat-independent but also caused by the compromised interaction with the CMG helicase.

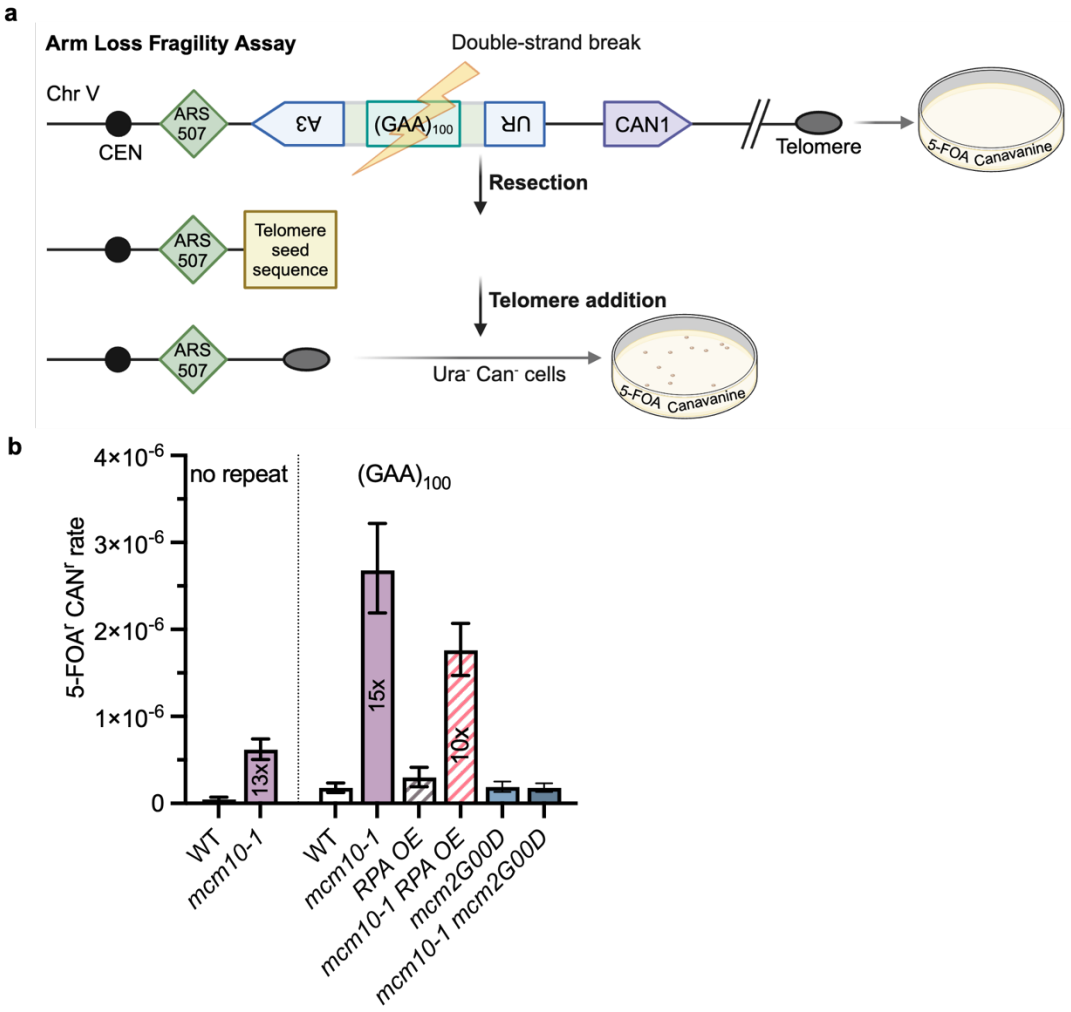

Figure 2-13: The *mcm10-1* mutation leads to repeat-independent increases in chromosomal fragility.

(a) Assay system to measure arm loss rates as a proxy for chromosomal fragility on Chromosome V. Created in BioRender.com. **(b)** 5-FOA^r CAN^r arm loss rates per cell per generation for strains containing either a no repeat control or the $(GAA)_{100}$ repeat at the semi-permissive temperature (27°C). Plotted value indicates the corrected rate calculated with FluCalc (https://flucalc.ase.tufts.edu/) [280] and the error bars represent 95% confidence intervals. Numbers within bars represent fold increases relative to the corresponding wild-type value. Raw data can be found in **Table 9**.

DISCUSSION

In this study, we investigated the role of the Mcm10 protein and other replisome factors in $(GAA)_n$ repeat instability. Mcm10 is a highly conserved protein that promotes DNA origin unwinding during replication initiation [253,311–314]. Several studies implicated Mcm10 in replication elongation, either as a replisome component or its accessory factor [272–274]. In addition, deficiencies in Mcm10 have been shown to lead to various types of genome instability and dysregulated DNA damage response [275,282,292,315], making it an attractive candidate to study in the context of repeat instability. We found that a temperature sensitive mutation in the Mcm10 protein (*mcm10-1*) dramatically increases length instability of $(GAA)_n$ repeats even at the semi-permissive temperatures **(Figure 2-3)**, at which other studies found no noticeable mutational phenotypes [253,274,275].

There are several mechanisms that could account for the elevated $(GAA)_n$ repeat instability in Mcm10 deficient cells. One possibility is that it could result from the degradation of Pol α that is known to occur in both yeast and human cells upon Mcm10 depletion [256,257,274,284,316]. Several of our results, however, argue against this possibility (**Figure 2-5**). First, we observed no loss of the catalytic subunit of Pol α, Pol1/Cdc17 at the semi-permissive temperatures. Second, a previous study showed that Pol α mutants resulted in the addition of a greater number of repeat units during expansions, but not an increase in the rate of $(GAA)_{100}$ repeat expansion, which is the opposite of what we observe in the *mcm10-1* mutant [83]. Third, Pol α mutations led to a strong increase in $(GAA)_{124}$ repeat contraction in a study conducted in our laboratory [64], while we observe that the *mcm10-1* effect on expansions is much greater than its effect on contractions. Fourth, an additional Mcm10 mutation that has been associated with low

levels of Pol1/Cdc17 but does not affect Mcm10 levels *(mcm10-G261D)* did not increase either $(GAA)_n$ repeat contractions or expansions **(Figure 2-6)**. Altogether, these results convincingly indicate that lower Polymerase α-primase complex levels are not responsible for the strong increase in instability that we observed in the *mcm10-1* mutant.

Another possibility is that repeat instability could primarily result from replication fork uncoupling between the leading and lagging strand syntheses, since Mcm10 contributes to fork integrity by bridging the CMG helicase, DNA Pol α and DNA Pol ε [253,272,317,318]. Uncoupling occurs when the helicase continues to unwind while leading strand synthesis by polymerase ε is halted, and lagging-strand synthesis continues [243]. We therefore compared the effects on repeat instability between the *mcm10-1* mutation and hydroxyurea (HU) treatment, which causes replication fork uncoupling (92–94). We found that HU treatment leads to an overall increase in $(GAA)_n$ repeat instability, with a much stronger effect on repeat contractions **(Figure 2-8 a and b)**. We acknowledge, however, that hydroxyurea treatment has pleiotropic effects on replication, as it also affects origin firing, checkpoint activation, and the function of Pol δ in lagging strand synthesis [295,319]. Thus, we cannot confidently attribute the observed effects on repeat instability to fork uncoupling alone. Along the same lines, deletion of the Ctf4 protein – another CMG helicase, DNA Pol α and DNA Pol ε adaptor – increases repeat contractions while leaving expansions unchanged **(Figure 2-8 c and d)**, and the *rfc1-1* mutation in the RFC complex, which was previously shown to increase small-scale expansion of $(CGG)_n$ repeats [320], also primarily increased $(GAA)_n$ repeat contractions **(Figure 2-8 e and f)**. Altogether, we conclude that replication fork uncoupling leading to the accumulation of single-stranded repetitive DNA combined with the massive RPA depletion in the *mcm10-1* mutant drives repeat contractions, which is in-line with our previous model [64].

How could Mcm10 counteract repeat expansions? Mcm10 is known to physically interact with and stabilize the CMG helicase via multiple MCM subunits [272,282], and it sits at the front edge of the CMG, which is optimal to face ssDNA/dsDNA junctions ahead of the fork [268,282]. Mcm10 regulates the ability of the CMG helicase to switch between an ssDNA and dsDNA encircling modes *in vitro* [268]. In the context of an uncoupled fork, the CMG-Mcm10 complex transitions to a dsDNA-encircling diffusive mode, which allows for fork re-entry and resuming of DNA synthesis upon encountering a lesion [268]. In addition, Mcm10 was shown to be important to promote the bypass of lagging strand blocks via its CMG isomerization functions [272]. To determine whether Mcm10 affects repeat instability via its interaction with the CMG helicase, we capitalized on a previous observation that Mcm10 and Mcm2 interact directly [272], and several mutations in Mcm2, including the *mcm2G400D* mutant, suppress the phenotypes of the *mcm10-1* allele [272,274]. We found that both types of elevated repeat instability in *mcm10-1* were rescued by the *mcm2G400D* mutation **(Figure 2-4)**.

We predicted that a consequence of perturbing the interaction between Mcm10 and the CMG helicase could be changes in replication fork speed through the $(GAA)_n$ repeat. Using live-cell microscopy, we indeed found that (i) there is a repeat-dependent slowdown of replication in the wild-type context, which is consistent with what we observed using 2D-gel electrophoresis of replication intermediates in the same orientation [65], and (ii) the *mcm10-1* mutation exacerbated this effect **(Figure 2-7)**. Since this assay system is located ~3 kb downstream of the origin, initiation should not contribute to the observed effects. These results indicate that Mcm10 promotes replication through the $(GAA)_n$ repeat and likely other hard-to-replicate genome regions, as was previously observed for replication termination in *Xenopus* egg extracts [321].

Increased fork stalling could result from more stalling events in the first place, or a difficulty of restart, which might be promoted by proper Mcm10-CMG interactions and be dependent on the ability of the CMG to switch from ssDNA to dsDNA encircling mode. Either of these events would lead to increased ssDNA exposure at the replication fork. While we observed a striking viability defect in Mcm10-deficient cells that contained expanded repeats at the essential chromosome III region, the viability defects were rescued by the *mcm2G400D* mutation as well as RPA overexpression, indicating that ssDNA accumulation is ultimately responsible for the viability defects **(Figure 2-9)**. Accumulation of single-stranded gaps combined with increased mutagenesis were indeed previously observed in *mcm10-1* mutants [292]. We therefore studied whether processes that are involved in the repair of ssDNA gaps, such as template switching (TS), are at play in our case. Rad5 knockout was previously shown to decrease the viability of *mcm10-1* mutants [292], implicating template switching and fork reversal in rescuing forks with Mcm10 function. In our case, deletion of the *RAD5* gene further impaired growth in the *mcm10-1* mutant. We identified a novel negative interaction between loss of Rad52, which is involved ssDNA gap filling in addition to homologous recombination [322], and Mcm10 deficiencies. Notably, none of the tested double mutants rescued the hyper-expansion phenotype. *RAD51* and *RAD52* deletion showed no change compared to *mcm10-1* alone, while *RAD5* knockout further increased repeat expansions. This indicates that while these pathways promote viability, they are not responsible for the elevated expansion rates observed in *mcm10-1*.

We reasoned that persistent exposure of ssDNA gaps in the *mcm10-1* mutant could activate the checkpoint and possibly also result in breakage at the fork. The checkpoint mediator Rad9 propagates checkpoint signaling during replication stress even in the presence of Mrc1 to promote repair processes at stalled forks, ssDNA gaps and DSBs [299,302,323]. Rad9 was previously

shown to be important for the survival of cells harboring a partial C-terminal deletion of Mcm10 [318]. A kinase-deficient Rad53 mutation as well as deletion the *RAD9* gene further impaired the survival of repeat-containing *mcm10-1* strains, pointing to persistent replication stress and checkpoint activation in the *mcm10-1* mutant, which we additionally confirmed by western blots showing the phosphorylated form of the Rad53 effector kinase [299]. In contrast to the DNA repair mutants described above, both *RAD9* knockout and kinase-deficient Rad53 led to a partial rescue of the hyper-expansion phenotype in *mcm10-1* mutants **(Figure 2-10 and Figure 2-12)**. This result suggests that the Rad9 checkpoint promotes the survival of the *mcm10-1* cells at the expense of triggering repeat expansions during DNA repair processes, which is very likely conducted by polymerase δ **(Figure 2-11)**. We also observed an increase in fragility at the $(GAA)_n$ repeat in the *mcm10-1* mutant, measured in the form of arm loss rates **(Figure 2-13)**. This is consistent with a proposed increase in DSBs in the *mcm10-1* mutant in general [256,275,285], and indeed we observe the same fold increase in fragility in the *mcm10-1* mutant with the control sequence, indicating that it is a repeat-independent effect. We would like to note that our arm loss assay system relies on the addition of a telomere at the available seed sequence following breakage to obtain a rate, and Mcm10 has been implicated in telomere maintenance in human cells [324,325]. Therefore, it might not be the best system to test the fragility of the *mcm10-1* mutant, and other systems such as Direct Duplication Recombination Assay (DDRA), might be a better assay to test it in to reach stronger conclusions [326].

The other checkpoint mediator, Mrc1, appeared to promote viability of Mcm10-deficient cells carrying the $(GAA)_{100}$ repeats via its function as part of the fork protection complex, rather than its checkpoint function **(Figure 2-12 d)**. This result is consistent with *in vitro* studies showing that in the absence of the replication function of Mrc1, Mcm10 becomes crucial for

proper progression of leading strand synthesis, and that Mrc1 and Mcm10 are partially redundant in the replication stress response[271,327,328]. We propose that in the absence of both Mcm10 and the replication-associated function of Mrc1, replication through DNA repeats becomes grossly inefficient, counteracting repeat expansions and resulting in cell death.

Overall, we propose a model in which Mcm10-deficiency causes replication problems such as increased fork stalling at the repeats and accumulation of ssDNA gaps genome-wide **(Figure 2-14 a)**. This leads to Rad9 checkpoint activation during late S-phase in a manner triggered by RPA binding to ssDNA gaps or conversion to double-stranded breaks **(Figure 2-14 b and c)**, in-line with previous studies describing a parallel role of Rad9 to Mrc1-sustained replication stress response [299,302]. Checkpoint activation results in the repair of under-replicated DNA, which is crucial for cell viability. Our data indicates that the massive repeat expansions we observed occur during the gap repair DNA fill-in synthesis conducted by DNA polymerase δ **(Figure 2-14 c)**. If ssDNA gaps remain unrepaired until the G2/M phase of the cell cycle due to deficient checkpoint activation or incomplete repair synthesis, under-replicated ssDNA would cause mitotic chromosome breakage, loss of essential DNA and ultimately cell death **(Figure 2-14 d)**. This model explains why RPA overexpression rescues cell viability by triggering checkpoint activation, while not rescuing hyper-expansion phenotype, which is downstream of checkpoint activation.

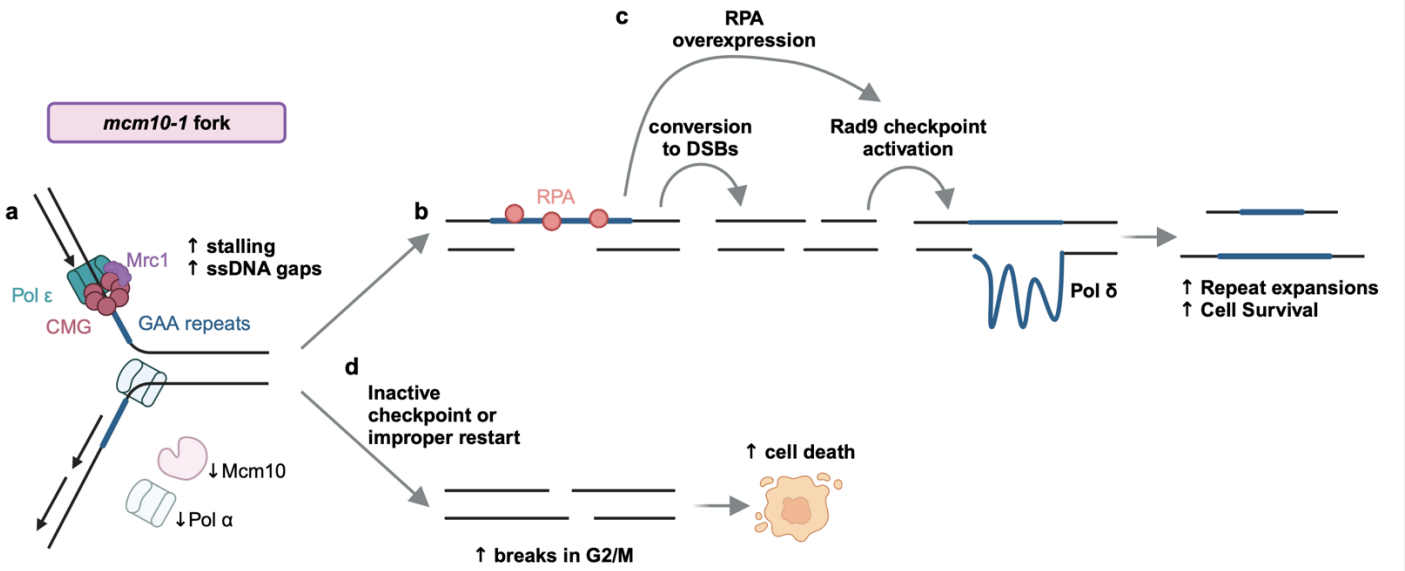

Figure 2-14: Model of the replication defect in the *mcm10-1* strain bearing the GAA repeat that leads to repeat instability and repeat-dependent loss of viability.

(a) Mcm10-deficiencies give rise to increased replication fork stalling at the expanded $(GAA)_n$ repeat, which results in ssDNA gap formation. The genome-wide increase in ssDNA levels can results in RPA depletion at the repeat, leading to **(b)** Rad9 checkpoint activation during late S-phase in a manner triggered by RPA binding to ssDNA gaps or conversion to double-stranded breaks. DNA repair and fill-in synthesis events lead to massive expansions during DNA synthesis conducted by polymerase δ, possibly due to reduced processivity and replication slippage. **(c)** This also occurs in the case of proper ssDNA coating upon RPA overexpression, minimizing conversions to DSBs. **(d)** In cases of insufficient checkpoint activation or incomplete synthesis, mitotic breakage can occur, leading to loss of essential DNA regions and cell death.

Our model partially agrees with previous observations of a replication fork block in *S. pombe* [329], where Rad51 and Rad52 knockouts led to a loss of viability due to accumulation of unrepaired ssDNA followed by chromosomal breakage. MCM10 was shown to suppress PRIMPOL-mediated gap formation in collaboration with BRCA2, suggesting that it could have a more general role in suppressing gap formation during DNA replication [315].

We found that Mcm10 deficiency destabilizes $(GAA)_n$ as well as $(GAGAAGAAA)_n$ repeats and cause repeat-mediated viability defect. This indicates that Mcm10 is vital for the maintenance of homopurine-homopyrimidine repeats regardless of their structure-forming potential. It is tempting, therefore, to speculate that Mcm10 could contribute to the maintenance of various other simple DNA repeats, which cumulatively account for ~3% of the human genome [330]. Indeed, *MCM10* haploinsufficiency in human cell lines led to chromosomal rearrangements overlapping with common fragile sites (CFSs) [324], which are overall rich in $(AT)_n$ repeats [331,332]. In addition, MCM10 promotes replication of telomeres [324], which are also highly repetitive DNA regions that can form alternative DNA structures. Further studies will be necessary to determine whether Mcm10 is also important for replication and maintenance of other disease-associated repetitive DNA sequences.

MATERIALS AND METHODS

All yeast strains used in this chapter are listed in **Table 2**. Point mutations were introduces using a CRISPR/Cas9[333] plasmids. Standard gene replacements were conducted using the pAG32 as template.

Oligonucleotides

All primers used for strain construction and to detect expansion events are listed in **Table 3**.

Fluctuation assays

All assays were conducted with at least two independent biological replicates per genotype. Mutation rates were calculated using the *FluCalc* software (https://flucalc.ase.tufts.edu/) [280].

Rates are considered meaningfully different if the 95% confidence intervals do not overlap. All rates can be found in **Table 7** and **Table 8**.

Expansion rates: Two independent biological replicates per genotype were plated for singles on YPDUA plates at 30°C (2 days), 27°C (2.5 days) or at 23°C (3 days). For the double mutants that displayed a severe growth defect, we incubated the non-selective plates at 27°C for up to 6 days, or until the colonies were as large as the wild-type colonies after 2.5 days. The initial repeat length was confirmed for 12 or more single colonies which were then serially diluted in 10^{-1} increments to a final dilution of 10^{-5}. 100 µl of the 10^{-5} dilution were plated on YPDUA to determine the total cell count (TCC). 100 µl of 10^{-1} or 10^{-2} dilution were plated on selective plates. For *ts* strains, the selective stage was conducted on 0.175% 5-FOA Glucose plates for 4 days at 23°C. For all other strains, the selective stage was conducted on 0.09% 5-FOA Glucose plates for 3 days at 30°C. Expansion events were confirmed by PCR using the A2 and B2 primer set and analyzed in ImageLab (BioRad).

Contraction rates: Fluctuation assays for contraction rates were conducted as described in (Khristich et al., 2020) using the appropriate temperatures at the non-selective stage and selective stages.

Arm loss rates (Fragility): Two independent biological replicates per genotype were plated for singles on YPDUA plates 5 days at 27°C or 6 days at 23°C. Repeat length was confirmed for 8-12 single colonies which were then serially diluted in 10^{-1} increments to a final dilution of 10^{-5}. 75 µl of the 10^{-5} dilution were plated on YPDUA to determine the total cell count (TCC) and grown for 3 days at 23°C. 150 µl of the undiluted sample were plated on selective plates (YC-Arg + 2% Glu + 0.1% 5-FOA + 0.006% Canavanine sulfate) and grown for 4 days at 23°C.

Viability Assays

Repeat tract length was determined for single colonies of each genotype grown at the permissive temperature of 23°C. Singles of the correct repeat length were grown to an OD_{600} of 0.6-1 in complete (YPDUA) media. An amount of 10^{-4} or 10^{-5} culture dilution corresponding to 100-500 cells was plated in duplicates on complete media in duplicates at both 23°C and 30°C. Colony forming units (CFUs) were counted after 3 days at 30°C and 4 days at 23°C. CFUs between duplicates were averaged and viability was calculated as (CFU 30°C)/(CFU 23°C) for each strain. Statistical differences were assayed using Welch's t-test in GraphPad Prism.

Spot tests

Yeast strains were grown overnight in YPDUA media at the permissive temperature of 23°C, diluted to an OD_{600}=0.2-0.3 and grown to log phase. *pGAL1-3xHA-POL32* strains, cells were grown in YPUA + 2% Raffinose overnight at 23°C and then diluted to OD_{600}=0.3 in YPUA media containing either 2% Glucose or 2% Galactose and grown until they reached an OD_{600}=0.8-1. 200 µl of OD_{600}=1 were collected and a serial 1:10 dilution was performed in sterile dH_2O. 5 µl of each dilution (1 to 10^{-4}) were spotted on YPDUA plates and placed at the temperatures indicated in the figure captions for 2-3 days and then imaged. Each spot test was conducted with at least two independent isolates of the same genotype.

Protein Isolation and Western Blotting

Total protein extraction was performed based on the protocol in [334]. For RPA overexpression blots, strains were grown to log phase at 23°C. For Pol1 protein measurements, cells were grown to OD_{600} 0.5-0.8 at 23°C and then shifted to 30°C or 37°C for 3 hours before sample collection. For measurements of *pGAL1-3xHA-POL32* expressions, cells were grown in YPUA + 2%

Raffinose overnight at 23°C and then diluted to OD_{600}=0.3 in YPUA media containing either 2% Glucose or 2% Galactose and grown until they reached an OD_{600}=0.8-1. 3-5 OD_{600} per sample were collected by centrifugation (2500 rpm, 5 min, 4°C) and the pellet was flash-frozen in liquid nitrogen. The pellet was then thawed and treated with 150 µl/OD denaturing lysis buffer (1.85 M NaOH, 7.5% β-mercaptoethanol) for 15 min on ice. The lysate was then mixed with 1 volume of 55% (w/v) trichloroacetic acid (TCA) for 15 min on ice. Proteins were then pelleted by centrifugation (13000 rpm, 4°C, 15 min) and washed with 1 volume of water (13000 rpm, 4°C, 15 min). The protein pellet was then resuspended in 50 µl/OD HU sample buffer (8 M urea, 5% SDS, 1 mM EDTA, 1.5% DTT, 1% bromophenol blue) and heated for 10 min at 65°C before running on 4-12% NuPAGE Bis-Tris gels in MOPS buffer for 90 min at 150V. We used Novex™ Sharp Pre-stained Protein Standard as a ladder. Transfer onto nitrocellulose membranes was conducted using an iBlot2 Gel Transfer device at 25 V for 7 min. RPA subunits were detected using the AS07-214 antibody (Agrisera, 1:5,000 dilution in 5% skim milk in TBS-T). Pol1-3xFLAG was detected using the ANTI-FLAG M2 antibody (Sigma Aldrich, 1:2,000 dilution in 5% skim milk in TBS-T). *pGAL1-3xHA-POL32* constructs were detected using Anti-HA Tag antibody 05-904 (Sigma Aldrich, 1 µg/ml dilution in 5% skim milk in TBS-T). Rad53 was detected using anti-Rad53 antibody EL7.E1 (Abcam, 1:2,000 dilution in 5% skim milk in TBS-T).

Cell Cycle Analysis by Flow Cytometry

Strains were grown overnight at 23°C in YPDUA at 200 rpm. Subsequently, cultures were diluted to OD_{600}=0.3 and grown for 150 min (23°C, 200 rpm). α-factor (Zymo Research, 10 mM) was added to a final concentration of 0.1 µM and cultures were grown for an additional 60 min (23°C, 200 rpm). ½ of the culture volume or the whole volume was then shifted to 30°C and

all cultures were grown for an additional 60 min (200 rpm) and cells were checked for shmoo formation with a light microscope. Cultures were then washed twice with 1 volume of YPDUA (2000 rpm, 2 min) and then resuspended in 1 volume of YPDUA containing Pronase E (Millipore Sigma # 537088) to a final concentration of 50 μg/ml and released at the respective temperatures.

0.5-2 x 10^7 cells were then collected at the indicated time points, washed with nuclease-free water and permeabilized with 1 ml 70% ethanol for overnight at 4°C. After centrifugation (5 min, 1500 rpm, 4°C), pellets were washed in 1 mL 50 mM Tris-HCl pH 7.5 and the cells were resuspended in 500 μL 50 mM Tris-HCl pH 7.5 containing 2 mg/ml RNaseA (Thermo Scientific, 10 mg/ml) and incubated overnight at 37°C. After centrifugation, the pellet was resuspended in 200 μL 50 mM Tris-HCl pH 7.5 containing 1 mg/ml Proteinase K (Thermo Scientific) and placed at 50 °C for 30 min. The cells were then pelleted by centrifugation (5 min, 5000 rpm, 4°C) and resuspended in 500 μL of 50 mM Tris-HCl pH 7.5 and stored at 4°C until measurements were acquired. Before analysis, 100 μL of the prepared sample were added to 1 ml 50 mM Tris-HCl pH 7.5 supplemented with 1 μM SYTOX™ Green Nucleic Acid Stain (Invitrogen). Data was acquired on an Attune^{TM} NxT machine using the Attune NxT software, and analyzed in FlowJo. We acquired 10,000 events on the ungated parameters.

Live-cell microscopy of replication fork progression

Yeast strains were grown overnight in synthetic complete (SC) medium at 23°C. Exponentially growing cultures were diluted to OD_{600}=0.2, 10 μg/mL α-factor was added to arrest cells at G1 phase and cultures were incubated for 1 hour at 23°C and then shifted to 30°C. Cells were then immobilized on microscopy chamber slides (Ibidi) coated with 2mg/mL concanavalin A (Sigma)

and washed thoroughly with warm SC medium to release the cells into S-phase. During imaging, cells were incubated in SC medium containing 4% glucose at 30°C. Live-cell imaging was performed on a CellDiscoverer 7 automated microscope (Zeiss) with an integrated LED light source, at 1 min intervals for 2 hours, using a 50x apochromatic water objective (NA=1.2) in 3D (8 z-sections 0.8 µm apart). Time-lapse measurements were collected using ZEN 3.0 software and analyzed using a custom-made, Python-based computational pipeline developed specifically for the analysis of replication rates, essentially as previously described [286]. Our pipeline identifies, tracks and quantifies the LacI-Envy and TetR-tdTomato dots in each cell. For each strain, at least 200 cells were measured in two independent experiments. Statistical analysis of replication rates was performed using Monte Carlo resampling with 1,000,000 iterations.

ACKNOWLEDGMENTS

The authors would like to thank the Van Deventer lab and Briana Lino for allowing us to use their flow cytometer for this study and members of the Mirkin and Freudenreich lab. We would like to thank Sarrah Hakimjee for her help with the *rfc1-1* experiments. pFA6a-TRP1-PGAL1 was a gift from John Pringle (Addgene plasmid # 41606). pAO139 (Apr LEU2 mrcl-AQ) was a gift from Stephen Elledge (Addgene plasmid # 41924). pRCC-N was a gift from Eckhard Boles (Addgene plasmid # 81192).

CRediT AUTHOR STATEMENT

Conceptualization and experimental design: C.M., S.M.M. and A.A. Data accumulation: C.M., Z.P., D.D, L.K.B., and S.M. Writing: C.M., S.M.M. All authors reviewed and edited the manuscript. Work in the S.M.M. lab was supported by NSF-BSF 2153071 and NIGMS R35GM130322 to S.M.M. Work in the A.A lab was supported by the Israeli Science foundation

(ISF) grant number 707/21 and the Binational Science Foundation (BSF-NSF) grant numbers 2019617 and 2021737 to A.A.

Table 2: List of yeast strains of Chapter 2.

Strain Name	Genotype	Comments
CM25	CM77 *mcm10-1 (P269L)*	*ts,* This study, CRISPR
CM26	CM77 *mcm10-1 (P269L)*	*ts, This study, CRISPR*
CM77	MATa leu2-Δ1, trp1-Δ63, ura3–52, his3–200 (CH1585), ade2::KanMX, MIP1-wt, HAP1-wt, ChrIII(75594-75641):: PGAL-UR-(GAA)128-A3-TRP1	PMID: 29447396
JAH112	MATa leu2-Δ1 trp1-Δ63 ura3-52 his3-200 (CH1585) bar1Δ::HIS3 ChrIII(75423-75715)::pUR-URA3-Int-GAA100d-TRP1	PMID: 23142667
CM71	MATa leu2-Δ1 trp1-Δ63 ura3-52 his3-200 (CH1585) bar1Δ::HIS3 ChrIII(75423-75715)::pUR-URA3-Int-GAA100d-TRP1	PMID: 23142667
CM72	CM71 *mcm10-1 (P269L)*	*ts,* This study, CRISPR
CM73	CM71 *mcm10-1 (P269L)*	*ts,* This study, CRISPR
CM85	CM71 *mcm2G400D*	This study, CRISPR
CM86	CM71 *mcm2G400D*	This study, CRISPR
CM88	CM72 *mcm2G400D*	This study, CRISPR
CM89	CM72 *mcm2G400D*	This study, CRISPR
CM59	CM77 *mcm2G400D*	This study, CRISPR
CM60	CM77 *mcm2G400D*	This study, CRISPR
CM69	CM25 *mcm2G400D*	This study, CRISPR
CM70	CM25 *mcm2G400D*	This study, CRISPR

CM195	CM25 *rad53K227A*	This study, CRISPR
CM197	CM25 *rad53K227A*	This study, CRISPR
CM200	CM72 *rad53K227A*	This study, CRISPR
CM201	CM72 *rad53K227A*	This study, CRISPR
CM221	CM71 rad5::HygB	This study
CM222	CM71 rad5::HygB	This study
CM218	CM72 rad5::HygB	This study
CM219	CM72 rad5::HygB	This study
CM227	CM71 *ctf4Δ*	This study, CRISPR
CM229	CM71 *ctf4Δ*	This study, CRISPR
SMYU170	CM77 ctf4::HygB	This study
CM148	CM77 *ctf4Δ*	This study, CRISPR
CM299	CM25 (pSK_UC57_2u_HIS3_RFA1_RFA2_RFA3)	This study
CM300	CM25 (pSK_UC57_2u_HIS3_RFA1_RFA2_RFA3)	This study
CM304	SMY706 (leu2Δ1, trp1Δ63, ura3Δ52, his3Δ200, MIP1, HAP1, MATa) ura3Δ bar1Δ + ChrIII(75423-75715)::pUR-UR-TTC100-A3-TRP1	This study
CM305	SMY706 (leu2Δ1, trp1Δ63, ura3Δ52, his3Δ200, MIP1, HAP1, MATa) ura3Δ bar1Δ + ChrIII(75423-75715)::pUR-UR-TTC100-A3-TRP1	This study
CM388	CM72 with CMP72 (pSK_UC57_2u_LEU2_RFA1_RFA2_RFA3)	This study
CM389	CM72 with CMP72 (pSK_UC57_2u_LEU2_RFA1_RFA2_RFA3)	This study, CRISPR
CM345	CM71 *rad51Δ*	This study, CRISPR
CM346	CM71 *rad51Δ*	This study, CRISPR
CM348	CM72 *rad51Δ*	This study, CRISPR

CM349	CM72 *rad51Δ*	This study, CRISPR
CM361	CM71 *rad52Δ*	This study, CRISPR
CM362	CM71 *rad52Δ*	This study, CRISPR
CM363	CM72 *rad52Δ*	This study, CRISPR
CM365	CM72 *rad52Δ*	This study, CRISPR
CM355	CM71 *exo1Δ*	This study, CRISPR
CM357	CM71 *exo1Δ*	This study, CRISPR
CM358	CM72 *exo1Δ*	This study, CRISPR
CM359	CM72 *exo1Δ*	This study, CRISPR
CM415	YEG150 *mcm10-1*	This study, CRISPR
CM416	YEG150 *mcm10-1*	This study, CRISPR
CM438	YEG200 *mcm10-1*	This study, CRISPR
CM439	YEG200 *mcm10-1*	This study, CRISPR
CM440	YEG200 *mcm10-1*	This study, CRISPR
CM315	CM304 *mcm10-1*	This study, CRISPR
CM316	CM305 *mcm10-1*	This study, CRISPR
YEG236/1	CM71 *rad53K227A*	PMID: 34849883
YEG236/2	CM71 *rad53K227A*	PMID: 34849883
YEG200/1	CM71 *mrc1::natMX*	PMID: 34849883
YEG200/3	CM71 *mrc1::natMX*	PMID: 34849883
YEG150/1	CM71 *rad9Δ*	PMID: 34849883
YEG150/2	CM71 *rad9Δ*	PMID: 34849883
CM153	CM72 GAA40	This study
CM154	CM72 GAA40	This study
CM155	CM71 GAA40	This study
CM156	CM71 GAA40	This study

SMYUD24	MATa, leu2-Δ1, trp1-Δ63, ura3-52, his3-200, ade2Δ::KanMX4, HAP1-wt, ChrIII(75594- 75641)::PGal1-UR-(GAGAAGAAA)41GAG-A3-TRP1	PMID: 31911468
SMYUD25	MATa, leu2-Δ1, trp1-Δ63, ura3-52, his3-200, ade2Δ::KanMX4, HAP1-wt, ChrIII(75594- 75641)::PGal1-UR-(GAGAAGAAA)41GAG-A3-TRP1	PMID: 31911468
CM281	SMYUD24 *mcm10-1*	This study, CRISPR
CM282	SMYUD24 mcm10-1	This study, CRISPR
JAH281	MATa leu2-Δ1, trp1-Δ63, ura3–52, his3–200 (CH1585), ade2::KanMX, MIP1-wt, HAP1-wt, ChrIII(75594-75641)::pUR-UR-no repeat control-A3-TRP1	This study
CM409	JAH281 *mcm10-1*	This study, CRISPR
CM410	JAH281 *mcm10-1*	This study, CRISPR
JAH293	MATa, leu2-1, trp1-63, his3-200, ura3Δ, bar1Δ, Chr. V arm loss centromere-facing no repeat control	This study
CM351	MATa, leu2-1, trp1-63, his3-200, ura3Δ, bar1Δ, ChrV(33,834-33,999)::UR-(GAA)100-A3 TRP1	This study
CM353	MATa, leu2-1, trp1-63, his3-200, ura3Δ, bar1Δ, ChrV(33,834-33,999)::UR-(GAA)100-A3 TRP1	This study
CM366	CM351 *mcm10-1*	This study, CRISPR
CM368	CM353 *mcm10-1*	This study, CRISPR
CM377	JAH293 *mcm10-1*	This study, CRISPR
CM378	JAH293 *mcm10-1*	This study, CRISPR
CM438	YEG200 *mcm10-1*	This study, CRISPR
CM439	YEG200 *mcm10-1*	This study, CRISPR
CM440	YEG200 *mcm10-1*	This study, CRISPR
CM441	CM71 *rev3::HygB*	This study

CM443	JAH112 *rev3::HygB*	This study
CM445	CM72 *rev3::HygB*	This study
CM447	CM73 *rev3::HygB*	This study
CM477	CM281 *mcm2G400D*	This study
CM478	CM282 *mcm2G400D*	This study
CM466	CM71 + pRS415	This study
CM467	CM71 + pRS415	This study
CM468	CM72 + pRS415	This study
CM469	CM72 + pRS415	This study
CM480	CM438 pRS415	This study
CM481	CM438 pRS415	This study
CM462	CM438 pAO139 (*mrc1AQ*)	This study
CM463	CM438 pAO139 (*mrc1AQ*)	This study
CM252	CM71 *Pol1-3xFLAG-KanMX*	This study
CM257	CM72 *Pol1-3xFLAG-KanMX*	This study
CM485	CM71 *pol3-Y708A*	This study
CM487	CM71 *pol3-Y708A*	This study
CM488	CM72 *pol3-y708A*	This study
CM490	CM72 *pol3-y708A*	This study
CM491	CM71 *kanMX-pGAL1-3xHA-POL32*	This study
CM492	CM71 *kanMX-pGAL1-3xHA-POL32*	This study
CM493	CM72 *kanMX-pGAL1-3xHA-POL32*	This study
CM494	CM72 *kanMX-pGAL1-3xHA-POL32*	This study
CM153	CM72 GAA40	This study
CM154	CM72 GAA45	This study
CM296	CM281 + *pSK_UC57_2u_HIS3_RFA1_RFA2_RFA3*	This study
CM297	CM281 + *pSK_UC57_2u_HIS3_RFA1_RFA2_RFA3*	This study

CM454	CM343 ChrIV(336,187)::URA3-Int-norpt-TRP1	This study
CM343	W1588, MATa, leu2-3,112, trp1-1, can1-100, ura3-1, ade2-1, chrIV(332,960)::LacOx128, chrIV(336,187)::NatMX, chrIV(352,560)::TetOx128, ade1Δ::LacI-Envy-KanMX-TetR-tdTomato	This study
CM425	CM343 ChrIV(336,187)::URA3-Int-GAA100d-TRP1	This study
CM471	CM454 *mcm10-1*	This study
CM436	CM425 *mcm10-1*	This study
CM371	CM71 + CMP72 (pSK_UC57_2u_LEU2_RFA1_RFA2_RFA3)	This study
CM372	CM71 + CMP73 (pSK_UC57_2u_LEU2_RFA1_RFA2_RFA3)	This study
CM373	CM71 + CMP74 (pSK_UC57_2u_LEU2_RFA1_RFA2_RFA3)	This study
CM315	CM304 *mcm10-1*	This study
CM316	CM305 *mcm10-1*	This study
CM317	CM305 *mcm10-1*	This study
CM163	CM77 *mcm10-G261D*	This study
CM164	CM77 *mcm10-G261D*	This study
CM240	CM71 *mcm10-G261D*	This study
CM241	CM71 *mcm10-G261D*	This study
CM137	CM71 *rfc1-1 (D513N)*	This study
CM138	CM71 *rfc1-1 (D513N)*	This study
CM143	CM77 *rfc1-1 (D513N)*	This study
CM144	CM77 *rfc1-1 (D513N)*	This study
CM351	SMY943, ChrV(33,834-33,999)::UR-(GAA)100-A3 TRP1	arm loss WT GAA100 strain, this study

CM353	SMY943, ChrV(33,834-33,999)::UR-(GAA)100-A3 TRP1	arm loss WT GAA100 strain, this study
CM366	CM351 *mcm10-1*	This study
CM368	CM353 *mcm10-1*	This study
JAH293	SMY943, ChrV(33,834-33,999)::UR-norpt-A3 TRP1	arm loss WT no repeat control strain, this study
JAH294	SMY943, ChrV(33,834-33,999)::UR-norpt-A3 TRP1	arm loss WT no repeat control strain, this study
CM377	JAH293 *mcm10-1*	This study
CM378	JAH293 *mcm10-1*	This study
CM390	CM353 + CMP72 (pSK_UC57_2u_LEU2_RFA1_RFA2_RFA3)	This study
CM393	CM353 + CMP72 (pSK_UC57_2u_LEU2_RFA1_RFA2_RFA3)	This study
CM391	CM366 + CMP72 (pSK_UC57_2u_LEU2_RFA1_RFA2_RFA3)	This study
CM392	CM366 + CMP72 (pSK_UC57_2u_LEU2_RFA1_RFA2_RFA3)	This study
CM418	CM351 *mcm2-G400D*	This study
CM420	CM351 *mcm2-G400D*	This study
CM422	CM366 *mcm2-G400D*	This study
CM423	CM366 *mcm2-G400D*	This study

Table 3: List of primers used in Chapter 2.

Name	**Sequence**	**Purpose**
pRCC plasmid amplification oligos to introduce CRISPR/Cas9 guides		
Mcm10_pRCC_F	GGTCTCCATGGTAGTACTTCgttttagagctagaaatagcaagttaaaataagg	

Mcm10_pRCC_R	GAAGTACTACCATGGAGACCcgatcatttatctttcactgcggag	*mcm10-1* (P269L) mutation
Mcm2_pRCC_F	GCTCCCGGAACCGTTCCTCCgttttagagctagaaatagcaagttaaaataagg	*mcm2G400D* mutation
Mcm2_pRCC_R	GGAGGAACGGTTCCGGGAGCcgatcatttatctttcactgcggag	
Rad53_pRCC_F	AACTACTGGGAAAACATTCGgttttagagctagaaatagcaagttaaaataagg	*rad53K227A* mutation
Rad53_pRCC_R	CGAATGTTTTCCCAGTAGTTtccgatcatttatctttcactgcggag	
CTF4_pRCC_F	AACAATAACGCGAACCGGGGgttttagagctagaaatagcaagttaaaataagg	CTF4 deletion
CTF4_pRCC_R	CCCCGGTTCGCGTTATTGTTcgatcatttatctttcactgcggag	
rad51_pRCC_F	TCAGCGGTTCGGATTAGACCgttttagagctagaaatagcaagttaaaataagg	RAD51 deletion
rad51_pRCC_R	GGTCTAATCCGAACCGCTGAcgatcatttatctttcactgcggag	
rad52_pRCC_F	ACCAGGTTCTTCGTCGAGTCgttttagagctagaaatagcaagttaaaataagg	RAD52 deletion
rad52_pRCC_R	GACTCGACGAAGAACCTGGTcgatcatttatctttcactgcggag	
Pol3_pRCC_F	TGGTTTTACAGGAGCGACGGgttttagagctagaaatagcaagttaaaataagg	*pol3-Y708A* mutation
Pol3_pRCC_R	CCGTCGCTCCTGTAAAACCAcgatcatttatctttcactgcggag	
Mcm10G261D_pRCC_F	AGATATTATAATCTTCGCCTgttttagagctagaaatagcaagttaaaataagg	*mcm10-G261D* mutation
Mcm10G261D_pRCC_R	AGGCGAAGATTATAATATCTcgatcatttatctttcactgcggag	
rfc1-1_pRCC_F	TACAAGAGGTGATATCCGCCgttttagagctagaaatagcaagttaaaataagg	*rfc1-1* mutation
rfc1-1_pRCC_R	GGCGGATATCACCTCTTGTAcgatcatttatctttcactgcggag	
CRISPR/Cas9 repair templates		
Mcm10-1_rep_temp	TAGAAAGATATTATAATCTTCGCCTGGGTGATGTGATAGCAATATTAAAtCTTGAAGTACTACCATGGAGACCCTCAGGGCGAGGAAATTTTATCAAATC	*mcm10-1* mutation

Mcm2-G400D_rep_temp	ATTATCAAAGGGTTACGCTCCAGGAAGCTCCCGGAACCGTTCCTCCAGACCGTCTACCAAGACATAGAGAAGTCATTTTGTTGGCGGATTTGGTAGATGT	*mcm2-G400D* mutation
Rad53_repairtemp_F	TTGCCACAGTAAAGAAAGCCATTGAAAGAACTACTGGGAAAACATTCGCCGTGGCGATTATAAGTAAACGCAAAGTAATAGGCAATATGGATGGTGTGAC	*Rad53-K227A* mutation
Rad53_repairtemp_R	GTCACACCATCCATATTGCCTATTACTTTGCGTTTACTTATAATCGCCACGGCGAATGTTTTCCCAGTAGTTCTTTCAATGGCTTTCTTTACTGTGGCAA	
CTF4del_rep_temp	AAATAATTGAGAAGGGCAAGAAGTGACGTAAATATACTAGACGTACTATTAAAAATGTAAAATATATATACGCAAGAGACAATTATTTGATACCTGTTCA	CTF4 deletion
rad51_repairtemp_F	CGACAAAGAGCAGACGTAGTTATTTGTTAAAGGCCTACTAATTTGTTATCGTCATGTATTTGGTCTCTTG	RAD51 deletion
rad51_repairtemp_R	AGAGGAGAATTGAAAGTAAACCTGTGTAAATAAATAGAGACAAGAGACCAAATACATGACGATAACAAAT	
rad52_repairtemp_F	TTAGTCTGTTAAGAAAAGACGAAAAATATAGCGGCGGGCGGGTTACGCGACCGGTATCGAAACGCTTCCTGGCCG	RAD52 deletion
rad52_repairtemp_R	AGGATTTTGGAGTAATAAATAATGATGCAAATTTTTTATTTGTTTCGGCCAGGAAGCGTTTCGATACCGGTCGCG	
Exo1_repairtemplate	ACCACATTAAAATAAAAGGAGCTCGAAAAAACTGAAAGGCGTAGAAAGGAAAGTTAAGTACTGCACGTTTCATATCGGAGGTATATTTTTCAAATGAAAA	EXO1 deletion
Pol3_Y708A_rep_temp	TTTTAAATGGTAGACAATTGGCTTTGAAGATTTCAGCTAACTCTGTCgcgGGTTTTACAGGAGCGACGGTtGGTAAATTGCCATGTTTAGCCATTTCTT	*pol3-Y708A* mutation

Mcm10G261D_rep_temp	ATATCTTCGGGAAAAAGGGTGTAGAAAGATATTATAATCTTCGCCTGgacGATGTGATAGCAATATTAAACCCAGAAGTACTACCATGGAGACCCTCAG	*mcm10-G261D* mutation
rfc1-1_rep_temp	TCAAACTTGATCCAAATGTCATTGATAGGTTGATACAGACTACAAGAGGTAATATCCGCCAAGTTATTAATCTACTTTCAACGATATCTACGACTACTA	*rfc1-1* mutation
Gene replacement primers		
Rev3_pAG_F	ATGTCGAGGGAGTCGAACGACACAATACAGAGCGATACGGTTAGATCATCCGGATCCCCGGGTTAATTAA	Rev3 deletion, used with pAG32
Rev3_pAG_R	GCGAGACATATCTGTGTCTAGATTACCAATCATTTAGAGATATTAATGCTGCATAGGCCACTAGTGGATCTG	
Gene tagging primers		
Pol1-PL3FLAG-KanMX_F	CGATTGTGGACGTCGCTACGTTGATATGACTAGCATATTTGATTTCATGCTAAATagggaacaaaagctggag	Pol1-3xFLAG-KanMX tagging
Pol1-PL3FLAG-KanMX_R	TTTTCTTACTATATAGAATATTCATGAGATCACACAACACATACAAAATACTTACctatagggcgaattgggt	
Pol32_F4	TTTTGAAAAGTACGCAGAAGTTCGTTACATCGCAATCAGATCAGCTCGAAgaattcgagctcgtttaaac	KanMX-pGAL1-3xHA-POL32 tagging
Pol32_R3	ACCTCAGTGAAGAGCTTCTCATTGATAAAATATGACGCCTTTTGATCCATgcactgagcagcgtaatctg	
Checking primers		
Mcm10_check_F	GCCTGAATACGCCAACTG	MCM10 amplification for sequencing
Mcm10_check_R	AATTCGAAGGTTGAAGGATTTG	
Mcm2_F	TTAGCGGAGTCCAAAGCCAT	MCM2 amplification for sequencing
Mcm2_R	TCACGTTCTTCTTCTTCAGTCCA	

CTF4_upstr_F	GCGACGCGTAATAAAGTTTCC	CTF4 external check
CTF4_dnstr_R	AAAATTCGCAAGGACGCTTC	
CTF4_F	GCCTGGAAAATCCAAGCGGAC	CTF4 ORF absence
CTF4_R	CACATCTTCCGCATTCGCTTC	
RAD51_chk_F	CAATTCGCAAGAAACGCACT	RAD51 ORF absence
RAD51_chk_R	AAGTAGTCATCGGGAAGAAGAGTA	
RAD51_upstr_F	TTTCTGTCCTGGTTTGTTTACAGTA	RAD51 external check
RAD52_chk_in_F	GCCAAGAAATCTGCCGTTAC	RAD52 ORF absence
RAD52_chk_in_R	TGAGCTTTCGCTGATTTCATCC	
Ex52_R	ACGTCGCTAAAGATGGTATGGTA	RAD52 external check
E52x-F	CTAGAGGATTTTGGAGTAATAAATAATGATG	
Exo1_F	CAGCGGGAGGGAAAACTGAT	EXO1 ORF absence
Exo1_R	CTCTGTTGGCTAGAGGTTGGTG	
exo1_KOchk_Fwd	GTATTACGTCCAAACTAAGTTCGCG	EXO1 external check
exo1_KOchk_rev	GACCGCTAGCGGCTTGATTAG	
Rad53_seq_F	TGGTGTAGGCGTGGAATCAG	RAD53 amplification for sequencing
Rad53_seq_R	TTTTGCCAGACCAAAGTCGG	
REV3_upstr_F	CGAGTGCAGTGCGTCTAGAAATAGTGT	REV3 external primers
Rev3_down_R	GGCGTTATTAATGCATCTGGGTCC	
JK183_hygRleft_rev	ACAGTCACATCATGCCCCTG	HYG internal primer for REV3 check
RFC1_in_chk_F	ATGCCGGTGTTAAAAACGCT	*rfc1-1* checking primers
RFC1_in_chk_R	CCATATGGCCGGCAACTTTT	
Repeat length primers		

A2	CTCGATGTGCAGAACCTGAAGCTTGATCT	Check repeat length of (GAA)100 strains
B2	GCTCGAGTGCAGACCTCAAATTCGATGA	
T1	GCCGCTCGAGTGAACCTGAAGCTTGATCT	Check repeat length of (TTC)100 strains
T2	AGCTGAGTGCAGACCTCAAATTCGATGA	
Live-cell microscopy strain construction		
cm343_int_F	CGCGTGATGTGAACATCAGCCCTTAGTTGAACTCTGTGGAC ACAAGACTAACGTCGCGGATCCGGAGATCTAGC	Inserting expansion cassettes in CM343 at the NAT locus
cm343_int_R	TAGCTCGTAGATAGATACTGACAGACTGATACTCTACGCTC ATACGATACCCGGGCGATAAGCTAGCAGG	
cm343_norpt_int	TAGCTCGTAGATAGATACTGACAGACTGATACTCTACGCTC ATACGATCCGGGCGATAAGCTAGGGCA	

CHAPTER 3 Deoxyribonucleotide pool availability and levels in the maintenance of $(GAA)_n$ repeat stability

Chiara Masnovo[1], Ayesha Lobo[1], Sergei M. Mirkin[1]

[1] Department of Biology, Tufts University, Medford, MA, USA 02155

ABSTRACT

Deoxyribonucleotides (dNTP) are the building blocks of DNA synthesis, and their cellular levels are regulated during the cell cycle and in response to stimuli such as DNA damage. Both high and low dNTP pools have been linked to genomic instability. We hypothesized that variations in dNTP levels would affect the stability of low-complexity regions, such as long $(GAA)_n$ repeats. In this chapter, we discovered that both an overall decrease and an increase in dNTP pool levels elevate $(GAA)_n$ repeat instability. While lower dNTP pools particularly promote repeat contraction, higher dNTP levels similarly affect both contractions and expansions. A specific mutation in the ribonucleotide reductase enzyme that increases dCTPs and dTTPs levels and decreases dATP and dGTP mildly increased expansions and had no effect on contractions. Collectively, these data indicate that imbalances in the dNTP levels, such as the ones observed in the context of several cancer states, can contribute to the instability of $(GAA)_n$ repeats.

INTRODUCTION

Deoxyribonucleotide (dNTP) pool levels are regulated in a cell-cycle dependent manner and the amount of dNTPs available during S-phase for DNA synthesis is tightly regulated to ensure a balance between complete genome replication and fidelity during synthesis. At the onset of S-phase in budding yeast, low dNTP levels transiently activate the Mec1-Rad53 checkpoint pathway, leading to activation of the Dun1 kinase and degradation of the ribonucleotide reductase (RNR) inhibitor Sml1, and therefore results in an increase in dNTP levels to support genome duplication [335–337]. RNR is the central enzyme complex involved in dNTP synthesis, as it catalyzes the rate-limiting step that converts ribonucleoside triphosphates to deoxyribonucleoside triphosphates **(Figure 3-1 a)**. RNR is regulated in multiple manners: transcriptional regulation, regulation through the cascade described above, and allosteric regulation of the large RNR subunits by NTPs/dNTPs themselves (reviewed in [338]).

Progression into S-phase with low dNTP levels, such as in the cause of hydroxyurea treatment, causes replication stress and DNA damage, particularly at hard-to-replicate regions such as chromosomal fragile sites[339] and non-canonical DNA secondary structures such as G4[340]. On the other hand, exceedingly high dNTP levels have also been associated with genome instability, cell cycle dysregulation, and as a potential driver of cancer [336,341,342]. This is mainly due to the reduction in fidelity of DNA polymerases and an increase in the usage of alternative polymerases such as translesion synthesis (TLS) polymerases, thereby increasing the genomic mutational load [343–345]. This notion was further corroborated by a study that showed that in the context of polymerase ε variants associated with tumorigenesis, their mutator phenotype was dependent on the availability of dNTPs, as deletion of the Dun1 kinase resulted in lower mutagenicity [346]. Nucleotide pool imbalances both give rise to replication stress and DNA

damage and regulate the cellular response to these signals. Generally, DNA damage leads to an increase in dNTP production to ensure timely DNA synthesis and repair [337,347] **(Figure 3-1 b)**. In mammalian cells, this does not seem to necessarily be the case [348,349].

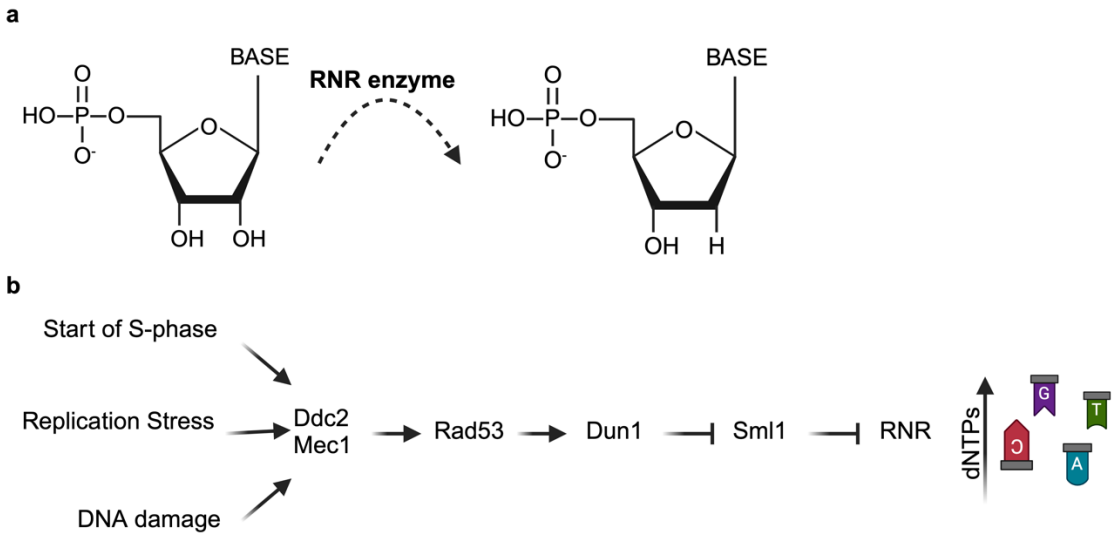

Figure 3-1: Mechanisms of dNTP pool regulation.

(a) The ribonucleotide reductase (RNR) enzyme converts ribonucleotides into deoxyribonucleotides. **(b)** Regulatory pathway in *S. cerevisiae* leading to an increase in dNTP production in response to stress signals via upregulation of RNR activity.

While dNTP pools can affect DNA replication and stress response globally, we and others hypothesized that these effects might be exacerbated in regions of low complexity, where there is a high local need of specific nucleotides. Long repetitive sequences are an example of these low-complexity regions, and nucleotide availability could therefore be a modulator of 1) fork stalling at repetitive sequences and 2) repeat instability during DNA replication [350,351]. For example, low nucleotide pools could leads to increased ssDNA gaps that promote the formation of secondary structures at the repeats and promote instability during synthesis if the polymerase

skips over a preformed structure or slips during replication, or by ribonucleotide excision repair due to an increase incorporation of ribonucleotides and DNA:RNA hybrids [352]. In the case of high nucleotide pools, translesion synthesis polymerases could lead to increased repeat instability and promote polymerase slippage. In the case of repetitive sequences, is also plausible that specific nucleotide imbalances could affect the two strands differentially.

In this chapter, we investigated the effect of nucleotide pool depletion, increase and imbalances on the instability of expanded $(GAA)_n$ repeats using three approaches. First, we studied the effects of HU-induced dNTP depletion on repeat stability. Next, we determined the role of Cyclin-dependent kinase inhibitor (CKI) *SIC1*, the functional homolog of $p27^{Kip1}$ in human cells [353], which regulates the transition from G1 to S-phase by repressing B-type cyclins [354]. Finally, we tested various point mutations in the Rnr1 subunit of the RNR complex. These mutations are known to cause specific nucleotide pool imbalances, including overall dNTP pool increase [355].

Overall, we find that nucleotide depletion increases repeat instability, with a stronger effect on contractions. On the other hand, increases in dNTP levels more strongly promotes $(GAA)_n$ repeat expansions. Future studies would be needed to determine the exact mechanism by which this occurs and whether the same would be observed in human cells.

RESULTS

Nucleotide depletion by hydroxyurea induces $(GAA)_n$ repeat instability

We first set out to test the effects of overall nucleotide pool depletion on the instability of expanded $(GAA)_n$ repeats using the instability reporters described in Chapter 2. A common agent used to induce replication stress is hydroxyurea (HU). HU inhibits the proton-coupled electron transfer step in in the RNR reaction, leading to an overall decrease of available dNTP pool levels. HU also inhibits enzymes containing Fe-S clusters, which includes all B-type polymerases (reviewed in [319]). It is important to note that not all nucleotides are affected equally by hydroxyurea, with stronger effects on purines[356].

As we presented in Chapter 2, exposure to 100 mM HU results in a 3.4-fold increase in repeat expansions and an 8.3-fold increase in repeat contractions when compared to the vehicle (DMSO) control (**Figure 3-2 b)**. To determine whether the instability is affected in a differential manner based on which sequence is on the leading or lagging strand, we also tested the inverted cassette for repeat contractions, in which $(TTC)_{124}$ serves as the template for lagging-strand synthesis, where we observed a comparable increase (10.7-fold) (**Figure 3-2 c)**. We also determined the general mutation rate by using the number bands of unchanged repeat size after the fluctuation assay as a proxy for URA3 mutations. Our rate for the vehicle treated control was comparable to the previously published mutation rate for the $(GAA)_{100}$ expansion cassette[83], and the rate was increased by 6.55-fold in the HU-treated condition **(Figure 3-2 d)**. Altogether, our results indicate that hydroxyurea treatment generally increases both types of repeat instability in a seemingly orientation-independent manner and suggests that low dNTP levels might be a driver of repeat instability.

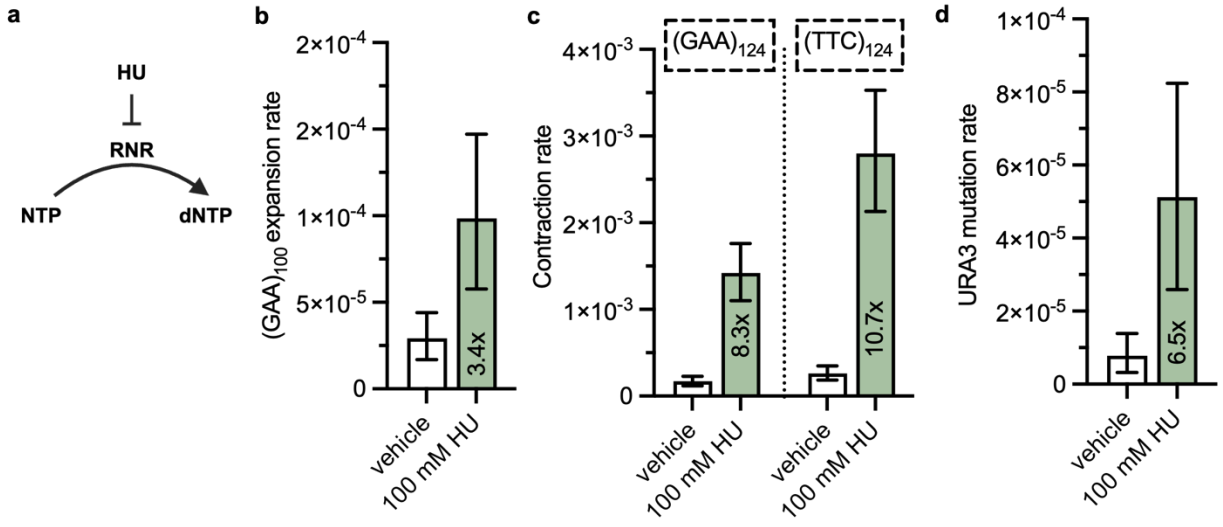

Figure 3-2: Effects on dNTP depletion caused by hydroxyurea treatment of $(GAA)_n$ repeat instability.

(a) Mechanism of nucleotide depletion by hydroxyurea (HU), which inhibits the ribonucleotide reductase (RNR) enzyme. **(b)** $(GAA)_{100}$ expansion rates under vehicle (DMSO) and 100 mM HU treatment conditions. **(c)** Contraction rates for $(GAA)_{124}$ and $(TTC)_{124}$ cassettes, where the indicated repeat corresponds to the lagging strand template sequence, under vehicle (DMSO) and 100 mM HU treatment conditions. **(d)** Rate of URA3 mutation under vehicle (DMSO) and 100 mM HU treatment conditions. Plotted values indicate the corrected rate and the error bars represent 95% confidence intervals. Numbers within bars indicate fold increase over the respective vehicle value. All experiments were conductive with a non-selective temperature of 30°C. All expansion and contraction data for this chapter can be found in **Table 7** and **Table 8**.

Defects in timing of S-phase entry increases $(GAA)_{124}$ repeat contractions

Right before S-phase, Sic1, which is an inhibitor of the Clb-Cdc28 kinases, is phosphorylated an degraded to allow for entry into S-phase and proper origin firing [353,357–359]. *SIC1* deletion causes uncontrolled entry into S-phase, resulting in perturbed origin firing and ultimately prolonged S-phase and an increase in chromosomal instability [358,360]. Deletion of *SIC1* did not have affect the 5-FOA rates in our expansion assay (**Figure 3-3 a)**. Note that we did not proceed with repeat PCRs to determine the actual expansion rate in the case of this mutant. On the other hand, the

sic1Δ strains showed a 2.37-fold increase in $(GAA)_{124}$ contraction rates (**Figure 3-3 b**). These results indicate that contraction levels might be more susceptible to perturbation of S-phase entry than expansion rates.

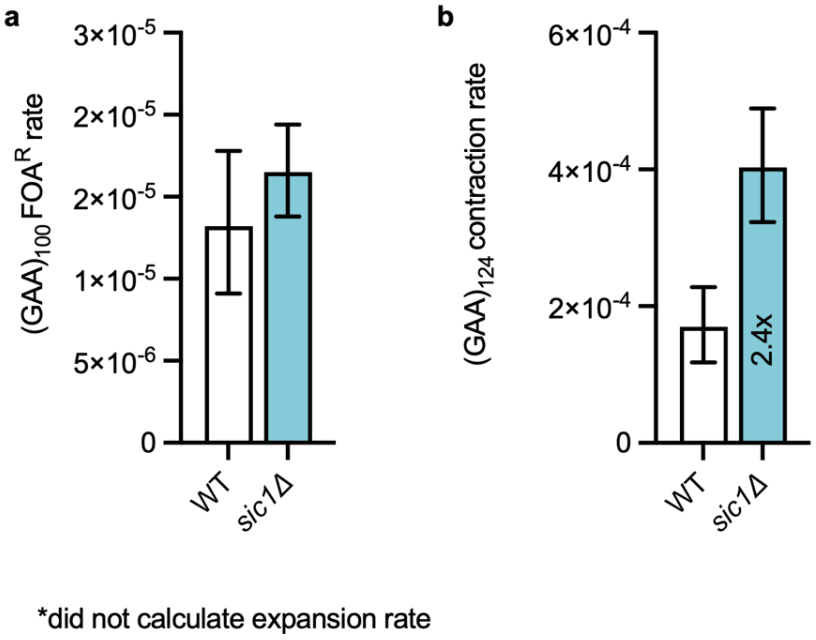

Figure 3-3: Effects of the deletion of the cell-cycle regulator Sic1 on $(GAA)_n$ repeat stability.

(a) 5-FOA resistance rates for the $(GAA)_{100}$ expansion assay. Actual expansion rate was not calculated via repeat PCR as we expected no effect based on the resistance rates. **(b)** $(GAA)_{124}$ contraction rates upon deletion of SIC1. Plotted values indicate the corrected rate and the error bars represent 95% confidence intervals. Numbers within bars indicate fold increase over the respective wild-type value. All experiments were conductive with a non-selective temperature of 30°C.

Changing the available dNTP pool levels by mutations in the ribonucleotide reductase subunit Rnr1 mildly affects $(GAA)_n$ repeat instability

dNTP pool imbalances can be achieved by mutating the RNR enzyme itself [355]. There are three major regulatory sites in the RNR enzyme: the "A-site" is the activity site, the "S-site" that

determines substrate specificity, and the "C-site" which is the catalytic core of the enzyme. Schmidt et al. performed a screen to identify mutations in the *RNR1* subunit of the RNR complex that displayed a mutator phenotypes and discovered that all of the identified alleles are linked to elevated or imbalanced dNTP pool levels. The mutant alleles were phenotypically categorized into four groups in a manner dependent on their interactions with additional genes, including polymerases and exonucleases. We selected a few of the mutations identified in this study to determine the role of specific nucleotide imbalances on the stability of $(GAA)_n$ repeats. A schematic representing the relative dNTP pool levels for the mutants we tested as measured in Schmidt et al. is provided in (**Figure 3-4**).

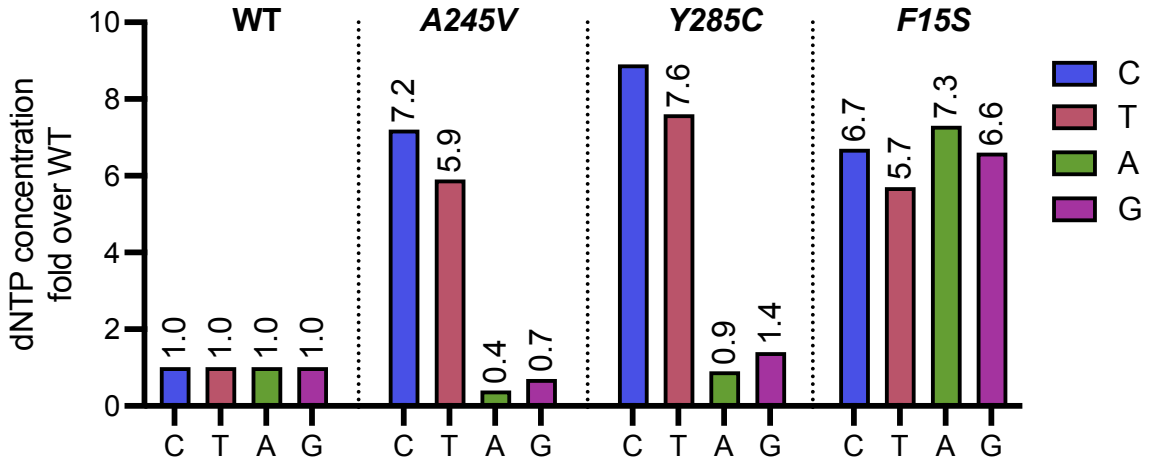

Figure 3-4: Relative dNTP pool levels compared to the wild-type levels.
Adapted from Schmidt et al. [355].

First, we tested the *rnr1-A245V* mutation primarily results in the depletion of purines (dATP and dGTP), which are on the lagging-strand template of the repeat, and a concomitant increase in pyrimidine (dTTP and dCTP) levels [355], which are on the leading strand template of

the repeat. We observed that the *rnr1-A245V* mutation does not affect $(GAA)_{124}$ repeat contractions, but leads to a modest but significant 2.5-fold increase in expansion rates (**Figure 3-5**).

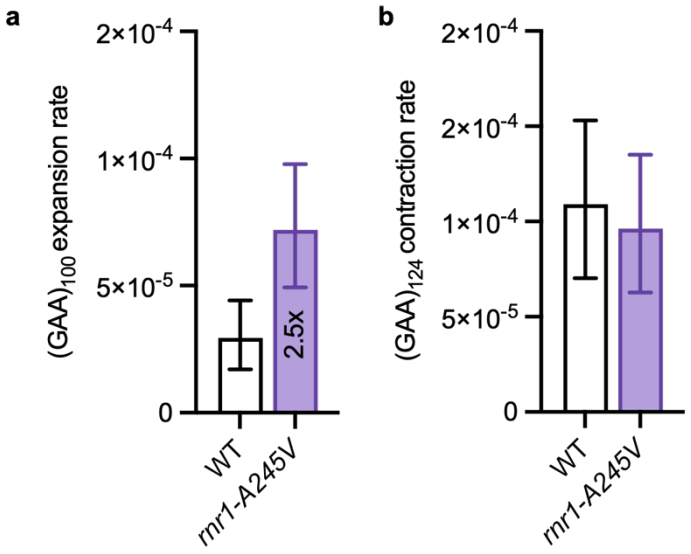

Figure 3-5: $(GAA)_n$ instability rates in the *rnr1-A245V* mutant.

(a) $(GAA)_{100}$ expansion rates for the *rnr1-A245V* mutant. **(b)** $GAA)_{124}$ contraction rates for the *rnr1-A245V* mutant. Plotted values indicate the corrected rate and the error bars represent 95% confidence intervals. Numbers within bars indicate fold increase over the respective wild-type value.All experiments were conductive with a non-selective temperature of 30°C.

The *rnr1-Y285C* mutation increases pyrimidine levels, while leaving purine levels unchanged. This mutation did not have an effect in our fluctuation assay with expansions strains (**Figure 3-6**). We note that we did not test the contraction rates for this mutant and it would be interesting to test this mutation in the future. This indicates that a mere increase in pyrimidine levels available for lagging strand synthesis does not affect the propensity of the repeat to expand.

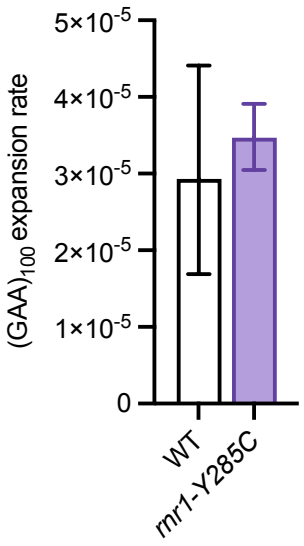

Figure 3-6: $(GAA)_{100}$ repeat expansions in the *rnr1-Y285C* mutant.

Plotted values indicate the corrected rate and the error bars represent 95% confidence intervals. All experiments were conductive with a non-selective temperature of 30°C.

Finally, we sought to determine what would happen in the case in which all nucleotide pools are increased, and can therefore mimic a circumstance of RNR complex upregulation as is observed in the context of DNA damage response. The *rnr1-F15S* mutation, located in the A-site, results in an increase of over 5-folds in the level of all four dNTPs (**Figure 3-4**). Interestingly, the *rnr1-F15S* mutation led to a 3.5-fold increase in expansion rates, as well as a mild 2-fold increase in contraction rates (**Figure 3-7**). This indicates that an overall increase in the nucleotide pool levels promotes an increase in $(GAA)_n$ repeat instability.

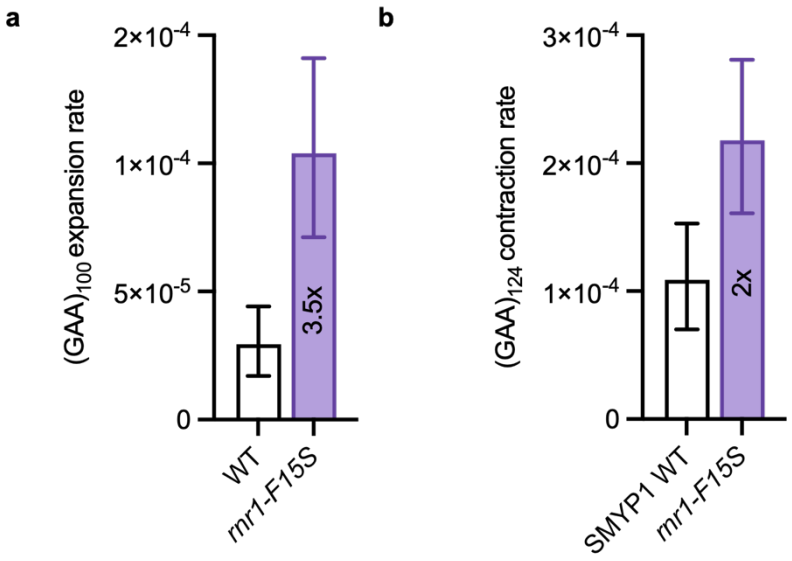

Figure 3-7: $(GAA)_n$ instability rates in the *rnr1-F15S* mutant.

(a) $(GAA)_{100}$ expansion rates for the *rnr1-F15S* mutant. **(b)** $GAA)_{124}$ contraction rates for the *rnr1-F15S* mutant. Plotted values indicate the corrected rate and the error bars represent 95% confidence intervals. Numbers within bars indicate fold increase over the respective wild-type value. All experiments were conductive with a non-selective temperature of 30°C.

DISCUSSION AND FUTURE DIRECTIONS

In this chapter, we aimed to determine whether deoxyribonucleotide (dNTP) levels can modulate $(GAA)_n$ repeat instability. There were several reasons why we hypothesized this could be the case. First, the level of available dNTPs can regulate replication fork speed as it influences the rate at which new nucleotides can by incorporated in the nascent strands [361]. Secondly, DNA repeats have an inherent sequence bias and low-complexity composition, which can create a high local need for specific dNTPs.

Overall, we have identified that a decrease of the overall nucleotide pool during S-phase as the one caused by HU treatment causes an increase in repeat instability, with a stronger effect

on repeat contractions, and this effect occurs independently of repeat orientation (**Figure 3-2**). This is consistent with our previous model published in Khristich et al. as well as the results presented in Chapter 2 [64]. Other studies have shown an increase in the instability of other repeats in HU treated cells: expanded $(CAG)_n$ repeats undergo increased expansions upon 100 mM HU treatment, whereas their contractions did not increase, indicating a that a different mechanism is responsible for the contraction events in the case of the CAG repeat [362]. A similar effect to that observed upon HU treatment was observed in the case of premature cell-cycle entry in the absence of the Sic1 protein (**Figure 3-3**), albeit to a much milder degree. We would like to emphasize that the nucleotide pool levels have not been determined by us or by others in the *sic1Δ* context and therefore we cannot conclude that the observed increase in contractions is due to changes in the dNTP pool in this case, but we hypothesize that premature cell-cycle entry would coincide with a temporary depletion of the nucleotide pool.

We also set out to study specific mutations in the Rnr1 subunit of the RNR complex, which result in different imbalances between nucleotides and have been related to mutator phenotypes, especially in the absence of the exonuclease Exo1 and mismatch repair [355]. Specific dNTP imbalances have been shown to affect the stability of telomeric sequences, which also have inherent low-complexity [363]. While we used strains derived from the same genetic background as the ones used in the original study that characterized the *rnr1* alleles[355], we did not directly measure the dNTP concentrations in our strains, and it is therefore possible that the specific nucleotide levels are slightly different in our study, which might affect the magnitude of the observed differences in our experiments. Nevertheless, we predict that the general trend of which nucleotide pools are decreased or increased in the mutants would be consistent with the

ones in Schmidt et al. and we therefore primarily attribute the observed differences to the changes in dNTP pool levels.

We tested the *rnr1-A245V* mutation, in which purine levels are increased and pyrimidine levels are decreased and we observed only a minor effect on $(GAA)_n$ repeat expansions. To determine whether this was due to the increase in purines or the decrease in pyrimidines, we turned to a different mutant, *rnr1-Y285C*, in which pyrimidine levels are comparably elevated, but purines are not depleted. In this case, we did not observe an increase in expansions, suggesting that the minor increase we observed in the *rnr1-A245V* mutant is caused by the depletion of dATPs and dGTPs, which constitute the newly synthesized leading strand. This indicates that there are strand specific effects of nucleotide imbalances on $(GAA)_n$ repeat instability. We were unable to obtain the *rnr1-Q288A* allele, which would have allowed us to test the effect of an increase in dATPs and dGTPs on the stability of the repeat [364].

Interestingly, we observed an increase in both types of instability when the dNTP pool was overall increased in the case of the *rnr1-F15S* mutation. This observation is particularly interesting because one of the cellular responses to DNA replication stress and DNA damage is Rad53 checkpoint activation through the Mec1-Rad53-Dun1 kinase cascade, which results in the degradation of the ribonucleotide reductase inhibitor Sml1, increasing dNTP pool levels to promote DNA repair, especially through the activity of more promiscuous polymerases such as translesion synthesis polymerases [365,366], and by promoting polymerase slippage. While this ensures genomic integrity, it can come at the cost of mutagenesis. It would be interesting to test whether removing the activity of these polymerases would rescue the increase in repeat instability that we observed. Similarly, we could directly test whether removing the RNR

inhibitor Sml1 results in an increase in repeat instability, since it also causes an increase in overall dNTP levels [367].

It is tempting to speculate that in contexts of elevated replication stress and DNA damage, changes to the dNTP pool could be an additional, underestimated modulator of DNA repeat stability. It is important to note, however, that dNTP pool regulation in human cells is quite different than the yeast pathways. In human cells, the dNTP triphosphohydrolase SAMHD1 is a major regulator of dNTP pools both during the normal cell cycle and in response to DNA damage, and mutations in *SAMHD1* have been associated with cancers [368–370]. It would be interesting to investigate whether mutations in SAMHD1 are associated with an increase in microsatellite instability.

MATERIALS AND METHODS

Yeast strains

All yeast strains used in this chapter are listed in **Table 4**. Point mutations were introduces using a CRISPR/Cas9[333] plasmid containing guides targeting *RNR1*, and repair templates were provided in the form of oligonucleotides. Sic1 was deleted by standard gene replacement using the pAG32 as template.

Oligonucleotides

All primers used for strain construction and to detect expansion events are listed in **Table *5***.

Fluctuation assays

All assays were conducted with at least two independent biological replicates per genotype. Mutation rates were calculated using the FluCalc software (https://flucalc.ase.tufts.edu/)[280]. Rates are considered meaningfully different if the 95% confidence intervals do not overlap. All raw rates are reported in **Table 7** and **Table 8**. For the HU experiment, we conducted the non-selective stage of the expansion or contraction assay on either YPDUA plates supplemented with 100 mM hydroxyurea (Sigma Aldrich) or the corresponding volume of DMSO (Sigma Aldrich).

Expansion rates. Two independent biological replicates per genotype were plated for singles on YPDUA plates at 30°C for 2 days. The initial repeat length was confirmed for 12 single colonies which were then serially diluted in 10^{-1} increments to a final dilution of 10^{-5} in sterile water. 100 µl of the 10^{-5} dilution were plated on YPDUA to determine the total cell count (TCC) for 2 days at 30°C. 100 µl of 10^{-1} dilution were plated on selective YC Glucose plates containing 0.09% 5-FOA and grown for 3 days at 30°C. Expansion events were confirmed by PCR using the A2 and B2 primer set and analyzed in ImageLab (BioRad).

Contraction rates. Fluctuation assays for contraction rates were conducted as described in (Khristich et al., 2020) at 30°C.

ACKNOWLEDGMENTS

We would like to thank all present and past members of the Mirkin laboratory for the useful comments and discussions. pRCC-N was a gift from Eckhard Boles (Addgene plasmid # 81192 ; http://n2t.net/addgene:81192 ; RRID:Addgene_81192).

CRediT AUTHOR STATEMENT

Conceptualization and experimental design: C.M. and S.M.M. **Data accumulation:** C.M. and A.L. **Writing:** C.M. This research was supported by the NSF-BSF 2153071 and NIGMS R35GM130322 grants to S.M.M.

Table 4: List of yeast strains used in Chapter 3.

Strain Name	Genotype	Comments
123 SMYP1	CH1585 + GalURA3-GAA128-XS	This study, made by Alexandra Khristich
166 SMYUI2	SMY710 + GalURA3-GAA124-XS-inverted	PMID: 31911468
168 SMYUI4	SMY710 + GalURA3-GAA124-XS-inverted	PMID: 31911468
CM115, CM116, CM117	CM71 *rnr1Y285C*	This study
CM157, CM159	CM71 *rnr1A245V*	This study
CM160, CM161	CM71 *rnr1F15S*	This study
CM38, CM39	SMYU *sic1::hygB*	This study
CM43, CM45	SMYP1 *rnr1A245V*	This study
CM52, CM53	SMYP1 *rnr1F15S*	This study
CM71	MATa leu2-Δ1 trp1-Δ63 ura3-52 his3-200 bar1Δ::HIS3 ChrIII(75423-75715)::URA3-Int-GAA100d-TRP1	PMID: 23142667
CM77 SMYU	MATa leu2-Δ1, trp1-Δ63, ura3–52, his3–200 (CH1585), ade2::KanMX, MIP1-wt, HAP1-wt, ChrIII(75594-75641):: PGAL-UR-(GAA)128-A3-TRP1	PMID: 31911468

Table 5: List of primers used in Chapter 3.

Name	Sequence	Purpose
pRCC plasmid amplification oligos to introduce CRISPR/Cas9 guides		
RNR1_pRCC_F	ACCAATACCACCAGCAGTTTgttttagagctagaaatagcaagttaaaataagg	*rnr1-A245V* and *rnr1-Y285C* mutations
RNR1_pRCC_R	AAACTGCTGGTGGTATTGGTCcgatcatttatctttcactgcggag	
RNR1F15S_pRCC_F	GAGCGGTAATCTTATCGAATgttttagagctagaaatagcaagttaaaataagg	*rnr1-F15S* mutation
RNR1F15S_pRCC_R	ATTCGATAAGATTACCGCTCcgatcatttatctttcactgcggag	
CRISPR/Cas9 repair templates		
RNR1_A245V_donor	GGACTCTATCGAGGGGATTTACGACACCTTGAAGGAATGTGCTTTGATTTCGAAAACTGTTGGTGGTATTGGTCTACATATCCATAACATTCGTTCAACT	*rnr1-A245V* mutation
Rnr1_F15S_rep_temp	ATTAACATCATGTACGTTTATAAAAGAGACGGTCGTAAAGAACCTGTaCAAtccGATAAGATTACCGCTCGTATATCACGCTTATGCTATGGTTTAGATC	*rnr1-F15S* mutation
Rnr1Y285C_rep_temp	TACTTCTAACGGTTTAATTCCTATGATTCGTGTTTTCAATAACACTGCCCGTtgcGTTGACcaaGGTGGTAATAAAAGACCTGGTGCGTTTGCCCTTTA	*rnr1-Y285C* mutation
Gene replacement primers		
SIC1_pAG_F	CCAAACCTCTACGGAATTTTGACCCTTGAAGCAGGGACTATTACACGAAACGGATCCCCGGGTTAATTAA	Sic1 deletion
SIC1_pAG_R	AGTAAGTAAATAAAATATAATCGTTCCAGAAACTTTTTTTTTTCATTTCTcataggccactagtggatctg	
Checking primers		
SIC1_chk_F	GCACTATTGCCGTGTTTCCT	Sic1 deletion check
SIC1_chk_R	CATCCTCCCCTAACTCGCTT	

SIC1_ORF_F	GCCAGGAGTCAAGAAAGTGA	
SIC1_ORF_R	GAAACAATGCCTTTGGCTTG	
RNR1_A245V_check_F	TATGAGAGTCGCACTAGGCA	*rnr1-A245V* check
RNR1_A245V_check_R	CTCTTTACCGTGGTTCTTCCTA	
RNR1_A245V_seq	AACGATCGGCTGCCATGTTA	
Rnr1_upstr_F	CTTGTTGCCTTTGTTAAGTCAG	rnr1 mutations check

CHAPTER 4 Cancer-associated $(GAAA)_n$ repeats stall DNA synthesis and form DNA triplexes *in vitro*

Chiara Masnovo[1], Sergei M. Mirkin[1]

[1] Department of Biology, Tufts University, Medford, MA, USA 02155

ABSTRACT

Triplex forming sequences are enriched in translocation events in cancers and somatic cancer-associated mutations are found adjacent to and within H-DNA forming sequences. Expansion of a $(GAAA)_n$ repeat in the first intron of the *UGT2B7* gene has been identified as the first cancer-associated repeat expansion event, occurring in 34% of all clear cell renal cell carcinoma cases. The $(GAAA)_n$ sequence has the potential to form a DNA triplex due to its homopurine-homopyrimidine mirror-repeat nature, as well as potentially serving as a DNA unwinding elements due to its A-rich composition. Here, we started investigating the structure forming abilities of the $(GAAA)_n$ repeat using *in vitro* DNA polymerization assays and chemical probing followed by primer extension assays. We show that the $(GAAA)_n$ can stall DNA synthesis by a B-family DNA polymerase in a manner that is promoted by increased Mg^{2+} ions and is consistent with the formation of an intramolecular triplex during DNA synthesis.

INTRODUCTION

Short tandem repeats (STRs), which consist of repetitions of 1-9 bp, represent about 3% of the human genome [1]. Historically, the identification and characterization of STRs has been hindered by several issues that arise when trying to sequence through them. First, many of these regions are too large to be covered in a single sequencing read using technologies such as Sanger sequencing and NGS, and therefore cannot be confidently mapped to the reference genome [371]. In addition, several sequencing methods encounter the same problems that occur during DNA replication, such as polymerase stalling and polymerase slippage. These issues have led to an underestimation of the disease states caused by repeat expansions, and advances in both sequencing and computational technologies have allowed the identification of several additional repeat expansion diseases (REDs), and more than 20 new REDs have been identified since 2009 [4,372]. The overwhelming majority of REDs are rare neurodegenerative diseases, but the involvement of repeat expansions in cancer has been hypothesized since 1995 [373]. Since then, instability of STRs has been observed in various cancer genomes, especially in the case of microsatellite instability (MSI), which is highly elevated in MMR-deficient colorectal cancers [374].

In 2023, Erwin et al., set out to identify recurrent repeat expansions (rREs) in cancer genomes using available genomic short-read sequencing datasets and a novel computational tool called ExpansionHunter Denovo (EHdn) [375], which does not rely on the presence of a reference genome [376]. They identified 160 rREs, which were non-randomly distributed, enriched near or at functional locations of the genome and highly sub-type specific. A large portion of the identified rREs consisted of homopurine-homopyrimidine repeats, with $(GAA)_n$ and $(GAAA)_n$ repeats being the mostly highly enriched sequences. In a specific example, $(GAAA)_n$ repeat expansions

in the first intron of the *UGT2B7* gene were identified in 34% of all clear cell renal cell carcinomas (ccRCCs). Healthy kidney cells harbor on average 26 repeat units, while affected cell lines have between 63 and 160 units. Targeting the expanded alleles with a synthetic transcription factor that targets $(GAAA)_n$ and recruits transcriptional machinery led to decreased proliferation and increased cell death in cell lines containing repeat expansions, as had previously been shown for $(GAA)_n$ repeats in Friedreich's ataxia cell lines, indicating a potentially shared pathogenic mechanism. In a separate study aimed at identifying genome-wide H-DNA forming sequences using the S1-END-Seq technique reviewed in Chapter 1, $(GAA)_n$ and $(GAAA)_n$ repeats were the two most highly identified sequences, indicating that they generally have a strong triplex forming potential [46]. In addition, the same sequences have been found to highly correlate with translocation breakpoints in cancers [377].

Together, these studies indicate that $(GAAA)_n$ repeats might be able to form triplex DNA structures and stall DNA replication, as has been shown to be the case for both $(GAA)_n$ repeats and $(A_2G_3)_n$ repeats, which are associated with the RED CANVAS (Cerebellar Ataxia, Neuropathy, Vestibular Areflexia Syndrome), *in vitro* as well as *in vivo* [50,53,378].

In this brief Chapter, we performed *in vitro* polymerization experiments using both A- and B-family polymerases and determined that $(GAAA)_{13}$ repeats can stall DNA polymerization *in vitro* when synthesis is performed with a B-family polymerase and in conditions which promote triplex formation, such as the increased presence of available Mg^{2+} cations.

RESULTS

The $(GAAA)_n$ repeat does not impede replication by an A-type polymerase and can act as a DNA unwinding element

We first set out to test whether $(GAAA)_n$ repeats would stall in vitro polymerization using the A-family polymerase Thermo Sequenase, an exonuclease deficient DNA polymerase I. We used a plasmid containing the $(GAAA)_{13}$ repeat as a template and carried out DNA sequencing reactions using a primer that either binds upstream or downstream of the repeat to initiate synthesis. We carried out the extension reactions at a temperature of 72°C using 3.5 mM Mg^{2+}. In this case, we did not observe a stall in either orientation of the repeat **(Figure 4-1 a)**.

We additionally used chemical probing to determine whether the $(GAA)_{13}$ forms a secondary structure using potassium permanganate ($KMnO_4$), which modifies single-stranded thymines, inhibiting the formation of Watson-Crick base pairing. We only conducted this analysis with the $(T_3C)_{13}$ template, as we expected that if there were to be a stall it would be more prominent when $(A_3G)_{13}$ is the template, as was observed for the $(A_2G_3)_n$ repeat [378]. We observed a pattern of repeat modifications consistent with a region of single-stranded DNA spanning the whole length of the repeat, whereas the unmodified plasmid had basically no signal at all **(Figure 4-1 a)**. These data collectively indicate that under the tested conditions (i) there is no repeat-mediated inhibition of DNA synthesis *in vitro* and (ii) the pattern observed in when conducting chemical probing is consistent with the whole repeat being unwound, a pattern consistent with the $(GAAA)_n$ repeat acting as a DNA unwinding element in supercoiled DNA **(Figure 4-1 b)**.

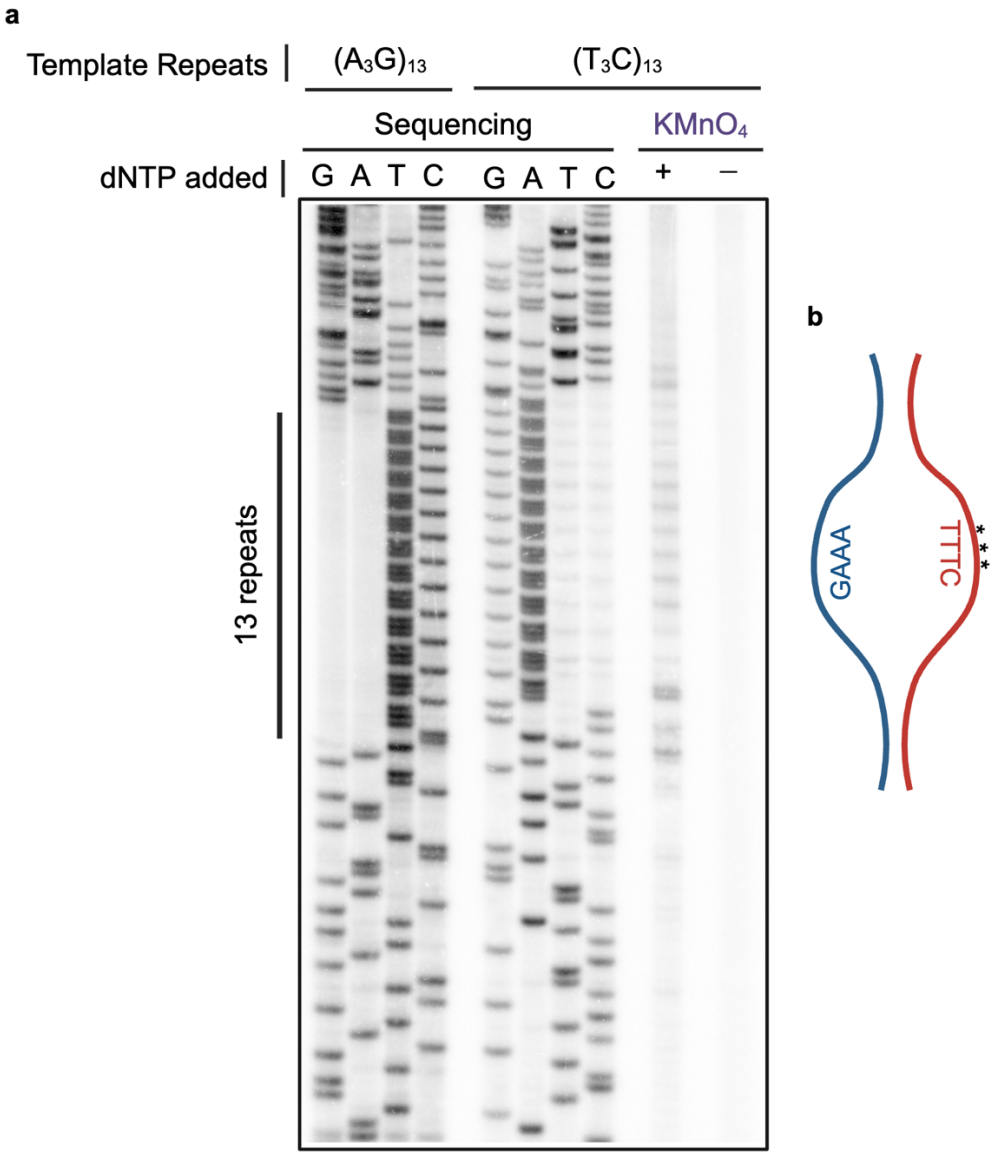

Figure 4-1: Thermo Sequenase does not stall during polymerization through the $(GAA)_{13}$ repeat.

(a) Sequencing reactions with Thermo Sequenase showing unperturbed replication through the $(GAAA)_{13}$ repeat in either orientation. On the right, the last two lanes show the primer extension reactions for potassium permanganate treated and water treated plasmids, pyrimidine strand was used as the template. More details on the experimental procedure are provided in the Materials and Methods section. **(b)** Schematic of a DNA unwinding element, such as the one that can be formed by the $(GAA)_{13}$ repeat. Asterisks indicate single-stranded thymines that can be modified by $KMnO_4$. Created in BioRender.

The $(GAAA)_n$ repeat impedes replication by a B-type polymerase and can form a DNA triplex

We then decided to conduct the in vitro polymerization experiments using a B-type polymerase instead, which are the family that the replicative eukaryotic DNA polymerases belong to. For these experiments, we used exonuclease-deficient Vent polymerase. In this experiment, we lowered the temperature to 50°C to mimic a more physiological reaction. In addition, we tested two different concentrations of Mg^{2+} (2 mM and 5 mM). We observed an orientation-dependent inhibition of DNA synthesis, in which there was a stall when the purine run served as a template, but not when the pyrimidine run did. The stall which was exacerbated in at 5 mM Mg^{2+} **(Figure 4-2 a).**

Since the signal accumulation occurred in the middle of the repeat tract and there is a dependence on the concentration of the magnesium cations, we conclude that under these conditions, Vent is stalled by the formation of an intramolecular triplex during DNA synthesis **(Figure 4-2 b)**.

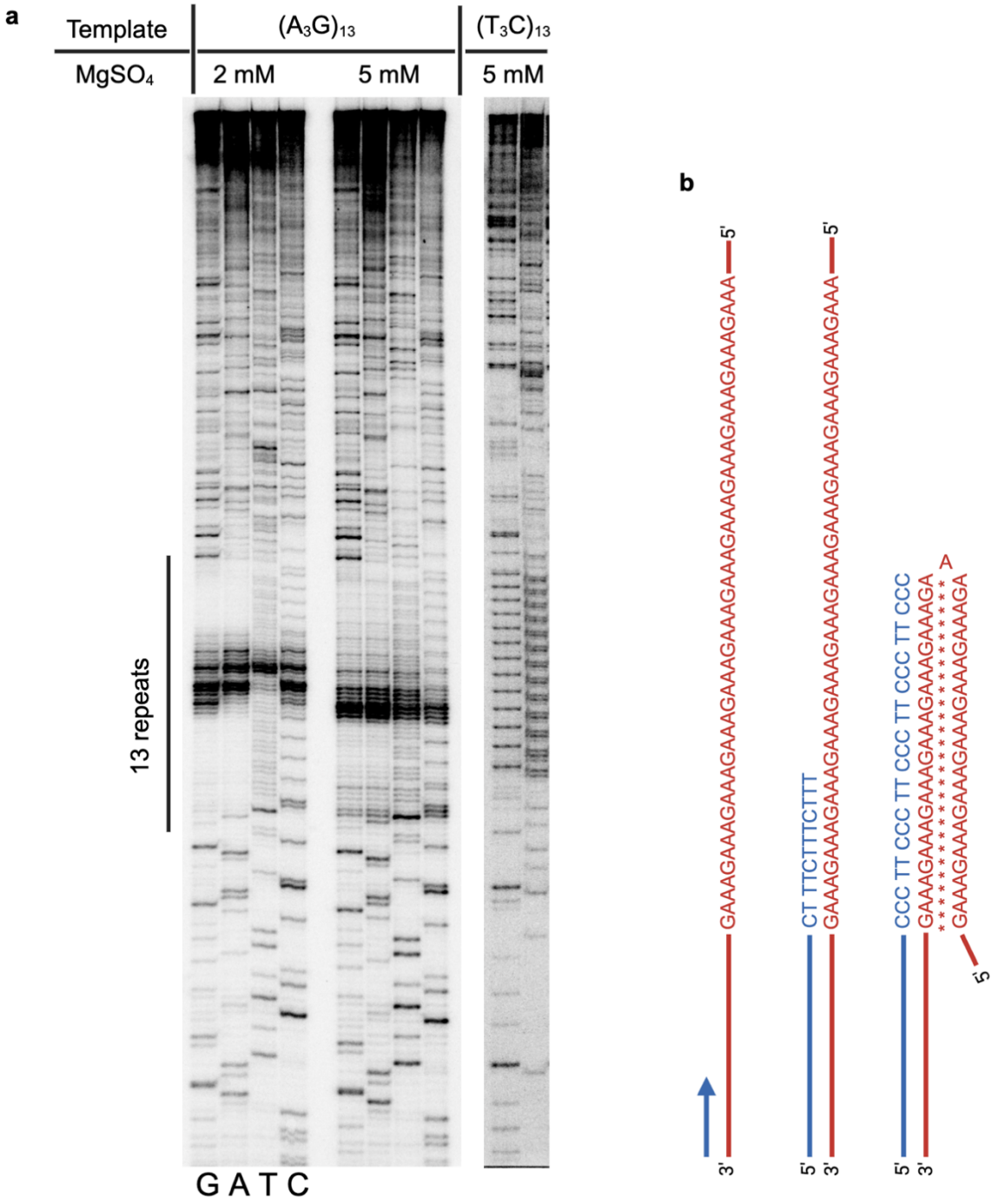

Figure 4-2: Vent (exo-) stalls during polymerization through the $(GAA)_{13}$ repeat when the homopurine run is the template for synthesis.

(a) Vent polymerase stalls in the middle of $(GAAA)_n$ repeats only when they serve as the template and the stalling is exacerbated by increased Mg^{2+} availability. More details on the experimental procedure are provided in the Materials and Methods section. **(b)** Schematic

representation of triplex formation during a single run of synthesis showing why the polymerase would stall in the middle of the repeat. Created with BioRender.com.

DISCUSSION

In this Chapter, we provide preliminary data regarding in vitro polymerization and structure formation for the $(GAAA)_n$ repeat, whose expansions in the first intro of the *UGT2B7* gene have recently been associated with the etiology of ~34% of all clear cell renal cell carcinoma cases, being responsible for ~110,000 yearly cases of kidney cancer worldwide [379]. Formation of DNA triplexes has been previously associated with cancers in the case of the Burkitt lymphoma translocation hotspot in the *c-MYC* gene. Several studies have shown that the triplex formed at the *c-MYCI* promoter sequence promotes the oncogenic translocation between *c-MYC* and an immunoglobulin gene, and this sequence is associated with increased mutagenesis [77,81,380,381]. More generally, triplex-forming sequences are enriched in translocation breakpoints in several types of cancer [80,377], and triplex forming sequences are enriched in mutations that are recurrent in cancer [60]. Collectively, this indicates that in many cases, triplex forming sequences associate with cancers.

The case of the $(GAAA)_n$ repeat in *UGT2B7* is of particular interest because in this case the expanded repeat itself could be the cancer-driving mutation, and it is the first ever case of a repeat expansion being associated with a cancer. While the rRE was not associated with overall decreased levels in UGT2B7 expression, Erwin et al., noted that there was a decrease in a specific transcript isoform, which has been previously associated with the same type of cancer [375]. Crucially, they also used a synthetic transcription factor Syn-TEF, which displayed anti-proliferative activity specifically on cell lines with the expanded repeats. Therefore,

understanding the behavior and features of this expansion could be crucial to the development of therapeutics. Specifically, we aimed at determining whether this sequence could form non B-DNA structures during *in vitro* replication.

Since the $(GAAA)_n$ repeat is both A-rich and is a homopurine-homopyrimidine mirror repeat, we hypothesized that it could either form a DNA unwinding element (DUE) or a DNA triplex, also called H-DNA. Our experiments with the A-family polymerase Thermo Sequenase showed no stalling the chemical probing pattern was consistent with the repeat behaving like a DUE. Another sequence containing just one more adenine, the $(A_4G)_n$ repeat, was shown to also behave like a DUE in a similar chemical probing and primer extension assay [378]. On the other hand, the experiment with the B-family polymerase Vent (exo-) Polymerase showed a stall when the purine run was the template for replication, and the stalling pattern in which the polymerase was able to replicate just the first half of the repeat, is consistent with the formation of a triplex during a single cycle of replication. Therefore, we conclude that the $(GAAA)_n$ repeat is able to form a DNA triplex at least *in vitro*.

Whether $(GAAA)_n$ repeats can form triplexes *in vivo* as well as the mechanisms of instability of this repeat remain to be determined, and it is work that will be complementary to the one presented in this chapter to form a complete study.

MATERIALS AND METHODS

All protocols in this Chapter were adapted from Hisey et al., 2024 [378], where I contributed as an author conducting the experiments found in Figure 1 b. Oligonucleotides used in this Chapter are

listed in **Table 6**. The CMP50 plasmid is identical to the pJH1 plasmid described in [378], but with the $(A_3G)_{13}$ repeat replacing the $(A_2G_3)_{60}$ repeat found in pJH1.

Chemical Probing with Potassium Permanganate

2 μg of supercoiled DNA from plasmid CMP50 were incubated in 10 mM Tris–HCl pH 7.5, 2 mM $MgCl_2$ buffer with either 6 mM $KMnO_4$ or an equal volume of water for 2 min at 37°C. The reaction was quenched with 1 M β-mercaptoethanol, precipitated using 100% ethanol, washed with 70% ethanol, and resuspended in 5 μl water, and used as a template for primer extension as described in the Thermo Sequenase section using the USBio Thermo Sequenase Cycle Sequencing Kit (Cat# 78500) with the following modifications: the DNA was pre-annealed with 0.25 pmol primer JH271. The final volume of the labeling reaction was 8.75 μl with the same ratios as described in the manufacturer's instructions, and dNTPs were added to a final concentration of 75 μM for each dNTP for the primer extension step.

DNA Polymerization Experiment with Thermo Sequenase

The double-stranded plasmid CMP50 was used as a template. Primers JH271 and JH272 were used to polymerize through the repeats with either the pyrimidine or purine repeat in the template strand, respectively. Reactions using Thermo Sequenase were carried out as described in [378]: The USBio Thermo Sequenase Cycle Sequencing Kit (Cat# 78500) was used according to the manufacturer's 3′-dNTP internal label cycling sequencing instructions with the following modifications. Instead of conducting multiple rounds of labeling and extending, we conducted only one round of extension. 5 ug of the plasmid were mixed with 0.5 pmol of either JH271 or JH272 primer and water was added to a total volume of 11 μl water, annealed at 95°C for 2 min and immediately submerged in an ice-water bath. The labeling components were then added and

the primer extension product was labeled at 60°C for 30'' with 0.625 μM [α-^{32}P]dATP. We then aliquoted the reaction into four pre-aliquoted termination mixes, and termination was carried out at 72°C for 5 min. 4 μl of stop solution were added to each reaction, mixed, and incubated at 95°C for 5 min and immediately submerged into an ice-water bath. Each sequencing termination reaction contained 80 μM dNTPs and 0.8 μM of the designated ddNTP. 8 μl of each reaction was loaded onto a 6% polyacrylamide gel with 7.5 M urea prepared according to manufacturer's instructions (National Diagnostics SEQUAGEL SEQUENCING SYSTEM 2.2 #EC-833). The gel was run at 1800 V for ~ 2.5 hours, dried, exposed and imaged using an Amersham Typhoon PhosphorImager.

DNA Polymerization Experiment with Vent (exo-) Polymerase

Vent (exo-) DNA polymerase (New England Biolabs # M0257S) was used according to primer extension experiments described in [382] with the following modifications. 0.5 pmol of primer and 5 μg (1.06 pmol) of DNA were pre-annealed as described above in Thermo Sequenase reactions, labeling was carried out at 60°C for 30 s, termination was carried out at 50°C for 5 min. Steps after termination were identical to those with Thermo Sequenase. Exact ddNTP and dNTP concentrations varied based on the exact nucleotide to yield optimal sequencing reaction resolution, but ddNTP concentrations ranged from 400 to 900 μM and dNTP concentrations ranged from 30 to 100 μM, as described in [378].

Table 6: List of primers used in Chapter 4.

Number	Name	Sequence
JH271	JH271_C_template_dATP_lab	AGCGCTATATGCGTTGATGC

JH272	JH272_G_template_dATP_lab	GCTTAAAAAGATTCTCTTTTTTTATGATATTTGTACA T

ACKNOWLEDGMENTS

We would like to thank Dr. Julia Hisey for the collaboration on this project and teaching me how to perform DNA sequencing gels and chemical probing.

CRediT AUTHOR STATEMENT

Conceptualization and experimental design: C.M. and S.M.M. Data accumulation: C.M. Writing: C.M.

CHAPTER 5 OVERALL CONCLUSIONS AND FUTURE DIRECTIONS

One of the most interesting results of our study on the role of Mcm10 in preventing homopurine-homopyrimidine repeat instability in Chapter 2 is that while Mcm10 deficiency affects both contractions and expansions of the repeats, the mechanisms appear to be fundamentally different, as we have determined by performing genetic controls and we discuss more in detail here.

In the case of $(GAA)_n$ contractions in *mcm10-1*, our data suggests that they occur due to increased ssDNA accumulation at the fork during lagging-strand synthesis, which promotes formation of slipped strand structures (including but not limited to DNA triplex formation), which lead to contractions during synthesis **(Figure 5-1)**. Our genetic data is consistent with this model: both the *mcm2-G400D* mutation and RPA overexpression, which counteract the accumulation of excessive uncoated ssDNA, rescue contractions in *mcm10-1*.

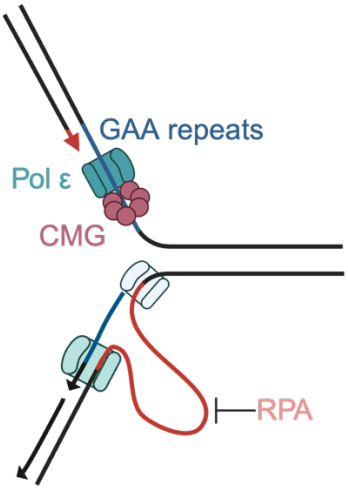

Figure 5-1: Overall model of repeat contractions in the mcm10-1 context.

Further, we also show that in the *mcm10-1 rev1Δ* double mutant, we observe a further increase in instability, as was previously observed in the wild-type context[64]. This is due to the fact that Rev1 is able to insert a G opposite to a C or lesion in the template strand, and our lagging strand template is the $(GAA)_n$ run[383], and therefore promotes lagging-strand synthesis through the repeats. Overall, we would like to emphasize that in the *mcm10-1* mutant, contractions happen through the same mechanism they do in the wild-type context[64], but since the basal levels of ssDNA at the replication fork is highly elevated, so are the contraction rates.

The overall increase in repeat instability in the *mcm10-1* mutant was accompanied by a repeat-length dependent decrease in viability at the semi-permissive temperatures. Interestingly, this held true both in the case of long GAA repeats as well as GAGAAGAAA repeats of the same length, which cannot form a strong DNA triplex. In both cases, the viability defect was rescued by overexpression of the ssDNA-binding protein RPA. Therefore, we conclude that exposure of long ssDNA gaps containing repetitive sequences is especially problematic for cell survival in the Mcm10-deficient context. Why is that the case? The global increase in ssDNA exposure in the *mcm10-1* mutant can cause depletion of RPA at the ssDNA repeat sequence. In addition, this effect is exacerbated by the fact that RPA has a 10-fold lower affinity for poly-purine tracts than poly-pyrimidine tracts[384]. In our case, the GAA poly-purine tract is the lagging-strand template. RPA overexpression can therefore attenuate the excessive ssDNA exposure on the lagging-strand template to counteract repeat contractions. In the case of viability, we believe that RPA coating prevents excessive conversion of both leading and lagging strand gaps into DSBs and promotes checkpoint activation and repair synthesis, thereby ensuring cell survival.

The mechanism of expansions in the *mcm10-1* mutant appears to be quite different. We conclude that they occur during post-replicative filling of ssDNA gaps generated during DNA replication. Those expansions are rescued by the *mcm2-G400D* mutation, which prevents the formation of the ssDNA gaps in the first place, but they are not rescued by RPA overexpression, indicating that ssDNA-coating is not sufficient to prevent the expansion events. We propose two slightly different models for how the expansions happen in the context of RPA depletion at the repeat, versus in the context of efficient coating by RPA **(Figure 5-2)**. Replication gaps in *mcm10-1* have indeed been reported before[292], and in human cells, Mcm10 suppresses PRIMPOL mediated gap accumulation[315]. ssDNA gaps left behind the replication fork lead to the checkpoint activation, and we indeed observe Rad9-dependent checkpoint activation during late

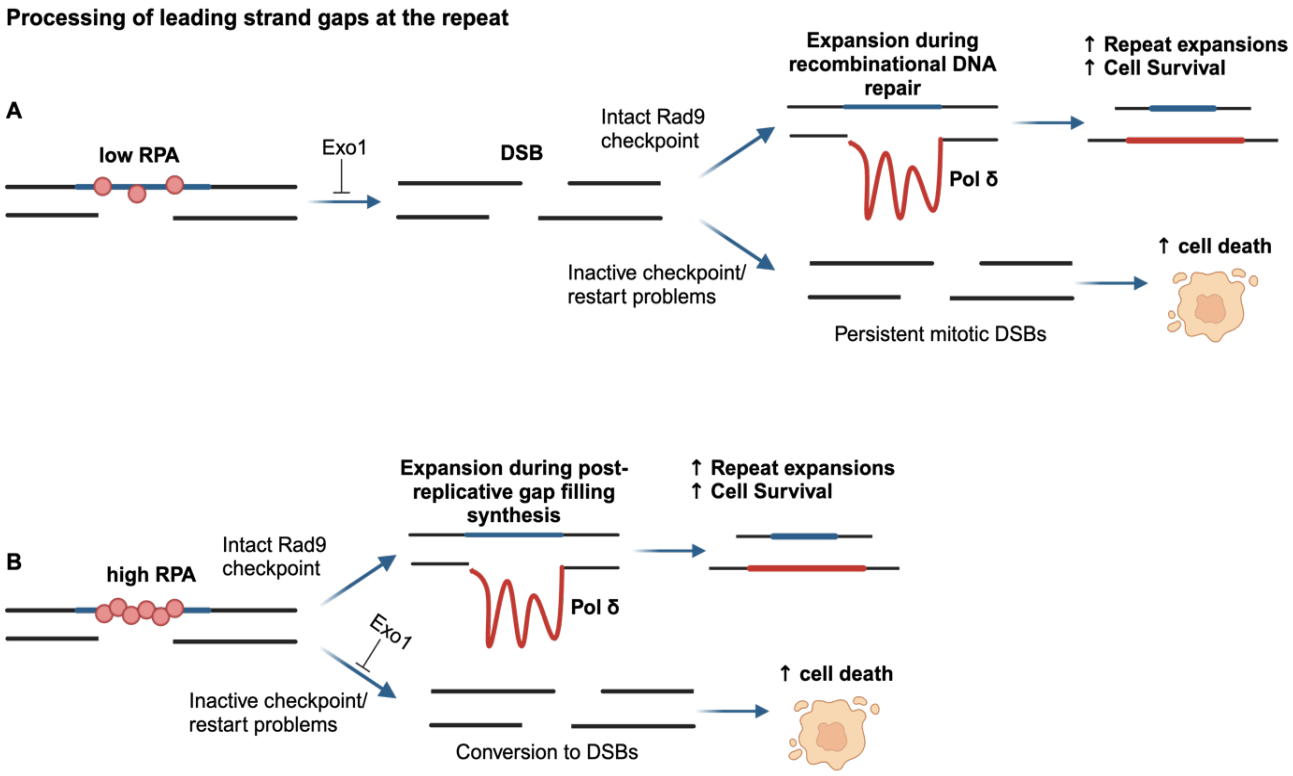

Figure 5-2: Models for processing of leading strand gaps in the mcm10-1 mutants leading to repeat expansions and ensuring cell viability in the context of low (A) vs. high (B) RPA levels.

S-phase, which promotes repeat expansions. On the other hand, the Mrc1 checkpoint function is not required for this process. This is consistent with persistence of the replication stress response at ssDNA gaps[299,302]. As was observed by García-Rodríguez et al., Exo1 processing of daughter-strand gaps can further exacerbate this effect, as we observed a mild rescue in growth in the *mcm10-1 exo1Δ* context.

Ultimately, DNA synthesis through the repeat (and other gaps at essential locations) needs to be performed to maintain genome integrity. We, thus, set out to determine which post-replicative pathway could be responsible for this synthesis. Elevated expansions in the *mcm10-1* mutant did not involve Rev3 – a subunit of DNA polymerase ζ needed for translesion synthesis (TLS), nor Rad5 – a ubiquitin ligase needed for template switching and TLS. On the other hand, our data indicated that Pol δ is responsible for the elevated expansions in *mcm10-1*, as reducing its fidelity partially rescued them. Furthermore, both fidelity and processivity of Pol δ were crucial to maintain the viability of *mcm10-1* strains. There is evidence that Pol δ can perform gap-filling in a manner distinct and complementary from its role in the DNA polymerase ζ, and our Pol3 phenotype supports this hypothesis, given the fact this subunit is unique to Pol δ. We propose that during post-replicative synthesis by Pol δ, there is an increase in polymerase slippage, as the polymerase is not stabilized by the accessory subunits of the replisome, leading to repeat expansions. We believe that this gap-filling occurs post-replication, either during late S-phase or during G2/M. Note that our proposed mechanism and genetics is generally consistent with a break-induced replication (BIR) mechanism, which was previously discussed by us for CAG repeats [385], but the elevated expansions in the *mcm10-1* mutant were not dependent on Rad52[386]. Nevertheless, BIR might still be a general contributor to ensuring genome integrity in

the *mcm10-1* mutant, as Rad52 (but not Rad51) deletion did have a large synthetic effect on viability with the *mcm10-1* mutation.

Our observations in Chapter 3 indicate that nucleotide pool imbalances, especially overall increase and decrease in dNTPs, can contribute to an increase in repeat instability. When dNTP pool decreases resulting from HU treatment or Rnr1 mutations, slower dNTP incorporation might lead to an increase in ssDNA gaps, including at the repeat, similarly to what we observed in the *mcm10-1* mutation. Thus, it would be interesting to test whether RPA overexpression in this context would attenuate the effect on contractions and expansions, and whether combining dNTP depletion with the *mcm10-1* mutation causes an exacerbation of the observed effects in the *mcm10-1* context. On the other hand, an increase in dNTP pool could lead to an overall increase in error-prone DNA synthesis, which we hypothesize could synergize with the *mcm10-1* mutation in the context of repeat expansions.

For future directions, it will be extremely interesting to 1) determine whether Mcm10 mutations gave the same effects on other homopurine-homopyrimidine repeats, such as the CANVAS-associated $(A_2G_3)_n$ repeat[378], and 2) investigate whether Mcm10 functioning is important to maintain other types of structure-forming tandem repeats, such as AT repeats, which can form cruciforms[387], CAG repeats, which can form hairpin-like structures[388], and G-quadruplex forming repeats. On the flip side, Mcm10 overexpression was shown to increase the instability of a poly-GT tract[316], making it foreseeable that the same could be true for other tandem repeats. Given that Mcm10 amplification is a prevalent feature of many cancers, such as bladder and ovarian cancer, future studies are needed to establish whether Mcm10 overexpression leads to tandem repeat instability in cancers. Supporting this idea, high Mcm10 levels have been correlated with microsatellite instability in seven tumor types[389]. These efforts

will be particularly important in cases in which the cancer type itself is associated with repeat expansions, such as in the case of the A_3G repeat we studied in Chapter 4.

CHAPTER 6 APPENDIX

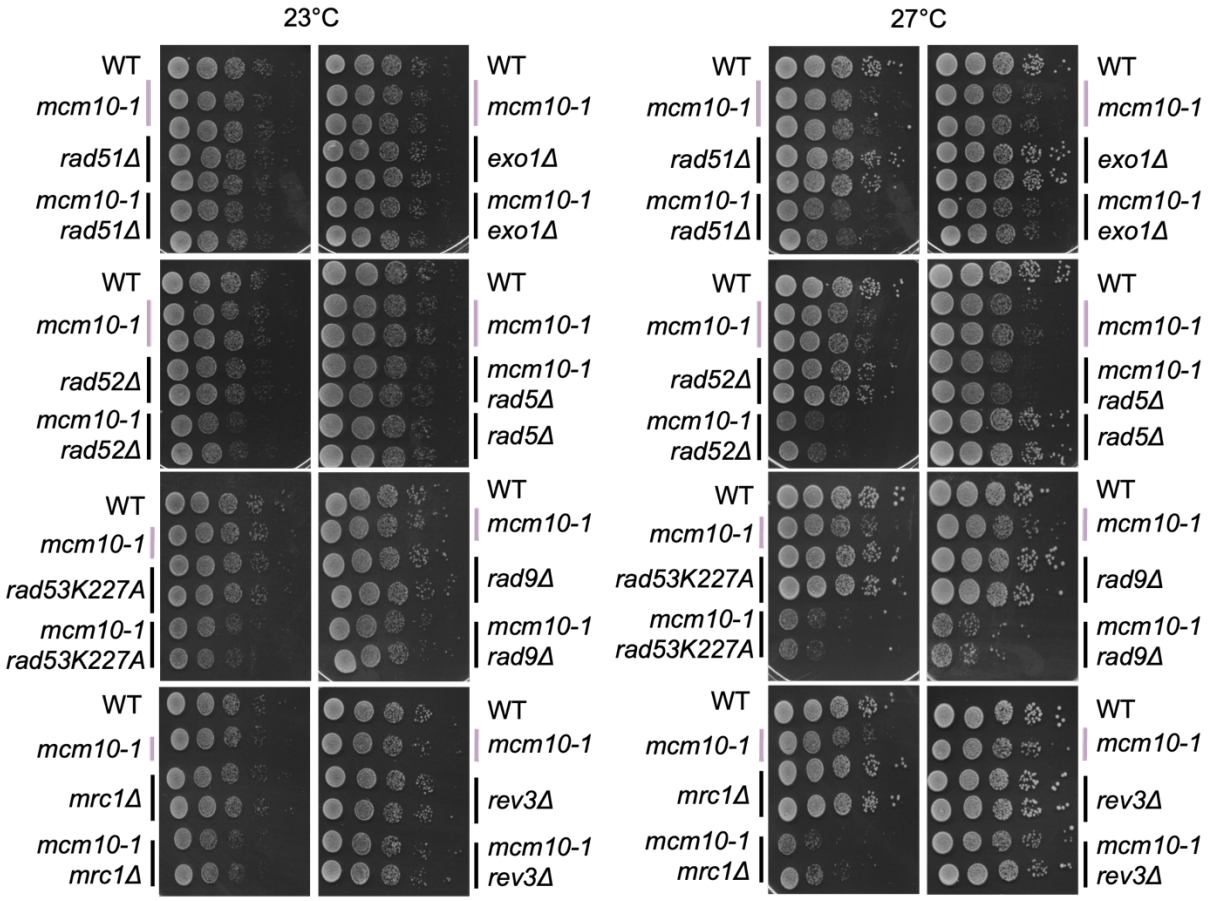

Figure 6-1: Spot tests at the permissive (23°C) and semi-permissive (27°C) temperatures for the genotypes tested in Figure 2-10.

Assay conducted as described in the referenced Figure.

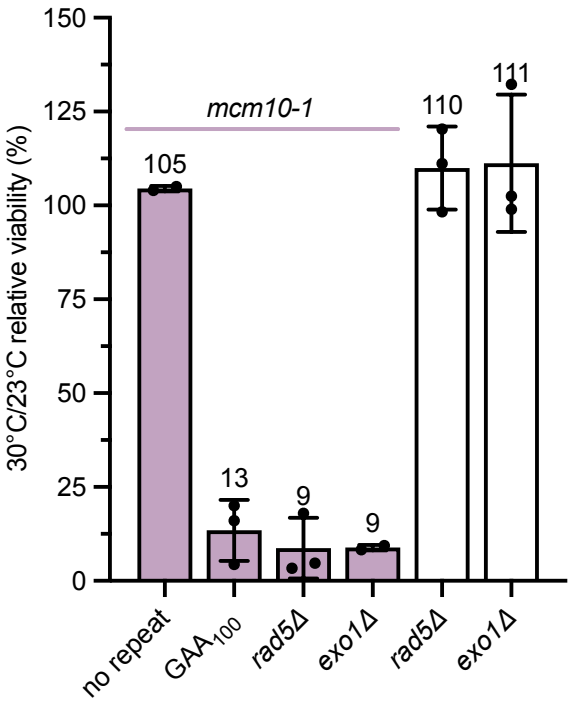

Figure 6-2: Colony forming unit viability assay for additional *mcm10-1* double mutants.

Conducted as described in **Figure 2-9**.

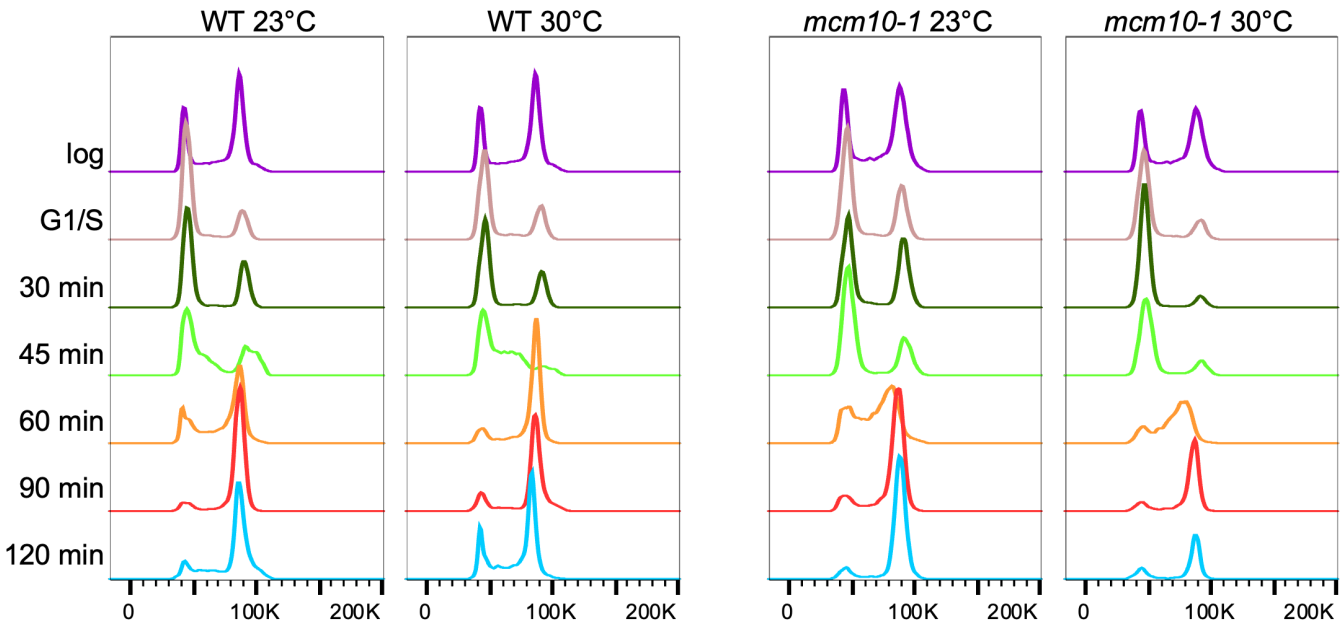

Figure 6-3: Flow Cytometry of DNA content for WT and *mcm10-1* at 23°C versus 30°C.

Experiment and analysis conducted as described for **Figure 2-12 f**.

Table 7: Raw expansion rates for data presented in Chapter 2 and Chapter 3.

Chapter 2					
Strain Number	**Genotype**	**rate**	**lower CI**	**upper CI**	**Temp/Treatment**
CM71	WT	5.35E-06	2.85E-06	8.38E-06	23°C
CM72, CM73	*mcm10-1*	3.13E-05	2.07E-05	4.36E-05	23°C
CM71	WT	1.30E-05	7.99E-06	1.89E-05	27°C
CM72, CM73	*mcm10-1*	4.28E-04	2.82E-04	5.94E-04	27°C
CM85, CM86	*mcm2-G400D*	5.20E-06	2.76E-06	8.18E-06	27°C
CM88, CM89	*mcm10-1 mcm2-G400D*	2.62E-05	1.68E-05	3.71E-05	27°C
CM388, CM389	*mcm10-1 RPA OE*	5.87E-04	3.39E-04	8.82E-04	27°C
CM71	WT	2.93E-05	1.69E-05	4.41E-05	30°C, DMSO
CM71	WT	9.83E-05	5.75E-05	1.47E-04	30°C, 100 mM HU
CM71	WT	2.93E-05	1.69E-05	4.41E-05	30°C
CM227, CM229	*ctf4Δ*	4.37E-05	2.77E-05	6.23E-05	30°C
CM137, CM138	*rfc1-1*	5.01E-05	3.04E-05	7.33E-05	30°C
CM137, CM138	*rfc1-1*	1.24E-04	8.33E-05	1.70E-04	27°C
CM218, CM219	*mcm10-1 rad5Δ*	1.19E-03	6.53E-04	1.84E-03	27°C
CM348, CM349	*mcm10-1 rad51Δ*	2.28E-04	1.30E-04	3.44E-04	27°C
CM363, CM365	*mcm10-1 rad52Δ*	8.09E-04	4.64E-04	1.22E-03	27°C
CM445, CM447	*mcm10-1 rev3Δ*	2.97E-04	1.81E-04	4.34E-04	27°C
CM358, CM359	*mcm10-1 exo1Δ*	4.16E-04	2.51E-04	6.10E-04	27°C
CM200, CM201	*mcm10-1 rad53-K227A*	1.46E-04	8.79E-05	2.14E-04	27°C
CM415, CM416	*mcm10-1 rad9Δ*	5.73E-05	3.18E-05	8.80E-05	27°C
CM438, CM440	*mcm10-1 mrc1Δ*	2.77E-05	1.22E-05	4.75E-05	27°C
CM466, CM467	WT + pRS415 plasmid	2.69E-05	1.60E-05	3.98E-05	27°C
CM468, CM469	*mcm10-1* + pRS415	1.24E-03	7.40E-04	1.84E-03	27°C
CM479, CM480, CM481	*mcm10-1 mrc1Δ* + pRS415	5.91E-05	3.38E-05	8.92E-05	27°C
CM462, CM464	*mcm10-1* + pAO139 (mrc1AQ)	7.76E-04	4.65E-04	1.14E-03	27°C
CM85, CM86	*mcm2-G400D*	7.51E-06	3.97E-06	1.18E-05	23°C
CM88, CM89	*mcm10-1 mcm2-G400D*	3.79E-05	2.47E-05	5.32E-05	23°C
CM240, CM241	*mcm10-G261D*	5.07E-05	3.33E-05	7.06E-05	30°C
Chapter 3					
Strain Number	**Genotype**	**rate**	**lower CI**	**upper CI**	**Temp/Treatment**
CM36, CM37	GAA100	7.35E-06	2.85E-06	3.34E-06	30°C
CM157, CM159	GAA100 *rnr1A245V*	7.20E-05	4.93E-05	9.78E-05	30°C
CM71	GAA100	2.93E-05	1.69E-05	4.41E-05	30°C
CM115, CM116, CM117	GAA100 *rnr1Y285C*	3.47E-05	3.05E-05	3.91E-05	30°C

CM160, CM161	GAA100 *rnr1F15S*	1.04E-04	7.12E-05	1.41E-04	30°C
CM71	GAA100 WT	2.93E-05	1.69E-05	4.41E-05	30°C, DMSO
CM71	GAA100 WT	9.83E-05	5.75E-05	1.47E-04	30°C, 100 mM HU

Table 8: Raw contraction rates for data presented in Chapter 2 and Chapter 3.

Chapter 2					
Strain Number	**Genotype**	**rate**	**lower CI**	**upper CI**	**Temp/Treatment**
CM77	WT	7.85E-05	4.92E-05	1.12E-04	23°C
CM25, CM26	mcm10-1	1.15E-04	7.89E-05	1.57E-04	23°C
CM77	WT	8.43E-05	5.74E-05	1.15E-04	27°C
CM25, CM26	mcm10-1	8.62E-04	6.97E-04	1.04E-03	27°C
CM59, CM60	mcm2-G400D	6.65E-05	4.09E-05	9.66E-05	27°C
CM69, CM70	mcm10-1 mcm2-G400D	2.25E-04	1.62E-04	2.96E-04	27°C
CM299, CM300	mcm10-1 RPA OE	1.84E-04	1.24E-04	2.52E-04	27°C
CM77	WT	1.70E-04	1.18E-04	2.28E-04	30°C, DMSO
CM77	WT	1.42E-03	1.10E-03	1.76E-03	30°C, 100 mM HU
CM77	WT	1.70E-04	1.18E-04	2.28E-04	30°C
SMYU170, CM148	ctf4Δ	8.30E-04	5.33E-04	1.17E-03	30°C
CM143, CM144	*rfc1-1*	2.52E-03	2.04E-03	3.04E-03	30°C
CM143, CM144	*rfc1-1*	2.79E-03	2.31E-03	3.31E-03	27°C
CM195, CM197	*mcm10-1 rad53-K227A*	8.90E-04	6.25E-04	1.19E-03	27°C
CM59, CM60	*mcm2-G400D*	3.73E-05	2.04E-05	5.79E-05	23°C
CM67, CM68	*mcm10-1 mcm2-G400D*	9.17E-05	5.66E-05	1.33E-04	23°C
SMYUD24, SMYUD25	WT	1.03E-05	5.87E-06	1.55E-05	27°C
CM281, CM282	mcm10-1	1.16E-04	9.48E-05	1.39E-04	27°C
CM296, CM297	mcm10-1 RPA OE	3.06E-05	2.18E-05	4.05E-05	27°C
CM163, CM164	mcm10-G261D	1.35E-04	1.02E-04	1.70E-04	30°C
Chapter 3					
Strain Number	**Genotype**	**rate**	**lower CI**	**upper CI**	**Temp/Treatment**
SMYUI2, SMYUI4	GAA124 inverted	2.62E-04	1.85E-04	3.48E-04	30°C, DMSO
SMYUI2, SMYUI4	GAA124 inverted	2.80E-03	2.13E-03	3.53E-03	30°C, 100 mM HU
SMYU	GAA124	1.70E-04	1.18E-04	2.28E-04	30°C, DMSO
SMYU	GAA124	1.42E-03	1.10E-03	1.76E-03	30°C, 100 mM HU
CM38, CM39	SMYU *sic1::HygB*	4.03E-04	3.23E-04	4.89E-04	30°C
SMYP1	GAA124 with WT ade	1.09E-04	7.02E-05	1.53E-04	30°C
CM43, CM45	SMYP1 *rnr1A245V*	9.63E-05	6.28E-05	1.35E-04	30°C

CM52, CM53	SMYP1 *rnr1F15S*	2.18E-04	1.61E-04	2.81E-04	30°C

Table 9: Raw fragility rates for data presented in Figure 2-13.

Chapter 2					
Strain Number	**Genotype**	**rate**	**lower CI**	**upper CI**	**Temp/Treatment**
JAH293, KAH294	no rpt WT	4.61E-08	2.58E-08	7.04E-08	27°C
CM377, CM378	no rpt *mcm10-1*	6.1767E-07	5.0319E-07	7.4061E-07	27°C
CM351, CM353	GAA100 WT	1.77E-07	1.24E-07	2.35E-07	27°C
CM366, CM368	GAA100 *mcm10-1*	2.68E-06	2.19E-06	3.22E-06	27°C
CM390, CM393	GAA100 *RPA OE*	2.97E-07	1.93E-07	4.16E-07	27°C
CM391, CM391	GAA100 *mcm10-1 RPA OE*	1.76E-06	1.47E-06	2.07E-06	27°C
CM418, CM420	GAA100 *mcm2-G400D*	1.89E-07	1.32E-07	2.53E-07	27°C
CM422, CM423	GAA100 *mcm10-1 mcm2G400D*	1.79E-07	1.33E-07	2.30E-07	27°C

REFERENCES

1. Gymrek, M. A genomic view of short tandem repeats. *Current Opinion in Genetics & Development* **44**, 9–16 (2017).
2. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Subramanian, S., Mishra, R. K. & Singh, L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol.* **4**, R13 (2003).
4. Depienne, C. & Mandel, J.-L. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *The American Journal of Human Genetics* **108**, 764–785 (2021).
5. Khristich, A. N. & Mirkin, S. M. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.* jbc.REV119.007678 (2020) doi:10.1074/jbc.REV119.007678.
6. Spada, A. R. L., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**, 77–79 (1991).
7. Kiliszek, A. & Rypniewski, W. Structural studies of CNG repeats. *Nucleic Acids Research* **42**, 8189–8199 (2014).
8. Petruska, J., Hartenstine, M. J. & Goodman, M. F. Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *J Biol Chem* **273**, 5204–5210 (1998).

9. Campuzano, V. *et al.* Friedreich's Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion. *Science* **271**, 1423–1427 (1996).
10. Clark, R. M. *et al.* Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. *Genomics* **83**, 373–383 (2004).
11. Chauhan, C., Dash, D., Grover, D., Rajamani, J. & Mukerji, M. Origin and Instability of GAA Repeats: Insights from Alu Elements. *Journal of Biomolecular Structure and Dynamics* **20**, 253–263 (2002).
12. Vankan, P. Prevalence gradients of Friedreich's Ataxia and R1b haplotype in Europe co-localize, suggesting a common Palaeolithic origin in the Franco-Cantabrian ice age refuge. *Journal of Neurochemistry* **126**, 11–20 (2013).
13. Al-Mahdawi, S. *et al.* Large Interruptions of GAA Repeat Expansion Mutations in Friedreich Ataxia Are Very Rare. *Frontiers in Cellular Neuroscience* **12**, 443 (2018).
14. Filla, A. *et al.* The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. *Am J Hum Genet* **59**, 554–560 (1996).
15. Montermini, L. *et al.* The Friedreich ataxia GAA triplet repeat: premutation and normal alleles. *Hum Mol Genet* **6**, 1261–1266 (1997).
16. La Pean, A., Jeffries, N., Grow, C., Ravina, B. & Di Prospero, N. A. Predictors of progression in patients with Friedreich ataxia. *Mov Disord* **23**, 2026–2032 (2008).
17. Sharma, R. *et al.* Friedreich ataxia in carriers of unstable borderline GAA triplet-repeat alleles: FRDA Unstable Borderline Alleles. *Ann Neurol.* **56**, 898–901 (2004).
18. Campuzano, V. *et al.* Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. *Hum Mol Genet* **6**, 1771–1780 (1997).

19. Punga, T. & Bühler, M. Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. *EMBO Mol Med* **2**, 120–129 (2010).

20. Rodden, L. N. *et al.* Methylated and unmethylated epialleles support variegated epigenetic silencing in Friedreich ataxia. *Hum Mol Genet* **29**, 3818–3829 (2021).

21. Soragni, E. *et al.* Long intronic GAA•TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia. *Nucleic Acids Res* **36**, 6056–6065 (2008).

22. Li, Y. *et al.* Expanded GAA repeats impede transcription elongation through the FXN gene and induce transcriptional silencing that is restricted to the FXN locus. *Hum Mol Genet* **24**, 6932–6943 (2015).

23. González-Cabo, P. & Palau, F. Mitochondrial pathophysiology in Friedreich's ataxia. *Journal of Neurochemistry* **126**, 53–64 (2013).

24. Koeppen, A. H. *et al.* The dorsal root ganglion in Friedreich's ataxia. *Acta Neuropathol* **118**, 763–776 (2009).

25. Koeppen, A. H., Davis, A. N. & Morral, J. A. The cerebellar component of Friedreich's ataxia. *Acta Neuropathol* **122**, 323–330 (2011).

26. Cook, A. & Giunti, P. Friedreich's ataxia: clinical features, pathogenesis and management. *Br Med Bull* **124**, 19–30 (2017).

27. Zesiewicz, T. A. *et al.* Emerging therapies in Friedreich's Ataxia. *Expert Review of Neurotherapeutics* **20**, 1215–1228 (2020).

28. Mirkin, S. M. *et al.* DNA H form requires a homopurine–homopyrimidine mirror repeat. *Nature* **330**, 495–497 (1987).

29. Mirkin, S. M. & Frank-Kamenetskii, M. D. H-DNA and Related Structures. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 541–576 (1994).

30. Paleček, E. Local Supercoil-Stabilized DNA Structure. *Critical Reviews in Biochemistry and Molecular Biology* **26**, 151–226 (1991).

31. Potaman, V. N. *et al.* Length-dependent structure formation in Friedreich ataxia (GAA)n·(TTC)n repeats at neutral pH. *Nucleic Acids Res* **32**, 1224–1231 (2004).

32. Frank-Kamenetskii, M. D. & Mirkin, S. M. Triplex Dna Structures. *Annual Review of Biochemistry* **64**, 65–95 (1995).

33. Bergquist, H. *et al.* Structure-specific recognition of Friedreich's ataxia (GAA)n repeats by benzoquinoquinoxaline derivatives. *Chembiochem* **10**, 2629–2637 (2009).

34. Sakamoto, N. *et al.* Sticky DNA: Self-Association Properties of Long GAA·TTC Repeats in R·R·Y Triplex Structures from Friedreich's Ataxia. *Molecular Cell* **3**, 465–475 (1999).

35. Faucon, B., Mergny, J. L. & Héléne, C. Effect of third strand composition on the triple helix formation: purine versus pyrimidine oligodeoxynucleotides. *Nucleic Acids Res* **24**, 3181–3188 (1996).

36. Sakamoto, N. *et al.* GGA·TCC-interrupted Triplets in Long GAA·TTC Repeats Inhibit the Formation of Triplex and Sticky DNA Structures, Alleviate Transcription Inhibition, and Reduce Genetic Instabilities. *Journal of Biological Chemistry* **276**, 27178–27187 (2001).

37. Son, L. S., Bacolla, A. & Wells, R. D. Sticky DNA: in vivo formation in E. coli and in vitro association of long GAA*TTC tracts to generate two independent supercoiled domains. *J Mol Biol* **360**, 267–284 (2006).

38. Poggi, L. & Richard, G.-F. Alternative DNA Structures In Vivo: Molecular Evidence and Remaining Questions. *Microbiol. Mol. Biol. Rev.* **85**, (2021).

39. Agazie, Y. M., Lee, J. S. & Burkholder, G. D. Characterization of a new monoclonal antibody to triplex DNA and immunofluorescent staining of mammalian chromosomes. *Journal of Biological Chemistry* **269**, 7019–7023 (1994).

40. Agazie, Y. M., Burkholder, G. D. & Lee, J. S. Triplex DNA in the nucleus: direct binding of triplex-specific antibodies and their effect on transcription, replication and cell growth. *Biochem J* **316**, 461–466 (1996).

41. Ohno, M., Fukagawa, T., Lee, J. S. & Ikemura, T. Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies. *Chromosoma* **111**, 201–213 (2002).

42. Ma, J. & Wang, M. D. DNA supercoiling during transcription. *Biophys Rev* **8**, 75–87 (2016).

43. Kouzine, F. *et al.* Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Syst* **4**, 344-356.e7 (2017).

44. Wu, W. *et al.* Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature* **593**, 440–444 (2021).

45. Maekawa, K., Yamada, S., Sharma, R., Chaudhuri, J. & Keeney, S. Triple-helix potential of the mouse genome. *Proceedings of the National Academy of Sciences* **119**, e2203967119 (2022).

46. Matos-Rodrigues, G. *et al.* S1-END-seq reveals DNA secondary structures in human cells. *Mol Cell* **82**, 3538-3552.e5 (2022).

47. Samadashwily, G. M., Raca, G. & Mirkin, S. M. Trinucleotide repeats affect DNA replication in vivo. *Nat Genet* **17**, 298–304 (1997).

48. Gacy, A. M. *et al.* GAA instability in Friedreich's Ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol Cell* **1**, 583–593 (1998).

49. Casas-Delucchi, C. S., Daza-Martin, M., Williams, S. L. & Coster, G. The mechanism of replication stalling and recovery within repetitive DNA. *Nat Commun* **13**, 3953 (2022).

50. Chandok, G. S., Patel, M. P., Mirkin, S. M. & Krasilnikova, M. M. Effects of Friedreich's ataxia GAA repeats on DNA replication in mammalian cells. *Nucleic Acids Res* **40**, 3964–3974 (2012).

51. Follonier, C., Oehler, J., Herrador, R. & Lopes, M. Friedreich's ataxia–associated GAA repeats induce replication-fork reversal and unusual molecular junctions. *Nat Struct Mol Biol* **20**, 486–494 (2013).

52. Gerhardt, J. *et al.* Stalled DNA Replication Forks at the Endogenous GAA Repeats Drive Repeat Expansion in Friedreich's Ataxia Cells. *Cell Rep* **16**, 1218–1227 (2016).

53. Krasilnikova, M. M. & Mirkin, S. M. Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. *Mol Cell Biol* **24**, 2286–2295 (2004).

54. Ohshima, K., Montermini, L., Wells, R. D. & Pandolfo, M. Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. *J Biol Chem* **273**, 14588–14595 (1998).

55. Kim, H.-M. *et al.* Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J* **27**, 2896–2906 (2008).

56. Rastokina, A. *et al.* Large-scale expansions of Friedreich's ataxia GAA•TTC repeats in an experimental human system: role of DNA replication and prevention by LNA-DNA oligonucleotides and PNA oligomers. *Nucleic Acids Research* **51**, 8532–8549 (2023).

57. Fanning, E. & Zhao, K. SV40 DNA replication: From the A gene to a nanomachine. *Virology* **384**, 352–359 (2009).

58. Kumari, D., Hayward, B., Nakamura, A. J., Bonner, W. M. & Usdin, K. Evidence for chromosome fragility at the frataxin locus in Friedreich ataxia. *Mutat Res* **781**, 14–21 (2015).

59. Cheloshkina, K. & Poptsova, M. Comprehensive analysis of cancer breakpoints reveals signatures of genetic and epigenetic contribution to cancer genome rearrangements. *PLoS Comput Biol* **17**, e1008749 (2021).

60. Georgakopoulos-Soares, I., Morganella, S., Jain, N., Hemberg, M. & Nik-Zainal, S. Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis. *Genome Res.* **28**, 1264–1271 (2018).

61. Lyu, R. *et al.* KAS-seq: genome-wide sequencing of single-stranded DNA by N3-kethoxal–assisted labeling. *Nat Protoc* **17**, 402–420 (2022).

62. Wu, T., Lyu, R., You, Q. & He, C. Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ. *Nat Methods* **17**, 515–523 (2020).

63. Nethisinghe, S. *et al.* Interruptions of the FXN GAA Repeat Tract Delay the Age at Onset of Friedreich's Ataxia in a Location Dependent Manner. *Int J Mol Sci* **22**, 7507 (2021).

64. Khristich, A. N., Armenia, J. F., Matera, R. M., Kolchinski, A. A. & Mirkin, S. M. Large-scale contractions of Friedreich's ataxia GAA repeats in yeast occur during DNA replication due to their triplex-forming ability. *Proc Natl Acad Sci USA* **117**, 1628–1637 (2020).

65. Shishkin, A. A. *et al.* Large-Scale Expansions of Friedreich's Ataxia GAA Repeats in Yeast. *Molecular Cell* **35**, 82–92 (2009).

66. De Biase, I. *et al.* Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. *Genomics* **90**, 1–5 (2007).

67. De Michele, G. *et al.* Parental gender, age at birth and expansion length influence GAA repeat intergenerational instability in the X25 gene: pedigree studies and analysis of sperm from patients with Friedreich's Ataxia. *Human Molecular Genetics* **7**, 1901–1906 (1998).

68. McMurray, C. T. Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet* **11**, 786–799 (2010).

69. Pollard, L. M. *et al.* Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. *Nucleic Acids Research* **32**, 5962–5971 (2004).

70. Sharma, R. *et al.* The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability in vivo, with a significant predilection for large contractions. *Hum Mol Genet* **11**, 2175–2187 (2002).

71. Long, A. *et al.* Somatic instability of the expanded GAA repeats in Friedreich's ataxia. *PLoS One* **12**, e0189990 (2017).

72. Montermini, L., Kish, S. J., Jiralerspong, S., Lamarche, J. B. & Pandolfo, M. Somatic mosaicism for Friedreich's ataxia GAA triplet repeat expansions in the central nervous system. *Neurology* **49**, 606–610 (1997).

73. Hellenbroich, Y., Schwinger, E. & Zühlke, C. Limited somatic mosaicism for Friedreich's ataxia GAA triplet repeat expansions identified by small pool PCR in blood leukocytes. *Acta Neurol Scand* **103**, 188–192 (2001).

74. Machkhas, H., Bidichandani, S. I., Patel, P. I. & Harati, Y. A mild case of Friedreich ataxia: lymphocyte and sural nerve analysis for GAA repeat length reveals somatic mosaicism. *Muscle Nerve* **21**, 390–393 (1998).

75. Shah, K. A. & Mirkin, S. M. The hidden side of unstable DNA repeats: Mutagenesis at a distance. *DNA Repair* **32**, 106–112 (2015).

76. Bidichandani, S. I. *et al.* Somatic Sequence Variation at the Friedreich Ataxia Locus Includes Complete Contraction of the Expanded GAA Triplet Repeat, Significant Length Variation in Serially Passaged Lymphoblasts and Enhanced Mutagenesis in the Flanking Sequence. *Human Molecular Genetics* **8**, 2425–2436 (1999).

77. Wang, G. & Vasquez, K. M. Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc Natl Acad Sci U S A* **101**, 13448–13453 (2004).

78. Wojciechowska, M., Napierala, M., Larson, J. E. & Wells, R. D. Non-B DNA Conformations Formed by Long Repeating Tracts of Myotonic Dystrophy Type 1, Myotonic Dystrophy Type 2, and Friedreich's Ataxia Genes, Not the Sequences per se, Promote Mutagenesis in Flanking Regions *. *Journal of Biological Chemistry* **281**, 24531–24543 (2006).

79. Tang, W., Dominska, M., Gawel, M., Greenwell, P. W. & Petes, T. D. Genomic deletions and point mutations induced in Saccharomyces cerevisiae by the trinucleotide repeats (GAA·TTC) associated with Friedreich's ataxia. *DNA Repair* **12**, 10–17 (2013).

80. Zhao, J. *et al.* Distinct Mechanisms of Nuclease-Directed DNA-Structure-Induced Genetic Instability in Cancer Genomes. *Cell Rep* **22**, 1200–1210 (2018).

81. Wang, G. & Vasquez, K. M. Non-B DNA structure-induced genetic instability. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **598**, 103–119 (2006).

82. Zhang, Y. *et al.* Genome-wide Screen Identifies Pathways that Govern GAA/TTC Repeat Fragility and Expansions in Dividing and Nondividing Yeast Cells. *Molecular Cell* **48**, 254–265 (2012).

83. Shah, K. A. *et al.* Role of DNA Polymerases in Repeat-Mediated Genome Instability. *Cell Reports* **2**, 1088–1095 (2012).

84. Grabowska, E. *et al.* Proper functioning of the GINS complex is important for the fidelity of DNA replication in yeast. *Mol Microbiol* **92**, 659–680 (2014).

85. Denkiewicz-Kruk, M. *et al.* Recombination and Pol ζ Rescue Defective DNA Replication upon Impaired CMG Helicase—Pol ε Interaction. *International Journal of Molecular Sciences* **21**, 9484 (2020).

86. Neil, A. J., Kim, J. C. & Mirkin, S. M. Precarious maintenance of simple DNA repeats in eukaryotes. *Bioessays* **39**, (2017).

87. Mirkin, S. M. & Smirnova, E. V. Positioned to expand. *Nat Genet* **31**, 5–6 (2002).

88. Spivakovsky-Gonzalez, E. *et al.* Rad9-mediated checkpoint activation is responsible for elevated expansions of GAA repeats in CST-deficient yeast. *Genetics* **219**, (2021).

89. Lahiri, M., Gustafson, T. L., Majors, E. R. & Freudenreich, C. H. Expanded CAG Repeats Activate the DNA Damage Checkpoint Pathway. *Molecular Cell* **15**, 287–293 (2004).

90. Guo, M. *et al.* A distinct triplex DNA unwinding activity of ChlR1 helicase. *J Biol Chem* **290**, 5174–5189 (2015).

91. McGinty, R. J. *et al.* Nanopore sequencing of complex genomic rearrangements in yeast reveals mechanisms of repeat-mediated double-strand break repair. *Genome Res.* **27**, 2072–2082 (2017).

92. Chiu, R., Rajan-Babu, I.-S., Friedman, J. M. & Birol, I. Straglr: discovering and genotyping tandem repeat expansions using whole genome long-read sequences. *Genome Biology* **22**, 224 (2021).

93. De Roeck, A. *et al.* NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. *Genome Biol* **20**, 239 (2019).

94. Dolzhenko, E. *et al.* Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* **27**, 1895–1903 (2017).

95. Dolzhenko, E. *et al.* ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* **35**, 4754–4756 (2019).

96. Liu, Q., Zhang, P., Wang, D., Gu, W. & Wang, K. Interrogating the "unsequenceable" genomic trinucleotide repeat disorders by long-read sequencing. *Genome Medicine* **9**, 65 (2017).

97. Liu, Q., Tong, Y. & Wang, K. Genome-wide detection of short tandem repeat expansions by long-read sequencing. *BMC Bioinformatics* **21**, 542 (2020).

98. Mantere, T., Kersten, S. & Hoischen, A. Long-Read Sequencing Emerging in Medical Genetics. *Front. Genet.* **0**, (2019).

99. Boemo, M. A. DNAscent v2: detecting replication forks in nanopore sequencing data with deep learning. *BMC Genomics* **22**, 430 (2021).

100. Müller, C. A. *et al.* Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nature Methods* **16**, 429–436 (2019).

101. Ciosi, M. *et al.* A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine* **48**, 568–580 (2019).

102. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell* **162**, 516–526 (2015).

103. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell* **178**, 887-900.e14 (2019).

104. Saini, N. *et al.* Fragile DNA Motifs Trigger Mutagenesis at Distant Chromosomal Loci in Saccharomyces cerevisiae. *PLOS Genetics* **9**, e1003551 (2013).

105. Neil, A. J., Liang, M. U., Khristich, A. N., Shah, K. A. & Mirkin, S. M. RNA-DNA hybrids promote the expansion of Friedreich's ataxia (GAA)n repeats via break-induced replication. *Nucleic Acids Res* **46**, 3487–3497 (2018).

106. Bourn, R. L. *et al.* Pms2 suppresses large expansions of the (GAA·TTC)n sequence in neuronal tissues. *PLoS One* **7**, e47085 (2012).

107. Halabi, A., Ditch, S., Wang, J. & Grabczyk, E. DNA Mismatch Repair Complex MutSβ Promotes GAA·TTC Repeat Expansion in Human Cells. *J Biol Chem* **287**, 29958–29967 (2012).

108. Halabi, A., Fuselier, K. T. B. & Grabczyk, E. GAA•TTC repeat expansion in human cells is mediated by mismatch repair complex MutLγ and depends upon the endonuclease domain in MLH3 isoform one. *Nucleic Acids Res* **46**, 4022–4032 (2018).

109. Krasilnikova, M. M. *et al.* Effects of Friedreich's ataxia (GAA)n*(TTC)n repeats on RNA synthesis and stability. *Nucleic Acids Res* **35**, 1075–1084 (2007).

110. Bidichandani, S. I., Ashizawa, T. & Patel, P. I. The GAA Triplet-Repeat Expansion in Friedreich Ataxia Interferes with Transcription and May Be Associated with an Unusual DNA Structure. *The American Journal of Human Genetics* **62**, 111–121 (1998).

111. Reddy, K. *et al.* Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res* **39**, 1749–1762 (2011).

112. Zhang, J., Fakharzadeh, A., Pan, F., Roland, C. & Sagui, C. Atypical structures of GAA/TTC trinucleotide repeats underlying Friedreich's ataxia: DNA triplexes and RNA/DNA hybrids. *Nucleic Acids Res* **48**, 9899–9917 (2020).

113. Zeng, C., Onoguchi, M. & Hamada, M. Association analysis of repetitive elements and R-loop formation across species. *Mobile DNA* **12**, 3 (2021).

114. Li, H. *et al.* The Cumulative Formation of R-loop Interacts with Histone Modifications to Shape Cell Reprogramming. *International Journal of Molecular Sciences* **23**, 1567 (2022).

115. Freudenreich, C. H. R-loops: targets for nuclease cleavage and repeat instability. *Curr Genet* **64**, 789–794 (2018).

116. García-Muse, T. & Aguilera, A. R Loops: From Physiological to Pathological Roles. *Cell* **179**, 604–618 (2019).

117. Groh, M., Lufino, M. M. P., Wade-Martins, R. & Gromak, N. R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome. *PLOS Genetics* **10**, e1004318 (2014).

118. Belotserkovskii, B. P. *et al.* Transcription blockage by homopurine DNA sequences: role of sequence composition and single-strand breaks. *Nucleic Acids Research* **41**, 1817–1828 (2013).

119. Belotserkovskii, B. P. *et al.* Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc Natl Acad Sci U S A* **107**, 12816–12821 (2010).

120. Ditch, S., Sammarco, M. C., Banerjee, A. & Grabczyk, E. Progressive GAA.TTC repeat expansion in human cell lines. *PLoS Genet* **5**, e1000704 (2009).

121. Shah, K. A., McGinty, R. J., Egorova, V. I. & Mirkin, S. M. Coupling transcriptional state to large-scale repeat expansions in yeast. *Cell Rep* **9**, 1594–1602 (2014).

122. McGinty, R. J. *et al.* A Defective mRNA Cleavage and Polyadenylation Complex Facilitates Expansions of Transcribed (GAA)n Repeats Associated with Friedreich's Ataxia. *Cell Rep* **20**, 2490–2500 (2017).

123. Zardoni, L. *et al.* Elongating RNA polymerase II and RNA:DNA hybrids hinder fork progression and gene expression at sites of head-on replication-transcription collisions. *Nucleic Acids Res* **49**, 12769–12784 (2021).

124. Alzu, A. *et al.* Senataxin Associates with Replication Forks to Protect Fork Integrity across RNA-Polymerase-II-Transcribed Genes. *Cell* **151**, 835–846 (2012).

125. Cohen, S. *et al.* Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nat Commun* **9**, 533 (2018).

126. Nakatani, R., Nakamori, M., Fujimura, H., Mochizuki, H. & Takahashi, M. P. Large expansion of CTG•CAG repeats is exacerbated by MutSβ in human cells. *Sci Rep* **5**, 11020 (2015).

127. Appanah, R., Lones, E. C., Aiello, U., Libri, D. & De Piccoli, G. Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability. *Cell Rep* **30**, 2094-2105.e9 (2020).

128. Al-Mahdawi, S. *et al.* GAA repeat instability in Friedreich ataxia YAC transgenic mice. *Genomics* **84**, 301–310 (2004).

129. Anjomani Virmouni, S., Sandi, C., Al-Mahdawi, S. & Pook, M. A. Cellular, molecular and functional characterisation of YAC transgenic mouse models of Friedreich ataxia. *PLoS One* **9**, e107416 (2014).

130. Anjomani Virmouni, S. *et al.* Identification of telomere dysfunction in Friedreich ataxia. *Mol Neurodegener* **10**, 22 (2015).

131. Clark, R. M. *et al.* The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. *Hum Genet* **120**, 633–640 (2007).

132. De Biase, I. *et al.* Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients. *Ann Neurol* **61**, 55–60 (2007).

133. Konopka, A. & Atkin, J. D. The Role of DNA Damage in Neural Plasticity in Physiology and Neurodegeneration. *Frontiers in Cellular Neuroscience* **16**, (2022).

134. Li, X. *et al.* Polymerases and DNA Repair in Neurons: Implications in Neuronal Survival and Neurodegenerative Diseases. *Frontiers in Cellular Neuroscience* **16**, (2022).

135. Merlo, D. *et al.* DNA repair in post-mitotic neurons: a gene-trapping strategy. *Cell Death Differ* **12**, 307–309 (2005).

136. Schwartz, E. I. *et al.* Cell cycle activation in postmitotic neurons is essential for DNA repair. *Cell Cycle* **6**, 318–329 (2007).

137. Shanbhag, N. M. *et al.* Early neuronal accumulation of DNA double strand breaks in Alzheimer's disease. *Acta Neuropathol Commun* **7**, 77 (2019).

138. Reid, D. A. *et al.* Incorporation of a nucleoside analog maps genome repair sites in postmitotic human neurons. *Science* **372**, 91–94 (2021).

139. Pećina-Šlaus, N., Kafka, A., Salamon, I. & Bukovac, A. Mismatch Repair Pathway, Genome Stability and Cancer. *Frontiers in Molecular Biosciences* **7**, 122 (2020).

140. Richard, G.-F. The Startling Role of Mismatch Repair in Trinucleotide Repeat Expansions. *Cells* **10**, 1019 (2021).

141. Schmidt, M. H. M. & Pearson, C. E. Disease-associated repeat instability and mismatch repair. *DNA Repair* **38**, 117–126 (2016).

142. Ezzatizadeh, V. *et al.* The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model. *Neurobiol Dis* **46**, 165–171 (2012).

143. Ezzatizadeh, V. *et al.* MutLα heterodimers modify the molecular phenotype of Friedreich ataxia. *PLoS One* **9**, e100523 (2014).

144. Ku, S. *et al.* Friedreich's ataxia induced pluripotent stem cells model intergenerational GAA·TTC triplet repeat instability. *Cell Stem Cell* **7**, 631–637 (2010).

145. Flower, M. *et al.* MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain* **142**, 1876–1886 (2019).

146. Flower, M. *et al.* Reply: The repeat variant in MSH3 is not a genetic modifier for spinocerebellar ataxia type 3 and Friedreich's ataxia. *Brain* **143**, e26 (2020).

147. Mirkin, S. M. Getting to the Core of Repeat Expansions by Cell Reprogramming. *Cell Stem Cell* **7**, 545–546 (2010).

148. Yau, W. Y. *et al.* The repeat variant in MSH3 is not a genetic modifier for spinocerebellar ataxia type 3 and Friedreich's ataxia. *Brain* **143**, e25 (2020).

149. Neil, A. J. *et al.* Replication-independent instability of Friedreich's ataxia GAA repeats during chronological aging. *PNAS* **118**, (2021).

150. Panigrahi, G. B., Lau, R., Montgomery, S. E., Leonard, M. R. & Pearson, C. E. Slipped (CTG)•(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. *Nat Struct Mol Biol* **12**, 654–662 (2005).

151. Guo, J., Gu, L., Leffak, M. & Li, G.-M. MutSβ promotes trinucleotide repeat expansion by recruiting DNA polymerase β to nascent (CAG)n or (CTG)n hairpins for error-prone DNA synthesis. *Cell Res* **26**, 775–786 (2016).

152. Iyer, R. R. & Pluciennik, A. DNA Mismatch Repair and its Role in Huntington's Disease. *J Huntingtons Dis* **10**, 75–94 (2021).

153. Kadyrova, L. Y., Gujar, V., Burdett, V., Modrich, P. L. & Kadyrov, F. A. Human MutLγ, the MLH1–MLH3 heterodimer, is an endonuclease that promotes DNA expansion. *Proc Natl Acad Sci USA* **117**, 3535–3542 (2020).

154. Owen, B. A. L. *et al.* (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat Struct Mol Biol* **12**, 663–670 (2005).

155. Zhang, T., Huang, J., Gu, L. & Li, G.-M. In vitro repair of DNA hairpins containing various numbers of CAG/CTG trinucleotide repeats. *DNA Repair (Amst)* **11**, 201–209 (2012).

156. Pinto, R. M. *et al.* Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. *PLoS Genet* **9**, e1003930 (2013).

157. Lee, J.-M. *et al.* A modifier of Huntington's disease onset at the MLH1 locus. *Human Molecular Genetics* **26**, 3859–3867 (2017).

158. Su, X. A. & Freudenreich, C. H. Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **114**, E8392–E8401 (2017).

159. Gomes-Pereira, M., Fortune, M. T., Ingram, L., McAbney, J. P. & Monckton, D. G. Pms2 is a genetic enhancer of trinucleotide CAG·CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Human Molecular Genetics* **13**, 1815–1825 (2004).

160. Roy, J. C. L. *et al.* Somatic CAG expansion in Huntington's disease is dependent on the MLH3 endonuclease domain, which can be excluded via splice redirection. *Nucleic Acids Research* **49**, 3907–3918 (2021).

161. Du, J. *et al.* Role of mismatch repair enzymes in GAA·TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells. *J Biol Chem* **287**, 29861–29872 (2012).

162. Kovalenko, M. *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**, e44273 (2012).

163. Manley, K., Shirley, T. L., Flaherty, L. & Messer, A. Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genetics* **23**, 471–473 (1999).

164. Kantartzis, A. *et al.* Msh2-Msh3 interferes with Okazaki fragment processing to promote trinucleotide repeat expansions. *Cell Rep* **2**, 216–222 (2012).

165. Tomé, S. *et al.* MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice. *PLoS Genet* **5**, e1000482 (2009).

166. Dragileva, E. *et al.* Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol Dis* **33**, 37–47 (2009).

167. Foiry, L. *et al.* Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Hum Genet* **119**, 520–526 (2006).

168. Keogh, N., Chan, K. Y., Li, G.-M. & Lahue, R. S. MutSβ abundance and Msh3 ATP hydrolysis activity are important drivers of CTG•CAG repeat expansions. *Nucleic Acids Res* **45**, 10068–10078 (2017).

169. Tomé, S. *et al.* MSH3 Polymorphisms and Protein Levels Affect CAG Repeat Instability in Huntington's Disease Mice. *PLOS Genetics* **9**, e1003280 (2013).

170. Williams, G. M. & Surtees, J. A. MSH3 Promotes Dynamic Behavior of Trinucleotide Repeat Tracts In Vivo. *Genetics* **200**, 737–754 (2015).

171. Cilli, P. *et al.* Oxidized dNTPs and the OGG1 and MUTYH DNA glycosylases combine to induce CAG/CTG repeat instability. *Nucleic Acids Research* **44**, 5190–5203 (2016).

172. Kovtun, I. V. *et al.* OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature* **447**, 447–452 (2007).

173. Lai, Y., Beaver, J. M., Laverde, E. & Liu, Y. Trinucleotide repeat instability via DNA base excision repair. *DNA Repair (Amst)* **93**, 102912 (2020).

174. Lai, Y. *et al.* Base excision repair of chemotherapeutically-induced alkylated DNA damage predominantly causes contractions of expanded GAA repeats associated with Friedreich's ataxia. *PLoS One* **9**, e93464 (2014).

175. Laverde, E. E. *et al.* R-loops promote trinucleotide repeat deletion through DNA base excision repair enzymatic activities. *Journal of Biological Chemistry* **295**, 13902–13913 (2020).

176. Lai, Y., Xu, M., Zhang, Z. & Liu, Y. Instability of CTG Repeats is Governed by the Position of a DNA Base Lesion through Base Excision Repair. *PLOS ONE* **8**, e56960 (2013).

177. La Rosa, P., Petrillo, S., Bertini, E. S. & Piemonte, F. Oxidative Stress in DNA Repeat Expansion Disorders: A Focus on NRF2 Signaling Involvement. *Biomolecules* **10**, 702 (2020).

178. Cobley, J. N., Fiorello, M. L. & Bailey, D. M. 13 reasons why the brain is susceptible to oxidative stress. *Redox Biology* **15**, 490–503 (2018).

179. Lupoli, F., Vannocci, T., Longo, G., Niccolai, N. & Pastore, A. The role of oxidative stress in Friedreich's ataxia. *FEBS Letters* **592**, 718–727 (2018).

180. Tang, W. *et al.* Friedreich's ataxia (GAA)n•(TTC)n repeats strongly stimulate mitotic crossovers in Saccharomyces cerevisae. *PLoS Genet* **7**, e1001270 (2011).

181. Napierala, M., Dere, R., Vetcher, A. & Wells, R. D. Structure-dependent recombination hot spot activity of GAA.TTC sequences from intron 1 of the Friedreich's ataxia gene. *J Biol Chem* **279**, 6444–6454 (2004).

182. Polleys, E. J. & Freudenreich, C. H. Homologous recombination within repetitive DNA. *Curr Opin Genet Dev* **71**, 143–153 (2021).

183. Balakrishnan, L. & Bambara, R. A. Flap Endonuclease 1. *Annual Review of Biochemistry* **82**, 119–138 (2013).

184. Michl, J., Zimmer, J. & Tarsounas, M. Interplay between Fanconi anemia and homologous recombination pathways in genome integrity. *EMBO J* **35**, 909–923 (2016).

185. Yang, J. & Freudenreich, C. H. Haploinsufficiency of Yeast FEN1 Causes Instability of Expanded CAG/CTG Tracts in a Length-Dependent Manner. *Gene* **393**, 110–115 (2007).

186. Freudenreich, C. H., Stavenhagen, J. B. & Zakian, V. A. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol Cell Biol* **17**, 2090–2098 (1997).

187. Tsutakawa, S. E. *et al.* Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. *Nat Commun* **8**, 15855 (2017).

188. Liu, Y., Zhang, H., Veeraraghavan, J., Bambara, R. A. & Freudenreich, C. H. Saccharomyces cerevisiae Flap Endonuclease 1 Uses Flap Equilibration to Maintain Triplet Repeat Stability. *Molecular and Cellular Biology* **24**, 4049–4064 (2004).

189. van den Broek, W. J. A. A., Nelen, M. R., van der Heijden, G. W., Wansink, D. G. & Wieringa, B. Fen1 does not control somatic hypermutability of the (CTG)(n)*(CAG)(n) repeat in a knock-in mouse model for DM1. *FEBS Lett* **580**, 5208–5214 (2006).

190. Spiro, C. & McMurray, C. T. Nuclease-deficient FEN-1 blocks Rad51/BRCA1-mediated repair and causes trinucleotide repeat instability. *Mol Cell Biol* **23**, 6063–6074 (2003).

191. Moe, S. E., Sorbo, J. G. & Holen, T. Huntingtin triplet-repeat locus is stable under long-term Fen1 knockdown in human cells. *J Neurosci Methods* **171**, 233–238 (2008).

192. Otto, C. J., Almqvist, E., Hayden, M. R. & Andrew, S. E. The 'flap' endonuclease gene FEN1 is excluded as a candidate gene implicated in the CAG repeat expansion underlying Huntington disease. *Clin Genet* **59**, 122–127 (2001).

193. Liu, G., Chen, X., Bissler, J. J., Sinden, R. R. & Leffak, M. Replication-dependent instability at (CTG)•(CAG) repeat hairpins in human cells. *Nat Chem Biol* **6**, 652–659 (2010).

194. Kim, K.-H. *et al.* Genetic and Functional Analyses Point to FAN1 as the Source of Multiple Huntington Disease Modifier Effects. *Am J Hum Genet* **107**, 96–110 (2020).

195. Deshmukh, A. L. *et al.* FAN1 exo- not endo-nuclease pausing on disease-associated slipped-DNA repeats: A mechanism of repeat instability. *Cell Reports* **37**, 110078 (2021).

196. Trost, B. *et al.* Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature* **586**, 80–86 (2020).

197. Zhao, X.-N. & Usdin, K. FAN1 protects against repeat expansions in a Fragile X mouse model. *DNA Repair (Amst)* **69**, 1–5 (2018).

198. Goold, R. *et al.* FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. *Hum Mol Genet* **28**, 650–661 (2019).

199. Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R. & Jiricny, J. Characterization of the interactome of the human MutL homologues MLH1, PMS1, and PMS2. *J Biol Chem* **282**, 2976–2986 (2007).

200. Smogorzewska, A. *et al.* A genetic screen identifies FAN1, a Fanconi anemia associated nuclease necessary for DNA interstrand crosslink repair. *Mol Cell* **39**, 36–47 (2010).

201. Goold, R. *et al.* FAN1 controls mismatch repair complex assembly via MLH1 retention to stabilize CAG repeat expansion in Huntington's disease. *Cell Rep* **36**, 109649 (2021).

202. Lahue, R. S. SPYing on triplet repeat expansions: Insights into FAN1-MLH1 interaction and regulation. *Cell Reports* **36**, 109736 (2021).

203. Loupe, J. M. *et al.* Promotion of somatic CAG repeat expansion by Fan1 knock-out in Huntington's disease knock-in mice is blocked by Mlh1 knock-out. *Hum Mol Genet* **29**, 3044–3053 (2020).

204. Porro, A. *et al.* FAN1-MLH1 interaction affects repair of DNA interstrand cross-links and slipped-CAG/CTG repeats. *Sci Adv* **7**, eabf7906 (2021).

205. McAllister, B. *et al.* Exome sequencing of individuals with Huntington's disease implicates FAN1 nuclease activity in slowing CAG expansion and disease onset. *Nat Neurosci* **25**, 446–457 (2022).

206. Deshmukh, A. L. *et al.* FAN1, a DNA Repair Nuclease, as a Modifier of Repeat Expansion Disorders. *J Huntingtons Dis* **10**, 95–122 (2021).

207. Calil, F. A. *et al.* Rad27 and Exo1 function in different excision pathways for mismatch repair in Saccharomyces cerevisiae. *Nat Commun* **12**, 5568 (2021).

208. Li, Y. *et al.* Excision of Expanded GAA Repeats Alleviates the Molecular Phenotype of Friedreich's Ataxia. *Molecular Therapy* **23**, 1055–1065 (2015).

209. Oura, S. *et al.* Precise CAG repeat contraction in a Huntington's Disease mouse model is enabled by gene editing with SpCas9-NG. *Commun Biol* **4**, 1–13 (2021).

210. Rocca, C. J. *et al.* CRISPR-Cas9 Gene Editing of Hematopoietic Stem Cells from Patients with Friedreich's Ataxia. *Mol Ther Methods Clin Dev* **17**, 1026–1036 (2020).

211. Ouellet, D. L., Cherif, K., Rousseau, J. & Tremblay, J. P. Deletion of the GAA repeats from the human frataxin gene using the CRISPR-Cas9 system in YG8R-derived cells and mouse models of Friedreich ataxia. *Gene Ther* **24**, 265–274 (2017).

212. Anjomani Virmouni, S. *et al.* A novel GAA-repeat-expansion-based mouse model of Friedreich's ataxia. *Dis Model Mech* **8**, 225–235 (2015).

213. Bird, M. J. *et al.* Functional characterization of Friedreich ataxia iPS-derived neuronal progenitors and their integration in the adult brain. *PLoS One* **9**, e101718 (2014).

214. Viventi, S. *et al.* In vivo survival and differentiation of Friedreich ataxia iPSC-derived sensory neurons transplanted in the adult dorsal root ganglia. *Stem Cells Transl Med* **10**, 1157–1169 (2021).

215. Mazzara, P. G. *et al.* Frataxin gene editing rescues Friedreich's ataxia pathology in dorsal root ganglia organoid-derived sensory neurons. *Nat Commun* **11**, 4178 (2020).

216. Sivakumar, A. & Cherqui, S. Advantages and Limitations of Gene Therapy and Gene Editing for Friedreich's Ataxia. *Frontiers in Genome Editing* **4**, (2022).

217. Ribeil, J.-A. *et al.* Gene Therapy in a Patient with Sickle Cell Disease. *N Engl J Med* **376**, 848–855 (2017).

218. Jacobson, S. G. *et al.* Safety and improved efficacy signals following gene therapy in childhood blindness caused by GUCY2D mutations. *iScience* **24**, 102409 (2021).

219. Nakamori, M. *et al.* A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat Genet* **52**, 146–159 (2020).

220. Li, Y. *et al.* Targeting 3′ and 5′ untranslated regions with antisense oligonucleotides to stabilize frataxin mRNA and increase protein expression. *Nucleic Acids Research* **49**, 11560–11574 (2021).

221. Lazaropoulos, M. *et al.* Frataxin levels in peripheral tissue in Friedreich ataxia. *Ann Clin Transl Neurol* **2**, 831–842 (2015).

222. Pianese, L. *et al.* Real time PCR quantification of frataxin mRNA in the peripheral blood leucocytes of Friedreich ataxia patients and carriers. *J Neurol Neurosurg Psychiatry* **75**, 1061–1063 (2004).

223. Bergquist, H. *et al.* Disruption of Higher Order DNA Structures in Friedreich's Ataxia (GAA)n Repeats by PNA or LNA Targeting. *PLOS ONE* **11**, e0165788 (2016).

224. Rastokina, A. *et al.* Large-scale expansions of Friedreich's ataxia GAA•TTC repeats in human cells are prevented by LNA-DNA oligonucleotides and PNA oligomers. 2022.07.04.498742 Preprint at https://doi.org/10.1101/2022.07.04.498742 (2022).

225. Grant, L. *et al.* Rational selection of small molecules that increase transcription through the GAA repeats found in Friedreich's ataxia. *FEBS Lett* **580**, 5399–5405 (2006).

226. Lufino, M. M. P. *et al.* A GAA repeat expansion reporter model of Friedreich's ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds. *Hum Mol Genet* **22**, 5173–5187 (2013).

227. Burnett, R. *et al.* DNA sequence-specific polyamides alleviate transcription inhibition associated with long GAA·TTC repeats in Friedreich's ataxia. *Proceedings of the National Academy of Sciences* **103**, 11497–11502 (2006).

228. Nance, E., Pun, S. H., Saigal, R. & Sellers, D. L. Drug delivery to the central nervous system. *Nat Rev Mater* **7**, 314–331 (2022).

229. Erwin, G. S. *et al.* Recurrent repeat expansions in human cancer genomes. *Nature* 1–7 (2022) doi:10.1038/s41586-022-05515-1.

230. Pellerin, D. *et al.* Intronic FGF14 GAA repeat expansions are a common cause of ataxia syndromes with neuropathy and bilateral vestibulopathy. *J Neurol Neurosurg Psychiatry* (2023) doi:10.1136/jnnp-2023-331490.

231. Rafehi, H. *et al.* An intronic GAA repeat expansion in FGF14 causes the autosomal-dominant adult-onset ataxia SCA50/ATX-FGF14. *The American Journal of Human Genetics* (2022) doi:10.1016/j.ajhg.2022.11.015.

232. Masnovo, C., Lobo, A. F. & Mirkin, S. M. Replication dependent and independent mechanisms of GAA repeat instability. *DNA Repair* **118**, 103385 (2022).

233. Bourn, R. L., Rindler, P. M., Pollard, L. M. & Bidichandani, S. I. E. coli mismatch repair acts downstream of replication fork stalling to stabilize the expanded (GAA.TTC)(n) sequence. *Mutat Res* **661**, 71–77 (2009).

234. Jedrychowska, M. *et al.* Defects in the GINS complex increase the instability of repetitive sequences via a recombination-dependent mechanism. *PLoS Genet* **15**, e1008494 (2019).

235. Lahue, R. S. & Slater, D. L. DNA repair and trinucleotide repeat instability. *Front Biosci* **8**, s653-665 (2003).

236. Amparo, C. *et al.* Duplex DNA from Sites of Helicase-Polymerase Uncoupling Links Non-B DNA Structure Formation to Replicative Stress. *Cancer Genomics & Proteomics* **17**, 101–115 (2020).

237. Devbhandari, S. & Remus, D. Rad53 limits CMG helicase uncoupling from DNA synthesis at replication forks. *Nat Struct Mol Biol* **27**, 461–471 (2020).

238. Ercilla, A. *et al.* Physiological Tolerance to ssDNA Enables Strand Uncoupling during DNA Replication. *Cell Reports* **30**, 2416-2429.e7 (2020).

239. Lopes, M., Foiani, M. & Sogo, J. M. Multiple Mechanisms Control Chromosome Integrity after Replication Fork Uncoupling and Restart at Irreparable UV Lesions. *Molecular Cell* **21**, 15–27 (2006).

240. Pagès, V. & Fuchs, R. P. Uncoupling of Leading- and Lagging-Strand DNA Replication During Lesion Bypass in Vivo. *Science* (2003) doi:10.1126/science.1083964.

241. Taylor, M. R. G. & Yeeles, J. T. P. Dynamics of Replication Fork Progression Following Helicase–Polymerase Uncoupling in Eukaryotes. *Journal of Molecular Biology* **431**, 2040–2049 (2019).

242. Kavlashvili, T., Liu, W., Mohamed, T. M., Cortez, D. & Dewar, J. M. Replication fork uncoupling causes nascent strand degradation and fork reversal. *Nat Struct Mol Biol* 1–10 (2023) doi:10.1038/s41594-022-00871-y.

243. Byun, T. S., Pacek, M., Yee, M., Walter, J. C. & Cimprich, K. A. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev.* **19**, 1040–1052 (2005).

244. Nedelcheva, M. N. *et al.* Uncoupling of unwinding from DNA synthesis implies regulation of MCM helicase by Tof1/Mrc1/Csm3 checkpoint complex. *J Mol Biol* **347**, 509–521 (2005).

245. Saldanha, J., Rageul, J., Patel, J. A. & Kim, H. The Adaptive Mechanisms and Checkpoint Responses to a Stressed DNA Replication Fork. *International Journal of Molecular Sciences* **24**, 10488 (2023).

246. Fanning, E., Klimovich, V. & Nager, A. R. A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res* **34**, 4126–4137 (2006).

247. Pellegrini, L. The CMG DNA helicase and the core replisome. *Current Opinion in Structural Biology* **81**, 102612 (2023).

248. Solomon, N. A. *et al.* Genetic and molecular analysis of DNA43 and DNA52: two new cell-cycle genes in Saccharomyces cerevisiae. *Yeast* **8**, 273–289 (1992).

249. Izumi, M. *et al.* The human homolog of Saccharomyces cerevisiae Mcm10 interacts with replication factors and dissociates from nuclease-resistant nuclear structures in G(2) phase. *Nucleic Acids Res* **28**, 4769–4777 (2000).

250. Thu, Y. M. & Bielinsky, A.-K. MCM10: one tool for all - integrity, maintenance and damage control. *Semin Cell Dev Biol* **0**, 121–130 (2014).

251. Yuan, Z. *et al.* Ctf4 organizes sister replisomes and Pol α into a replication factory. *eLife* **8**, e47405 (2019).

252. Porcella, S. Y. *et al. Separable, Ctf4-Mediated Recruitment of DNA Polymerase α for Initiation of DNA Synthesis at Replication Origins and Lagging-Strand Priming during Replication Elongation*. http://biorxiv.org/lookup/doi/10.1101/352567 (2018) doi:10.1101/352567.

253. Homesley, L. *et al.* Mcm10 and the MCM2–7 complex interact to initiate DNA synthesis and to release replication factors from origins. *Genes Dev* **14**, 913–926 (2000).

254. Merchant, A. M., Kawasaki, Y., Chen, Y., Lei, M. & Tye, B. K. A lesion in the DNA replication initiation factor Mcm10 induces pausing of elongation forks through chromosomal replication origins in Saccharomyces cerevisiae. *Mol Cell Biol* **17**, 3261–3271 (1997).

255. Ricke, R. M. & Bielinsky, A.-K. Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. *Mol Cell* **16**, 173–185 (2004).

256. Chattopadhyay, S. & Bielinsky, A.-K. Human Mcm10 Regulates the Catalytic Subunit of DNA Polymerase-α and Prevents DNA Damage during Replication. *MBoC* **18**, 4085–4095 (2007).

257. Zhu, W. *et al.* Mcm10 and And-1/CTF4 recruit DNA polymerase α to chromatin for initiation of DNA replication. *Genes Dev.* **21**, 2288–2299 (2007).

258. Fien, K. *et al.* Primer Utilization by DNA Polymerase α-Primase Is Influenced by Its Interaction with Mcm10p*. *Journal of Biological Chemistry* **279**, 16144–16153 (2004).

259. Simon, A. C. *et al.* A Ctf4 trimer couples the CMG helicase to DNA polymerase α in the eukaryotic replisome. *Nature* **510**, 293–297 (2014).

260. Villa, F. *et al.* Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase. *Molecular Cell* **63**, 385–396 (2016).

261. Im, J.-S. *et al.* Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. *Proc Natl Acad Sci U S A* **106**, 15628–15632 (2009).

262. Perez-Arnaiz, P. & Kaplan, D. L. An Mcm10 Mutant Defective in ssDNA Binding Shows Defects in DNA Replication Initiation. *J Mol Biol* **428**, 4608–4625 (2016).

263. Warren, E. M., Huang, H., Fanning, E., Chazin, W. J. & Eichman, B. F. Physical Interactions between Mcm10, DNA, and DNA Polymerase α*. *Journal of Biological Chemistry* **284**, 24662–24672 (2009).

264. Das-Bradoo, S., Ricke, R. M. & Bielinsky, A.-K. Interaction between PCNA and diubiquitinated Mcm10 is essential for cell growth in budding yeast. *Mol Cell Biol* **26**, 4806–4817 (2006).

265. Wang, J., Wu, R., Lu, Y. & Liang, C. Ctf4p facilitates Mcm10p to promote DNA replication in budding yeast. *Biochemical and Biophysical Research Communications* **395**, 336–341 (2010).

266. Jones, M. L., Aria, V., Baris, Y. & Yeeles, J. T. P. How Pol α-primase is targeted to replisomes to prime eukaryotic DNA replication. *Molecular Cell* (2023) doi:10.1016/j.molcel.2023.06.035.

267. Baris, Y., Taylor, M. R. G., Aria, V. & Yeeles, J. T. P. Fast and efficient DNA replication with purified human proteins. *Nature* **606**, 204–210 (2022).

268. Wasserman, M. R., Schauer, G. D., O'Donnell, M. E. & Liu, S. Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. *Cell* **178**, 600-611.e16 (2019).

269. Langston, L. D., Georgescu, R. E. & O'Donnell, M. E. Mechanism of eukaryotic origin unwinding is a dual helicase DNA shearing process. *Proceedings of the National Academy of Sciences* **120**, e2316466120 (2023).

270. Henrikus, S. S. *et al.* Unwinding of a eukaryotic origin of replication visualized by cryo-EM. *Nat Struct Mol Biol* **31**, 1265–1276 (2024).

271. Langston, L. D. *et al.* Mcm10 promotes rapid isomerization of CMG-DNA for replisome bypass of lagging strand DNA blocks. *eLife* **6**, e29118 (2017).

272. Lõoke, M., Maloney, M. F. & Bell, S. P. Mcm10 regulates DNA replication elongation by stimulating the CMG replicative helicase. *Genes Dev.* **31**, 291–305 (2017).

273. Kawasaki, Y., Hiraga, S. & Sugino, A. Interactions between Mcm10p and other replication factors are required for proper initiation and elongation of chromosomal DNA replication in Saccharomyces cerevisiae. *Genes Cells* **5**, 975–989 (2000).

274. Lee, C., Liachko, I., Bouten, R., Kelman, Z. & Tye, B. K. Alternative Mechanisms for Coordinating Polymerase α and MCM Helicase. *Molecular and Cellular Biology* **30**, 423–435 (2010).

275. Alver, R. C. *et al.* The N-terminus of Mcm10 is important for interaction with the 9-1-1 clamp and in resistance to DNA damage. *Nucleic Acids Res* **42**, 8389–8404 (2014).

276. Chadha, G. S., Gambus, A., Gillespie, P. J. & Blow, J. J. Xenopus Mcm10 is a CDK-substrate required for replication fork stability. *Cell Cycle* **15**, 2183–2195 (2016).

277. Thu, Y. M. *et al.* Slx5/Slx8 Promotes Replication Stress Tolerance by Facilitating Mitotic Progression. *Cell Rep* **15**, 1254–1265 (2016).

278. Schmit, M. M. *et al.* A critical threshold of MCM10 is required to maintain genome stability during differentiation of induced pluripotent stem cells into natural killer cells. *Open Biology* **14**, 230407 (2024).

279. Yu, X. & Gabriel, A. Patching Broken Chromosomes with Extranuclear Cellular DNA. *Molecular Cell* **4**, 873–881 (1999).

280. Radchenko, E. A., McGinty, R. J., Aksenova, A. Y., Neil, A. J. & Mirkin, S. M. Quantitative Analysis of the Rates for Repeat-Mediated Genome Instability in a Yeast Experimental System. *Methods Mol Biol* **1672**, 421–438 (2018).

281. Thu, Y. M. & Bielinsky, A.-K. Enigmatic roles of Mcm10 in DNA replication. *Trends in Biochemical Sciences* **38**, 184–194 (2013).

282. Mayle, R. *et al.* Mcm10 has potent strand-annealing activity and limits translocase-mediated fork regression. *PNAS* **116**, 798–803 (2019).

283. Bochman, M. L. & Schwacha, A. The Mcm Complex: Unwinding the Mechanism of a Replicative Helicase. *Microbiol Mol Biol Rev* **73**, 652–683 (2009).

284. Ricke, R. M. & Bielinsky, A.-K. A Conserved Hsp10-like Domain in Mcm10 Is Required to Stabilize the Catalytic Subunit of DNA Polymerase-α in Budding Yeast. *Journal of Biological Chemistry* **281**, 18414–18425 (2006).

285. Park, J. H., Bang, S. W., Jeon, Y., Kang, S. & Hwang, D. S. Knockdown of human MCM10 exhibits delayed and incomplete chromosome replication. *Biochemical and Biophysical Research Communications* **365**, 575–582 (2008).

286. Dovrat, D. *et al.* A Live-Cell Imaging Approach for Measuring DNA Replication Rates. *Cell Reports* **24**, 252–258 (2018).

287. Bianchi, V., Pontis, E. & Reichard, P. Changes of deoxyribonucleoside triphosphate pools induced by hydroxyurea and their relation to DNA synthesis. *Journal of Biological Chemistry* **261**, 16037–16042 (1986).

288. Shiomi, Y. & Nishitani, H. Control of Genome Integrity by RFC Complexes; Conductors of PCNA Loading onto and Unloading from Chromatin during DNA Replication. *Genes (Basel)* **8**, 52 (2017).

289. Beckwith, W. H. *et al.* Destabilized PCNA Trimers Suppress Defective Rfc1 Proteins in Vivo and in Vitro. *Biochemistry* **37**, 3711–3722 (1998).

290. Beckwith, W. & McAlear, M. A. Allele-specific interactions between the yeast RFC1 and RFC5 genes suggest a basis for RFC subunit-subunit interactions. *Mol Gen Genet* **264**, 378–391 (2000).

291. McAlear, M. A., Howell, E. A., Espenshade, K. K. & Holm, C. Proliferating cell nuclear antigen (pol30) mutations suppress cdc44 mutations and identify potential regions of interaction between the two encoded proteins. *Mol Cell Biol* **14**, 4390–4397 (1994).

292. Becker, J. R., Nguyen, H. D., Wang, X. & Bielinsky, A.-K. Mcm10 deficiency causes defective-replisome-induced mutagenesis and a dependency on error-free postreplicative repair. *Cell Cycle* **13**, 1737–1748 (2014).

293. Putnam, C. D., Pennaneach, V. & Kolodner, R. D. Saccharomyces cerevisiae as a Model System To Define the Chromosomal Instability Phenotype. *Mol Cell Biol* **25**, 7226–7238 (2005).

294. Yan, Z. *et al.* Rad52 Restrains Resection at DNA Double-Strand Break Ends in Yeast. *Molecular Cell* **76**, 699-711.e6 (2019).

295. Fuchs, J., Cheblal, A. & Gasser, S. M. Underappreciated Roles of DNA Polymerase δ in Replication Stress Survival. *Trends in Genetics* (2021) doi:10.1016/j.tig.2020.12.003.

296. Pavlov, Y. I., Shcherbakova, P. V. & Kunkel, T. A. In vivo consequences of putative active site mutations in yeast DNA polymerases alpha, epsilon, delta, and zeta. *Genetics* **159**, 47–64 (2001).

297. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14**, 953–961 (1998).

298. Saxena, S. & Zou, L. Hallmarks of DNA replication stress. *Mol Cell* **82**, 2298–2314 (2022).

299. Bacal, J. *et al.* Mrc1 and Rad9 cooperate to regulate initiation and elongation of DNA replication in response to DNA damage. *The EMBO Journal* **37**, e99319 (2018).

300. Granata, M. *et al.* Dynamics of Rad9 Chromatin Binding and Checkpoint Function Are Mediated by Its Dimerization and Are Cell Cycle–Regulated by CDK1 Activity. *PLOS Genetics* **6**, e1001047 (2010).

301. Weinert, T. A. & Hartwell, L. H. The RAD9 Gene Controls the Cell Cycle Response to DNA Damage in Saccharomyces cerevisiae. *Science* **241**, 317–322 (1988).

302. García-Rodríguez, N., Morawska, M., Wong, R. P., Daigaku, Y. & Ulrich, H. D. Spatial separation between replisome- and template-induced replication stress signaling. *EMBO J* **37**, e98369 (2018).

303. Waterman, D. P., Haber, J. E. & Smolka, M. B. CHECKPOINT RESPONSES TO DNA DOUBLE-STRAND BREAKS. *Annu Rev Biochem* **89**, 103–133 (2020).

304. Pizzul, P. *et al.* The DNA damage checkpoint: A tale from budding yeast. *Frontiers in Genetics* **13**, (2022).

305. Galanti, L. & Pfander, B. Right time, right place—DNA damage and DNA replication checkpoints collectively safeguard S phase. *The EMBO Journal* **37**, e100681 (2018).

306. Lanz, M. C., Dibitetto, D. & Smolka, M. B. DNA damage kinase signaling: checkpoint and repair at 30 years. *The EMBO Journal* **38**, e101801 (2019).

307. Liu, Y. *et al.* The intra-S phase checkpoint directly regulates replication elongation to preserve the integrity of stalled replisomes. *PNAS* **118**, (2021).

308. Hoch, N. C. *et al.* Molecular Basis of the Essential S Phase Function of the Rad53 Checkpoint Kinase. *Molecular and Cellular Biology* **33**, 3202–3213 (2013).

309. Osborn, A. J. & Elledge, S. J. Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53. *Genes Dev* **17**, 1755–1767 (2003).

310. Jh, P., Sw, B., Sh, K. & Ds, H. Knockdown of human MCM10 activates G2 checkpoint pathway. *Biochemical and biophysical research communications* **365**, (2008).

311. Watase, G., Takisawa, H. & Kanemaki, M. T. Mcm10 Plays a Role in Functioning of the Eukaryotic Replicative DNA Helicase, Cdc45-Mcm-GINS. *Current Biology* **22**, 343–349 (2012).

312. van Deursen, F., Sengupta, S., De Piccoli, G., Sanchez-Diaz, A. & Labib, K. Mcm10 associates with the loaded DNA helicase at replication origins and defines a novel step in its activation. *EMBO J* **31**, 2195–2206 (2012).

313. Kanke, M., Kodama, Y., Takahashi, T. S., Nakagawa, T. & Masukata, H. Mcm10 plays an essential role in origin DNA unwinding after loading of the CMG components. *The EMBO Journal* **31**, 2182–2194 (2012).

314. Yeeles, J. T. P., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. X. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**, 431–435 (2015).

315. Zhihua Kang *et al.* BRCA2 associates with MCM10 to suppress PRIMPOL-mediated repriming and single-stranded gap formation after DNA damage. *Nature Communications* **12**, 1–12 (2021).

316. Haworth, J., Alver, R. C., Anderson, M. & Bielinsky, A.-K. Ubc4 and Not4 Regulate Steady-State Levels of DNA Polymerase-α to Promote Efficient and Accurate DNA Replication. *Mol Biol Cell* **21**, 3205–3219 (2010).

317. Quan, Y. *et al.* Cell-Cycle-Regulated Interaction between Mcm10 and Double Hexameric Mcm2-7 Is Required for Helicase Splitting and Activation during S Phase. *Cell Rep* **13**, 2576–2586 (2015).

318. Douglas, M. E. & Diffley, J. F. X. Recruitment of Mcm10 to Sites of Replication Initiation Requires Direct Binding to the Minichromosome Maintenance (MCM) Complex. *J Biol Chem* **291**, 5879–5888 (2016).

319. Musiałek, M. W. & Rybaczek, D. Hydroxyurea—The Good, the Bad and the Ugly. *Genes (Basel)* **12**, 1096 (2021).

320. Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R. & Mirkin, S. M. Replication and Expansion of Trinucleotide Repeats in Yeast. *Mol Cell Biol* **23**, 1349–1357 (2003).

321. Campos, L. V. *et al.* RTEL1 and MCM10 overcome topological stress during vertebrate replication termination. *Cell Reports* **42**, 112109 (2023).

322. Cabello-Lobato, M. J. *et al.* Physical interactions between MCM and Rad51 facilitate replication fork lesion bypass and ssDNA gap filling by non-recombinogenic functions. *Cell Reports* **36**, 109440 (2021).

323. Gangavarapu, V., Santa Maria, S. R., Prakash, S. & Prakash, L. Requirement of Replication Checkpoint Protein Kinases Mec1/Rad53 for Postreplication Repair in Yeast. *mBio* **2**, 10.1128/mbio.00079-11 (2011).

324. Baxley, R. M. *et al.* Bi-allelic MCM10 variants associated with immune dysfunction and cardiomyopathy cause telomere shortening. *Nature Communications* **12**, (2021).

325. Baxley, R. M. *et al.* Chronic Mcm10 deficiency causes defects in telomere maintenance in human cells. *bioRxiv* 844498 (2019) doi:10.1101/844498.

326. Polleys, E. J. & Freudenreich, C. H. Genetic Assays to Study Repeat Fragility in Saccharomyces cerevisiae. *Methods Mol Biol* **2056**, 83–101 (2020).

327. Lewis, J. S. *et al.* Single-molecule visualization of Saccharomyces cerevisiae leading-strand synthesis reveals dynamic interaction between MTC and the replisome. *Proc Natl Acad Sci U S A* **114**, 10630–10635 (2017).

328. McClure, A. & Diffley, J. Rad53 checkpoint kinase regulation of DNA replication fork rate via Mrc1 phosphorylation. *eLife* **10**, e69726 (2021).

329. Ait Saada, A. *et al.* Unprotected Replication Forks Are Converted into Mitotic Sister Chromatid Bridges. *Mol Cell* **66**, 398-410.e4 (2017).

330. Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

331. Ji, F. *et al.* New Era of Mapping and Understanding Common Fragile Sites: An Updated Review on Origin of Chromosome Fragility. *Front Genet* **13**, 906957 (2022).

332. Kaushal, S. *et al.* Sequence and Nuclease Requirements for Breakage and Healing of a Structure-Forming (AT)n Sequence within Fragile Site FRA16D. *Cell Reports* **27**, 1151-1164.e5 (2019).

333. Generoso, W. C., Gottardi, M., Oreb, M. & Boles, E. Simplified CRISPR-Cas genome editing for Saccharomyces cerevisiae. *Journal of Microbiological Methods* **127**, 203–205 (2016).

334. Psakhye, I., Castellucci, F. & Branzei, D. SUMO-Chain-Regulated Proteasomal Degradation Timing Exemplified in DNA Replication Initiation. *Mol Cell* **76**, 632-645.e6 (2019).

335. Forey, R. *et al.* Mec1 Is Activated at the Onset of Normal S Phase by Low-dNTP Pools Impeding DNA Replication. *Molecular Cell* **78**, 396-410.e4 (2020).

336. Chabes, A. & Stillman, B. Constitutively high dNTP concentration inhibits cell cycle progression and the DNA damage checkpoint in yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **104**, 1183–1188 (2007).

337. Zhao, X. & Rothstein, R. The Dun1 checkpoint kinase phosphorylates and regulates the ribonucleotide reductase inhibitor Sml1. *Proc Natl Acad Sci U S A* **99**, 3746–3751 (2002).

338. Greene, B. L. *et al.* Ribonucleotide Reductases: Structure, Chemistry, and Metabolism Suggest New Therapeutic Targets. *Annual Review of Biochemistry* **89**, 45–75 (2020).

339. Tsantoulis, P. K. *et al.* Oncogene-induced replication stress preferentially targets common fragile sites in preneoplastic lesions. A genome-wide study. *Oncogene* **27**, 3256–3264 (2008).

340. Papadopoulou, C., Guilbaud, G., Schiavone, D. & Sale, J. E. Nucleotide Pool Depletion Induces G-Quadruplex-Dependent Perturbation of Gene Expression. *Cell Reports* **13**, 2491–2503 (2015).

341. Mullen, N. J. & Singh, P. K. Nucleotide metabolism: a pan-cancer metabolic dependency. *Nat Rev Cancer* **23**, 275–294 (2023).

342. Aird, K. M. & Zhang, R. Nucleotide metabolism, oncogene-induced senescence and cancer. *Cancer Lett* **356**, 204–210 (2015).

343. Fleck, O. *et al.* Spd1 accumulation causes genome instability independently of ribonucleotide reductase activity but functions to protect the genome when deoxynucleotide pools are elevated. *J Cell Sci* **126**, 4985–4994 (2013).

344. Prakash, S. & Prakash, L. Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. *Genes Dev* **16**, 1872–1883 (2002).

345. Pai, C. & Kearsey, S. A Critical Balance: dNTPs and the Maintenance of Genome Stability. *Genes* **8**, 57 (2017).

346. Williams, L. N. *et al.* dNTP pool levels modulate mutator phenotypes of error-prone DNA polymerase ε variants. *PNAS* **112**, E2457–E2466 (2015).

347. Chabes, A. *et al.* Survival of DNA Damage in Yeast Directly Depends on Increased dNTP Levels Allowed by Relaxed Feedback Inhibition of Ribonucleotide Reductase. *Cell* **112**, 391–401 (2003).

348. Håkansson, P., Hofer, A. & Thelander, L. Regulation of mammalian ribonucleotide reduction and dNTP pools after DNA damage and in resting cells. *J Biol Chem* **281**, 7834–7841 (2006).

349. Das, B., Mishra, P., Pandey, P., Sharma, S. & Chabes, A. dNTP concentrations do not increase in mammalian cells in response to DNA damage. *Cell Metabolism* **34**, 1895–1896 (2022).

350. Leffak, M. Hypothesis: Local dNTP depletion as the cause of microsatellite repeat instability during replication (Comment on DOI 10.1002/bies.201200128). *BioEssays* **35**, 305–305 (2013).

351. Kuzminov, A. Inhibition of DNA synthesis facilitates expansion of low-complexity repeats: Is strand slippage stimulated by transient local depletion of specific dNTPs? *Bioessays* **35**, 306–313 (2013).

352. Meroni, A. *et al.* RNase H activities counteract a toxic effect of Polymerase η in cells replicating with depleted dNTP pools. *Nucleic Acids Res* **47**, 4612–4623 (2019).

353. Barberis, M. *et al.* The yeast cyclin-dependent kinase inhibitor Sic1 and mammalian p27Kip1 are functional homologues with a structurally conserved inhibitory domain. *Biochem J* **387**, 639–647 (2005).

354. Bertoli, C., Skotheim, J. M. & de Bruin, R. A. M. Control of cell cycle transcription during G1 and S phases. *Nat Rev Mol Cell Biol* **14**, 518–528 (2013).

355. Schmidt, T. T. *et al.* A genetic screen pinpoints ribonucleotide reductase residues that sustain dNTP homeostasis and specifies a highly mutagenic type of dNTP imbalance. *Nucleic Acids Research* **47**, 237–252 (2019).

356. Koç, A., Wheeler, L. J., Mathews, C. K. & Merrill, G. F. Hydroxyurea Arrests DNA Replication by a Mechanism That Preserves Basal dNTP Pools *. *Journal of Biological Chemistry* **279**, 223–230 (2004).

357. Barberis, M. Sic1 as a timer of Clb cyclin waves in the yeast cell cycle – design principle of not just an inhibitor. *The FEBS Journal* **279**, 3386–3410 (2012).

358. Lengronne, A. & Schwob, E. The Yeast CDK Inhibitor Sic1 Prevents Genomic Instability by Promoting Replication Origin Licensing in Late G1. *Molecular Cell* **9**, 1067–1078 (2002).

359. Nishizawa, M., Kawasumi, M., Fujino, M. & Toh-e, A. Phosphorylation of Sic1, a Cyclin-dependent Kinase (Cdk) Inhibitor, by Cdk Including Pho85 Kinase Is Required for Its Prompt Degradation. *Mol Biol Cell* **9**, 2393–2405 (1998).

360. Tripodi, F., Zinzalla, V., Vanoni, M., Alberghina, L. & Coccetti, P. In CK2 inactivated cells the cyclin dependent kinase inhibitor Sic1 is involved in cell-cycle arrest before the onset of S phase. *Biochem Biophys Res Commun* **359**, 921–927 (2007).

361. Poli, J. *et al.* dNTP pools determine fork progression and origin usage under replication stress. *EMBO J* **31**, 883–894 (2012).

362. Sundararajan, R. & Freudenreich, C. H. Expanded CAG/CTG Repeat DNA Induces a Checkpoint Response That Impacts Cell Proliferation in Saccharomyces cerevisiae. *PLOS Genetics* **7**, e1001339 (2011).

363. Gupta, A. *et al.* Telomere Length Homeostasis Responds to Changes in Intracellular dNTP Pools. *Genetics* **193**, 1095–1105 (2013).

364. Kumar, D. *et al.* Mechanisms of mutagenesis in vivo due to imbalanced dNTP pools. *Nucleic Acids Research* **39**, 1360–1371 (2011).

365. Gon, S., Napolitano, R., Rocha, W., Coulon, S. & Fuchs, R. P. Increase in dNTP pool size during the DNA damage response plays a key role in spontaneous and induced-mutagenesis in Escherichia coli. *Proc Natl Acad Sci U S A* **108**, 19311–19316 (2011).

366. Davidson, M. B. *et al.* Endogenous DNA replication stress results in expansion of dNTP pools and a mutator phenotype. *EMBO J* **31**, 895–907 (2012).

367. Zhao, X., Muller, E. G. D. & Rothstein, R. A Suppressor of Two Essential Checkpoint Genes Identifies a Novel Protein that Negatively Affects dNTP Pools. *Molecular Cell* **2**, 329–340 (1998).

368. Akimova, E. *et al.* SAMHD1 restrains aberrant nucleotide insertions at repair junctions generated by DNA end joining. *Nucleic Acids Research* **49**, 2598–2608 (2021).

369. Franzolin, E. *et al.* The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proc Natl Acad Sci U S A* **110**, 14272–14277 (2013).

370. Rentoft, M. *et al.* Heterozygous colon cancer-associated mutations of SAMHD1 have functional significance. *PNAS* **113**, 4723–4728 (2016).

371. Ho, S. S., Urban, A. E. & Mills, R. E. Structural variation in the sequencing era. *Nat Rev Genet* **21**, 171–189 (2020).

372. Liao, X. *et al.* Repetitive DNA sequence detection and its role in the human genome. *Commun Biol* **6**, 1–21 (2023).

373. Panzer, S., Kuhl, D. P. & Caskey, C. T. Unstable triplet repeat sequences: a source of cancer mutations? *Stem Cells* **13**, 146–157 (1995).

374. Bonneville, R. *et al.* Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precis Oncol* 1–15 (2017) doi:10.1200/PO.17.00073.

375. Erwin, G. S. *et al.* Recurrent repeat expansions in human cancer genomes. *Nature* **613**, 96–102 (2023).

376. Dolzhenko, E. *et al.* ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biology* **21**, 102 (2020).

377. Bacolla, A., Tainer, J. A., Vasquez, K. M. & Cooper, D. N. Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. *Nucleic Acids Res* **44**, 5673–5688 (2016).

378. Hisey, J. A. *et al.* Pathogenic CANVAS (AAGGG)n repeats stall DNA replication due to the formation of alternative DNA structures. *Nucleic Acids Research* **52**, 4361–4374 (2024).

379. Schiavoni, V. *et al.* Recent Advances in the Management of Clear Cell Renal Cell Carcinoma: Novel Biomarkers and Targeted Therapies. *Cancers (Basel)* **15**, 3207 (2023).

380. Kompella, P. *et al.* Obesity increases genomic instability at DNA repeat-mediated endogenous mutation hotspots. *Nat Commun* **15**, 6213 (2024).

381. Umek, T. *et al.* Oligonucleotide Binding to Non-B-DNA in MYC. *Molecules* **24**, 1000 (2019).

382. Krasilnikov, A. S. *et al.* Mechanisms of Triplex-Caused Polymerization Arrest. *Nucleic Acids Research* **25**, 1339–1346 (1997).

383. Masuda, Y. *et al.* Deoxycytidyl transferase activity of the human REV1 protein is closely associated with the conserved polymerase domain. *J Biol Chem* **276**, 15051–15058 (2001).

384. Patrick, S. M. & Turchi, J. J. Stopped-flow Kinetic Analysis of Replication Protein A-binding DNA: DAMAGE RECOGNITION AND AFFINITY FOR SINGLE-STRANDED DNA REVEAL DIFFERENTIAL CONTRIBUTIONS OF kon ANDkoff RATE CONSTANTS*. *Journal of Biological Chemistry* **276**, 22630–22637 (2001).

385. Kim, J. C., Harris, S. T., Dinter, T., Shah, K. A. & Mirkin, S. M. The role of break-induced replication in large-scale expansions of (CAG)n/(CTG)n repeats. *Nat Struct Mol Biol* **24**, 55–60 (2017).

386. Signon, L., Malkova, A., Naylor, M. L., Klein, H. & Haber, J. E. Genetic requirements for RAD51- and RAD54-independent break-induced replication repair of a chromosomal double-strand break. *Mol Cell Biol* **21**, 2048–2056 (2001).

387. Brázda, V., Laister, R. C., Jagelská, E. B. & Arrowsmith, C. Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Molecular Biology* **12**, 33 (2011).

388. Hartenstine, M. J., Goodman, M. F. & Petruska, J. Base stacking and even/odd behavior of hairpin loops in DNA triplet repeat slippage and expansion with DNA polymerase. *J Biol Chem* **275**, 18382–18390 (2000).

389. Chen, D. *et al.* MCM10, a potential diagnostic, immunological, and prognostic biomarker in pan-cancer. *Sci Rep* **13**, 17701 (2023).