Helicase Requirements for Prevention of Fragility at the Flex1 AT Repeat

Samantha Regan

Biology Senior Honors Thesis 2018
Tufts University

Research Mentor/Principal Investigator:
Dr. Catherine Freudenreich

Acknowledgements

I have been so lucky over the course of this project to have the support of so many incredible people. I would like to thank Dr. Catherine Freudenreich, for your guidance and for the opportunity to work on this fascinating project. Thank you also to Dr. Mitch McVey for all your help, advice and support, as an academic advisor as well as a thesis committee member. I want to give a huge thank you to Simran Kaushal, for your dedication, patience, and unbeatable mentorship at every step of the way. Also thank you to the entire Freudenreich lab, especially to Erica, Jenna and Anisha, and the rest of the FRA16D team Charlie, Julia and Ruby. I owe a big thank you to my parents for their never-ending support and love. I'm also grateful for my sister, my family, and my friends, for always being there for me even when I'm sure you're tired of hearing about yeast.

TABLE OF CONTENTS

Abstract

Common fragile sites are regions of the genome that are present in all individuals, and prone to breakage under replicative stress. FRA16D, one of the most highly expressed common fragile sites, is located in the WWOX gene. The WWOX protein is a putative tumor suppressor which has been linked to cancer proliferation. Flex1, an AT-rich sequence within the FRA16D site, contributes to much of FRA16D's fragility (Zhang and Freudenreich, 2007). It is possible that a secondary structure may form in Flex1, causing fragility. These secondary structures may need to be unwound, or they may stall replication and cause a reversed fork; either scenario can prevent proper replication progression and result in DNA breakage. In this project, genetic assays in *Saccharomyces cerevisiae* were used to indirectly measure fragility of Flex1 in the absence of various helicases associated with replication and DNA damage and repair. The rate of fragility in the presence and absence of helicases Rrm3, Srs2, Sgs1, and Mph1 was evaluated. The absence of Rrm3 seems to cause a decrease in fragility at Flex1. The presence of Sgs1 and Mph1 have a protective effect at Flex1, and in their absence there is an increase in fragility. Srs2 does not seem to have an effect fragility at Flex1. Based on these results, it is possible to predict that some of these helicases may prevent breakage at Flex1 through the formation of regressed forks (especially Mph1). Fork regression could lead to eventual fork restart or mitotic DNA synthesis across the AT repeat at Flex1, and be a way to avoid DNA fragility.

INTRODUCTION

Common Fragile Sites and FRA16D

Common Fragile Sites

Common fragile sites (CFSs) are regions of DNA that are prone to breakage. CFSs are large regions of DNA (hundreds of kilobases to over a megabase of DNA), and tend to be rich in AT base pairs. Although common fragile site sequences are always present they are not always broken, or expressed, unless under replicative stress (Glover 2006). Replicative stressors include aphidicholin (which inhibits polymerase), caffeine (which inhibits kinase ATR), carcinogens, and oncogene expression. They all can induce breakage at common fragile sites (Guven et al. 1999 and Macheret & Halazonetis 2015). Under replication stress, CFSs display gaps and breaks in metaphase chromosomes (Figure 1, Glover et al. 2005). These sites are present in all humans and are conserved across species. For example, the *Saccharomyces cerevisiae* equivalent of a common fragile site is a 'replication slow zone', a region that replicates more slowly and is more prone to breakage in the absence of Mec1 (human ATR), a replication checkpoint kinase (Cimprich 2003). Other natural fragile sites in *S. cerevisiae* include Ty elements, which are inverted repeats and can form secondary structures. With two Ty elements head-to-head, double stranded breaks can occur when replication is compromised (Lemoine et al. 2005).

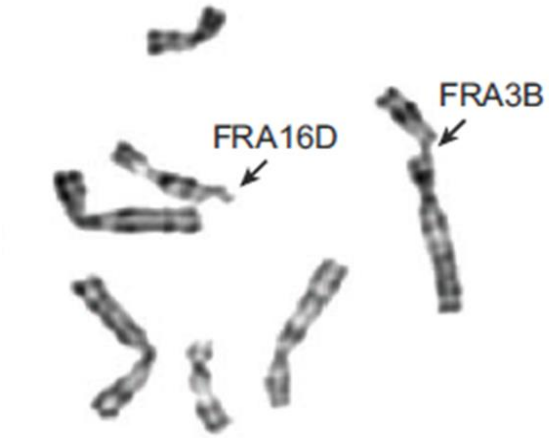


Figure 1. These images show two of the most commonly expressed common fragile sites, FRA16D and FRA3B, on metaphase chromosomes. The right broken site, FRA3B, appears to have one broken sister chromatid, possibly due to nuclease cleavage. The left broken site, FRA16D, looks like a gap. MiDAS (mitotic DNA synthesis) is likely occurring here (Glover et al 2005).

Common fragile sites are, in many ways, linked to cancer. CFSs are often sites of translocations, rearrangements, deletions, and loss of heterozygosity in many cancers (Glover et al. 2017). More than half of cancer-specific chromosomal translocations occur within common fragile sites, so understanding common fragile site breakage will lead to a better comprehension of carcinogenesis (Dillon et al. 2010). Also, two of the most commonly expressed fragile sites, FRA3B and FRA16D, are located in putative tumor suppressors (FHIT and WWOX respectively). Breakage at these fragile sites can lead to deletions, or loss of function of these tumor suppressors, and therefore cancer proliferation.

Many common fragile sites are present in large genes. FHIT spans 1.5 Mb, and WWOX is 1.1 Mb long (Glover et al. 2017). Some common fragile sites are cell-specific, while others are found in all cell types (Debatisse et al. 2012). CFSs have a dearth of replication origins, meaning that within many common fragile sites there are less origin firing events than would be expected for a region of that size (Letessier et al. 2011).

Related to the concept of replication, many common fragile sites have been proven to be late replicating. For example, FRA3B does not replicate until the cell is well into S phase (Le Beau et al. 1998), and others have similar delays in replication (Durkin & Glover 2007). With added replication stress in the form of APH treatment, replication could be delayed even into M phase. The Hickson group demonstrated that when incompletely replicated common fragile sites pass into M phase and therefore do not condense properly, a special M phase replication process is activated. They call this process MiDAS, or Mitotic DNA Synthesis (Minocherhomji et al. 2015). Because of such late replication, regions containing common fragile sites are often unable to go through normal chromosome condensation (Glover et al. 2017).

Suspected Mechanisms of Breakage

Because of the high levels of fragility at common fragile sites, it is possible that multiple mechanisms of breakage are at play. One such factor contributing to breakage is repetitive sequences that are present in many common fragile sites. Because common fragile sites are AT rich (and contain repetitive AT sequences) they are more likely to form secondary structures. In AT-rich sequences, the DNA strands are more likely to separate, since AT base pairs only have two hydrogen bonds. GC base pairs have three hydrogen bonds, making their connection stronger. AT repeats are especially prone to forming cruciforms, and have been proven to do so in *E. coli* (Dayn et al. 1991). The formation of secondary structures within common fragile sites could lead to fragility in multiple ways. Replication forks encountering these structures could stall, eventually leading to double stranded breaks. A structure-forming AT repeat within FRA16D has been shown to stall replication (Zhang & Freudenreich 2007). Alternatively, proteins such as nucleases could cleave DNA at the secondary structure. Secondary structures

could cause the formation of reversed forks, which could also be cleaved by nucleases (Kaushal et al. in prep).

Patterns of replication at common fragile sites also contribute to their fragility. These sites are late replicating, and do not go through replication until the end of S phase. When under replication stress, such as when treated with aphidicolin, CFS can replicate as late as M phase. Although common fragile sites can go through M phase replication (Minocherhomji et al. 2015), if that does not occur common fragile sites have the potential to break or to form breakage prone structures such as ultra fine bridges (UFBs). With UFBs, sister chromatids do not separate completely because they are connected by unreplicated DNA. Ultra fine bridges can lead to breakage, chromosome rearrangements, or nondisjunction (Glover et al. 2017).

Transcription itself also seems to be related to common fragile sites fragility. Transcription is known to impede replication fork progress in yeast (Prado and Aguilera 2005), which can lead to breakage and recombination-mediated deletions. Many common fragile sites are located within large genes so it would take a longer time to complete transcription, and therefore transcription and replication may occur at the same time (Franchitto 2013). Collisions between transcription machinery and the replication fork are possible, and would lead to fragility. Additionally, with transcription DNA:RNA hybrids called R-loops can form. The presence of such structures have been linked to increased fragility in common fragile sites (Helmrich et al. 2011, Madireddy et al. 2016).

Common Fragile Site FRA16D and the WWOX tumor suppressor

In this project a specific common fragile site, FRA16D, is being examined. This particular fragile site is highly expressed, meaning that breakage occurs in a high percentage of

cells and various cell types (Helmrich et al. 2011). FRA16D is located in an intron between exons eight and nine of the large human WW domain-containing oxidoreductase (WWOX) gene. The WWOX gene is considered a putative tumor suppressor, and the presence of a functional WWOX protein has been proven to inhibit breast cancer tumor growth (Bednarek et al. 2001). Additionally, lower levels of the WWOX protein in cancer patients with tumors correlate with poorer prognosis for ovarian cancer and breast cancer (Nunez et al. 2005, Aqeilan et al. 2007). Humans with lower levels of WWOX are predisposed to lung cancer and glioma (Yang et al. 2013, Yu et al. 2014).

The WWOX protein's actual function has been linked to metabolism. Genome-wide studies associate the WWOX protein and triglyceride and high density lipoprotein cholesterol levels in humans (Saez et al. 2010). Additionally, *Drosophila melanogaster* without the WWOX gene have alterations in metabolic mRNA and proteins. Specifically, based on correlations of mRNA in studies in *Drosophila* with a highly conserved WWOX orthologue, the WWOX protein seems to work in pathways with superoxide dismutase (SOD1) and isocitrate dehydrogenase (IDH) (O'Keefe et al. 2011). IDH is a part of the citric acid cycle, and alterations in cellular metabolism are a hallmark of cancer. SOD1 has a role in regulating sensitivity to ionizing radiation, and lower levels of WWOX could possibly lead to radioresistance in tumors (Richards et al. 2015). The role of WWOX as a tumor suppressor seems to work in conjunction with other genes as well, since lower levels of WWOX lead to a predisposition to cancer and poorer prognosis. Even if WWOX does not initiate cancer, it seems clear that it does have a role in cancer progression. Understanding what could cause loss of function in WWOX, such as breakage at FRA16D, could lead to a better understanding of cancer cell proliferation.

Subregions of FRA16D and Fragility

Previously, using the Flexstab program, the Freudenreich laboratory identified sub-regions of FRA16D that had a higher likelihood of forming secondary structures. Regions of interest of FRA16D were chosen based on their palindrome length, because long palindromes are more likely to form secondary structures such as hairpins and cruciforms. A number of sequences were identified that were predicted to form stable secondary structures (Figure 2A). These secondary structures could be hairpins or, more likely, cruciforms (Figure 3, Zhang and Freudenreich 2007). Some sequences of interest include Flex1 and Flex5 (Figure 2B).

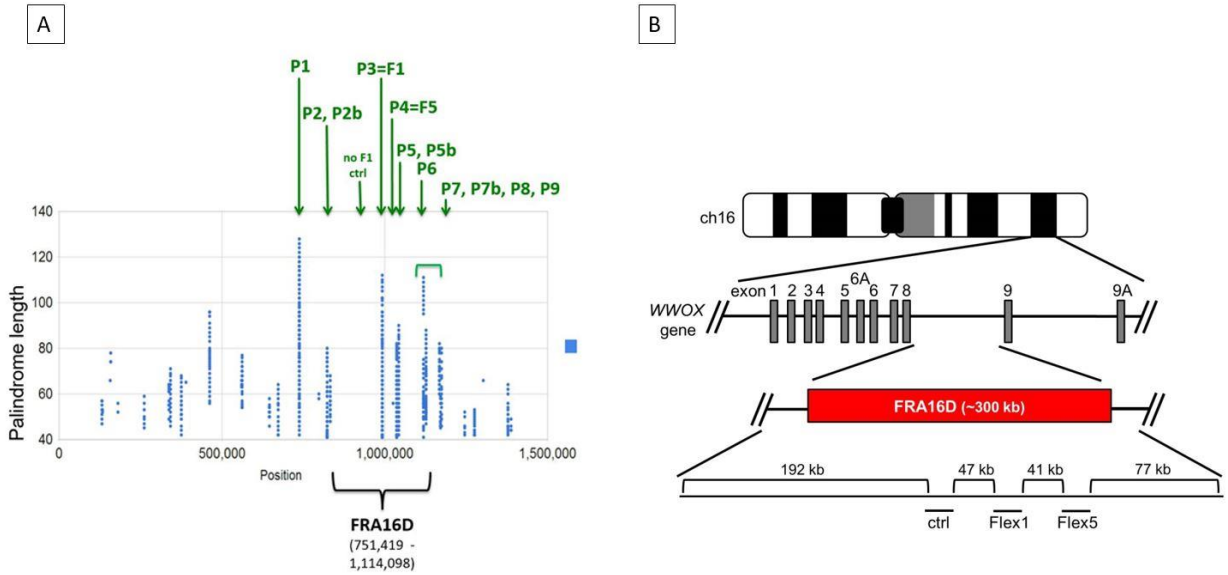


Figure 2. A) Palindromic sequences within FRA16D were compared and evaluated. The sequence that is being focused on in this project to evaluate helicases, Flex1, is one of the more stable palindromic sequences of the region. Because of that, is more likely to form secondary structures (palindrome program by Catherine Freudenreich, Sergei Mirkin, Anoop Kumar, Lenore Cowen; figure by CHF & SK). B) Flex1 is a region within common fragile site FRA16D, which is located in an intron between exons eight and nine of the *WWOX* gene, on human chromosome II (Kaushal et al. in prep).

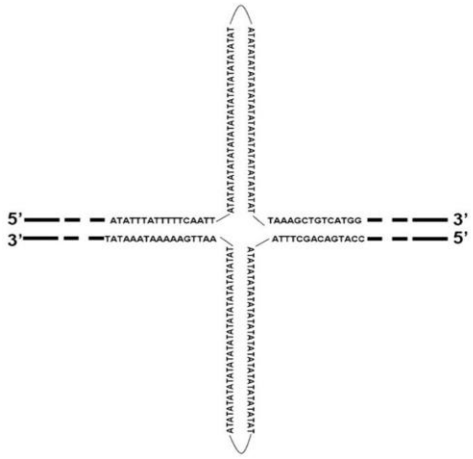


Figure 3. Cruciforms form when repeats, especially AT repeats, bind to each other instead of the other strand of DNA. They appear as two hairpins across from one another. The cruciform is a secondary structure that is likely to be forming at the Flex1 AT repeat (Zhang & Freudenreich 2007).

These subregions of interest were then evaluated with a fragility assay using a Yeast Artificial Chromosome (YAC) containing regions of FRA16D. Using this YAC-based fragility assay, it was determined that the Flex1 region of FRA16D had a higher rate of fragility than other sub-regions of FRA16D (Figure 4, Zhang and Freudenreich, 2007). Based on this information, a subtractive assay was performed. The Flex1 sequence was deleted from a YAC containing DNA from human chromosome 16, including the entire FRA16D sequence. The rate of fragility of a strain containing this YAC was compared with that of strains containing all of FRA16D, including Flex1. The strain that was deleted for Flex1 had a significantly lower rate of breakage, implying that Flex1 is important for fragility of FRA16D. The rate of fragility of the strain without Flex1 was still higher than the rate of a strain that contained human DNA from a region not containing FRA16D. Therefore, while Flex1 contributes to the fragility of FRA16D, it may not be the only source of fragility (Figure 5A, Kaushal et al. in prep).

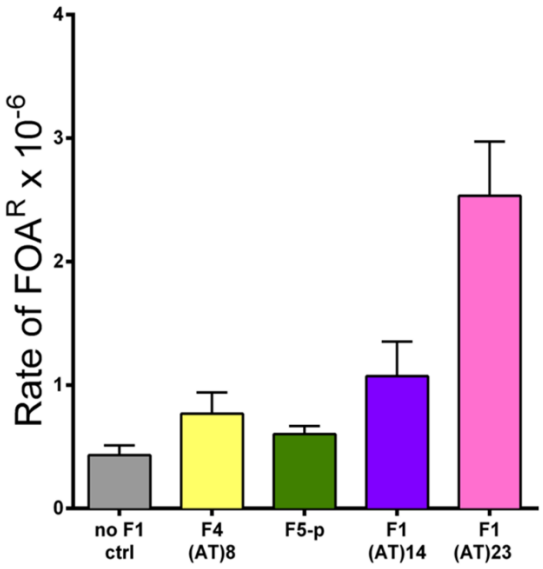


Figure 4. Strains of S. cerevisiae with Flex1 on a Yeast Artificial Chromosome were measured to have higher rates of FOA resistance, which implies higher rates of breakage. Rate of breakage seems correlate with the number of AT repeats. The F stands for Flex (such as Flex1), and each of these strains contains a palindromic region identified by the Flexstab program. The strains in this assay are *rad52*Δ (Adapted from Zhang and Freudenreich 2007).

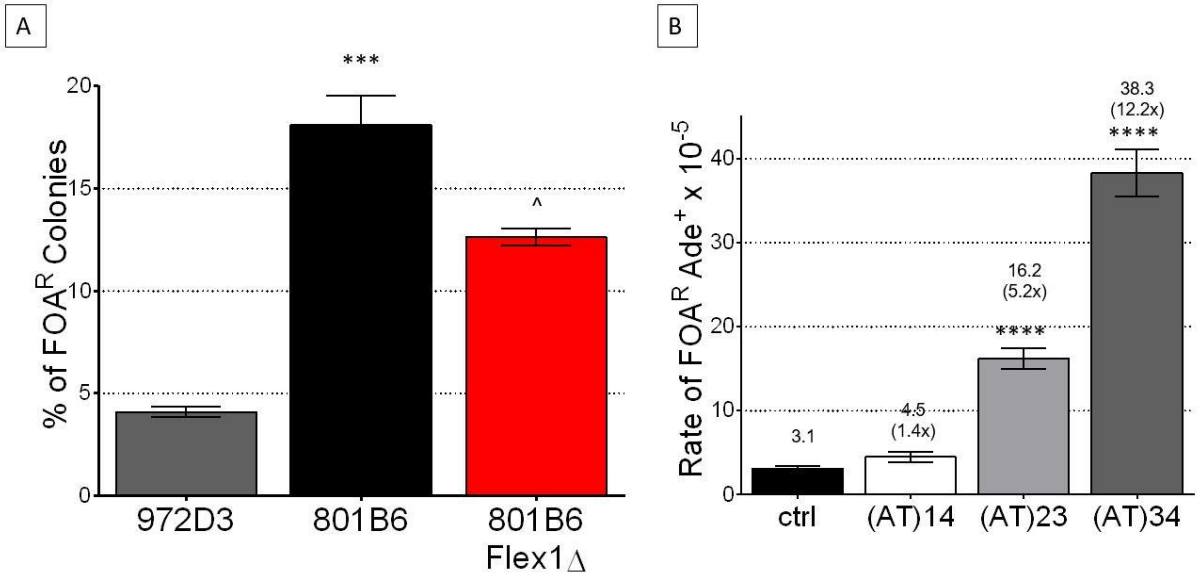


Figure 5. A) YAC 801B6 contains human DNA from chromosome 16, including the full FRA16D sequence. YAC 972D3 contains a human DNA that is adjacent to FRA16D on chromosome 16 (but does not contain any part of FRA16D). When the Flex1 sequence is not included on the YAC, rate of fragility decreases significantly but not to the rate of the control sequence. This implies that Flex1 is important for fragility within the context of FRA16D, but it is not the only factor that contributes to fragility. *** = $p<0.001$ compared to FRA16D 972D3 p

= 0.0003, ^ = p<0.05 compared to FRA16D 801B6 with Flex1 B) Rate of recombination increases as tract length increases for AT repeats in Flex1, as measured in the direct duplication recombination assay. This adds validity to the data of Figure 4, which was done in a YAC system. **** = p<0.0001 compared to a control sequence with no AT repeats (Kaushal et al., in prep).

The Flex1 sequence contains polymorphic ATs, and can have anywhere from 11 to 88 perfect AT repeats (Finnis et al. 2005). As the tract length of AT repeats within Flex1 increases, rate of fragility tends to increase in an AT length dependent manner. The AT length dependent fragility was also evaluated in a recombination-based system, which showed a more clear relationship between AT length and breakage (Figure 5B, Kaushal et al. in prep). As AT length increases, there is a greater chance of more stable secondary structures forming, which could increase the chance of chromosome breakage. AT repeats are predicted to form stable secondary structures beginning at 23 AT repeats, which is the point at which the breakage rate significantly increases (Zhang & Freudenreich 2007). This implies a connection between the formation of a stable secondary structure and double stranded DNA breakage.

Based on this information, it can be concluded that Flex1 is fragile and important for fragility within the context of FRA16D as a whole. Because of this, the Flex1 sequence provides a useful model, or proxy, for evaluating fragility of FRA16D. The sequence used in this project contains 34 perfect AT repeats, as well as flanking sequences of DNA from the Flex1 region on either side. Using this construct, we are able to determine how the fragility of a secondary-structure-forming Flex1 sequence would be affected by factors that act on DNA.

The Role of Helicases

Helicases are proteins that prepare the DNA for the replication fork in various ways. They clear the genome of obstacles that could cause collisions between transcription or

replication proteins, and therefore minimize DNA breakage. Although the main replicative helicase in *Saccharomyces cerevisiae* is the Mcm2-7 complex, there are other proteins that have the ability to help the fork move through hard-to-replicate regions such as secondary structures or recombination intermediates. Especially because of the role of many helicases in unwinding secondary structures, this group of proteins could play a key part in preventing or mitigating fragility at Flex1.

Rrm3

Rrm3 is a 5' to 3' helicase that moves with the replication fork to prevent stalling. It is a member of the Pif1 family helicases. In humans there is only one protein of this family, Pif1, but in *Saccharomyces cerevisiae* there are two proteins: Pif1 and Rrm3. Although they do have some overlapping roles and can act as backups for one another in certain functions, Rrm3 does have certain roles independent of Pif1. Although it is only active at certain locations in the genome, Rrm3 always stays with DNA polymerase as it moves through the DNA (Azvolinsky et al. 2006), and is known to remove bound proteins from the genome ahead of the replication fork. Strains knocked out for *rrm3* are more likely to have fork stalling at around 1,400 different places in the genome. Some of Rrm3's target sequences include the genes that code for tRNA (tDNA), inactive origins, and transcriptional silencers (Ivessa et al. 2002). It also has a role at multiple sites in ribosomal DNA, including at the replication fork barrier (Torres et al. 2004). At all of these locations, the Rrm3 protein contributes to the prevention of replication defects by removing bound protein complexes. Rrm3 also catalytically promotes the replication of telomeres, and in its absence there is an increase in slowing of replication forks (Ivessa et al. 2002).

Rrm3 seems to interact with the yeast cell cycle checkpoint protein Mec1, which is part of the damage response pathway. When *MEC1* is mutated, increased DNA fragility occurs, since Mec1 prevents double stranded break formation at replication slow zones in yeast (similar to the human common fragile site). However, when *rrm3* is also knocked out, this breakage phenotype is rescued somewhat and the rate of breakage decreases (Hashash et al 2010).

Rrm3 does not have a primary role in unwinding DNA secondary structures. For example, it does not seem to have a role in unwinding hairpins, as it does not affect fork stalling at hairpin-forming sequences (Anand et al. 2011). However, it does act as a backup to the protein Pif1 and, in its absence, can bind to G4 DNA and suppress G4-induced genome instability (Paeschke et al. 2013). Rrm3 is able to suppress damage at tDNA caused by DNA:RNA hybrids, or R loops, so it may have a role in unwinding these structures (Tran et al. 2017).

Rrm3 also has a role in recombination and repair. Rrm3 is involved in repair of replication-induced double stranded breaks, and accumulates at such breaks (Muñoz-Galván et al 2017). The name *RRM3* stands for rDNA recombinational mutation, and this holds true. In the absence of *rrm3*, mitotic recombination increases at rDNA and *CUP1* tandem repeats, but not at other repeated (but not tandem repeat) sequences, such as Ty elements. This increased recombination is likely due to defects in replication, since strains knocked out for *rrm3* are more likely to have fork stalling events (Keil & McWilliams 1993). In the absence of either *sgs1* or *srs2* (both proteins involved in recombination-based repair) *RRM3* is required for healthy growth (Schmidt & Kolodner 2004). It is likely that in the absence of Rrm3, stalled or broken forks form that could form toxic recombination intermediates (when acted on by Rad51) in the absence of Sgs1 or Srs2 (Torres et al. 2004). Additionally, Rrm3 is shown to associate with stalled forks in cells that are *rad53*Δ, in which the Rad53 checkpoint is not active (Rossi et al. 2015). In *rad53*Δ

cells, Rrm3 seems to have a role in the accumulation of both reversed and collapsed forks. (Rossi et al. 2015).

Rrm3 was chosen to study because of its main role in removing proteins bound to the DNA (and not secondary structures). The goal is to compare it to other helicases that are, in contrast, known to unwind secondary structures that might be forming in Flex1.

Srs2

Srs2 is a 3' to 5' helicase that is known to have a helicase and a strippase role in the cell. Its human homolog is not known, although it seems to have some overlapping functions with human proteins RTEL and FBH1. Srs2 has been shown to unwind secondary structures such as CTG and CGG hairpins to in order to facilitate replication and to prevent repeat fragility (Kerrest et al 2009, Anand et al 2011). Srs2 does not mitigate fork stalling at G4T4 sequences nor at protein barriers, so its helicase role is not universal across all secondary structures (Anand et al. 2011). Since cruciforms appear as two hairpins across from one another, the role of Srs2 at hairpins could be relevant at Flex1. Srs2 has the ability to unwind Y form DNA, 5' flaps, and most efficiently 3' flaps in vitro, and has a binding preference for ssDNA-dsDNA junctions. However, it is inefficient at unwinding double stranded DNA unless a nick is introduced (Marini & Krejci 2012). Srs2 may have a role in the formation of reversed forks, perhaps by removing Rad51 from damaged forks to bypass damage and allow fork restart (Kerrest et al. 2009).

The strippase role of Srs2 is defined as being able to displace, or strip, Rad51 from the DNA (Macris & Sung 2005). Since Rad51 is important for recombination-based DNA repair pathways, Srs2 effectively can inhibit and regulate DNA repair. After a double strand break, nucleases resect the DNA. Rad51 binds to the 3' end of the single stranded DNA (ssDNA)

present after this resection, and the nucleoprotein filament (Rad51 and ssDNA) searches for a homologous DNA sequence and promotes the formation of a D loop (Sung and Klein 2006). Srs2 physically causes the displacement of Rad51 from ssDNA by stimulating ATP hydrolysis in Rad51 (Antony et al. 2009) and it also antagonizes Rad52, which has a role in mediating the Rad51-ssDNA nucleoprotein filament (Seong et al. 2009). When Rad51 is removed from the ssDNA by Srs2, this homology search and D loop formation cannot occur, and recombination will not occur. Because of its role in preventing the formation of the nucleoprotein filament and preventing the formation of D loops, Srs2 can be considered an anti-recombinase (Niu & Klein 2017, Figure 6).

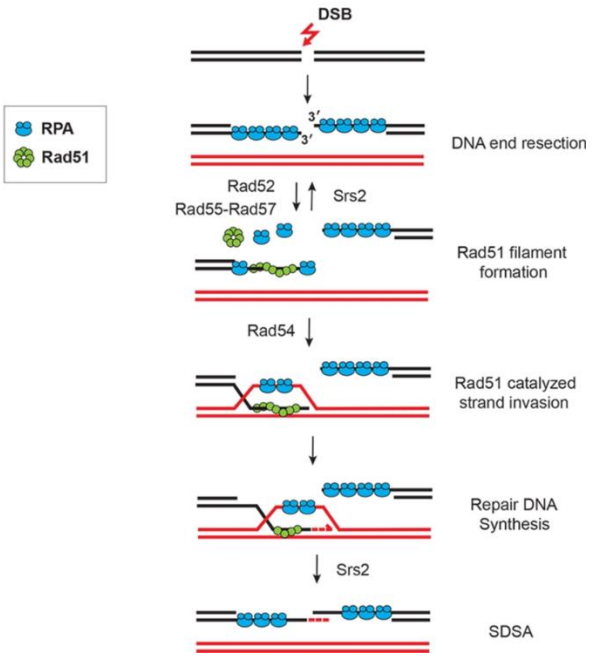


Figure 6. The strippase role of Srs2 allows it to disassemble the nucleofilament of Rad51 attached to single stranded DNA. This disassembly can occur as an anti-recombinase role, right after the formation of the Rad51 filament, before the strand invasion step. It can also occur in a pro-recombinase capacity after the D loop has already been extended, as part of the SDSA or other repair pathways (Niu & Klein 2017).

However, Srs2 can also be considered a pro-recombinase in its strippase role because it appears to promote non-crossing over in the synthesis-dependent single-strand annealing (SDSA)

pathway of recombination-based repair. Srs2 could have a role in preventing crossing over through displacement of the D loop, and minimization of the formation of toxic recombination intermediates (Marini & Krejci 2010). The ability of Srs2 to remove Rad51 from ssDNA filaments may also be useful in other repair pathways such as single strand annealing (SSA) or break induced replication (BIR), and especial Nonhomologous End Joining, or NHEJ (Carter et al. 2009). Srs2 can prevent toxic joint molecules and recombination intermediates from forming by removing Rad51 from ssDNA and allowing these repair pathways to move forward (Elango et al. 2017).

Outside of its direct role in recombination, Srs2 has many other functions involved in replication, some of which may be because of its strippase ability. Srs2 is also known to have a role in post replication repair (PRR), and specifically in the error-free gap repair pathway. It influences lesions into the Rad6 repair pathway instead of alternative recombination repair pathways, perhaps by removing Rad51 (Marini & Krejci 2010). Srs2 is also required for recovery from checkpoint arrest, as it is involved in turning off the DNA damage checkpoint and inactivating the Rad53 kinase (Vaze et al. 2002).

Additionally, Srs2 promotes the activity of Mus81, an endonuclease that has a role in cleaving many structures produced by replication and recombination (Chavdarova et al 2015). Rad51 can inhibit Mus81 activity, and with its removal Srs2 mitigates this inhibition. Both proteins colocalize to damaged regions of DNA, and they could be working together in order to resolve recombination intermediates (Chavdarova et al 2015).

Previously, *SRS2* has been tested by the Freudenreich laboratory for fragility in an assay system which measures fragility based on recombination rate. The rates of fragility of a *srs2*Δ strain that were found were 16.9 x 10^{-6} $FOA^{R}Ade^{+}$ and 18.7 x 10^{-6} $FOA^{R}Ade^{+}$, which was a

decreased rate from wildtype and would have implied that the presence of Srs2 actually increases fragility (Figure 7). However, Srs2 does have an effect on the process of single strand annealing, which is the method through which this assay functions. The Haber laboratory has shown that the absence of *srs2* causes a three-fold reduction in single strand annealing in their system (Sugawara et al. 2000), which mirrors the results that we are seeing. Additionally, because of its role in recombination (both as a recombinase and an antirecombinase) the assay system was used less than ideal for evaluating its effect on fragility. For more reliable data, Srs2 will now be evaluated in a YAC-based end loss system because it does not inherently involve recombination.

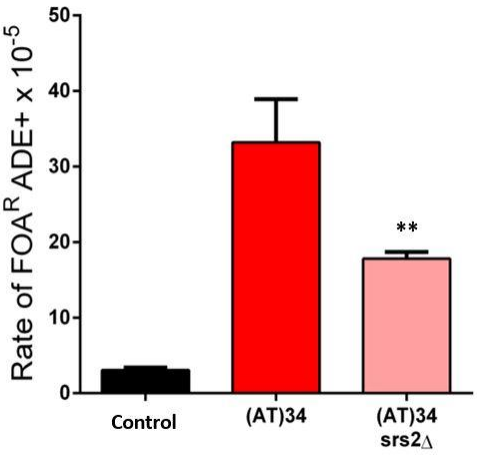


Figure 7. Knocking out *srs2* appears to decrease the rate of recombination at Flex1 (** = p=0.0025). However, since Srs2 has a role in recombination and single strand annealing, it should be evaluated in the YAC system with a fragility assay (SK & CHF unpublished).

Sgs1

Sgs1 is a 3' to 5' helicase and the only protein from the RecQ family that is present in *Saccharomyces cerevisiae*. The RecQ family is a group of proteins that are highly conserved across species and involved in genome stability. The direct homolog of Sgs1 in humans is BLM, which is also a RecQ protein and is implicated in causing Bloom's syndrome. This disorder is characterized by predisposition to cancer and genomic instability, because of the role of BLM

(and Sgs1) in protecting the genome from breakage. Sgs1 stands for Slow Growth Suppressor, because it suppresses the slow growth phenotype of *top3* mutants. Sgs1 has a protective role in genome maintenance, and helps to stabilize DNA polymerase at stalled forks (Cobb et al. 2003).

Sgs1 has the ability to unwind certain secondary structures in vitro. It preferentially unwinds G4 DNA structures (Sun et al. 1998) and can also unwind hairpins (Anand et al. 2011). However, there is not increased fork stalling at either G4T4 sequences or CGG repeats, implying that it is not essential for the unwinding of either G-quadraplex structures or hairpins (Anand et al. 2011). It has not previously been implicated in having a role in unwinding cruciform structures.

The Bloom helicase, a human homolog to Sgs1, has the ability to cause the formation of regressed forks in vitro (Ralf et al. 2006). A regressed, or reversed, replication fork occurs when the replisome "backs up", and the newly synthesized DNA binds to itself, forming a chicken-foot shaped structure. A reversed fork could serve the purpose of allowing the replisome to pause safely while repair machinery removes a lesion. Fork reversal can also be a manner of repair in itself. Since the reversed fork resembles a Holliday junction, recombination-based repair is often able to occur. Also, the reversed fork could create a substrate for nuclease cleavage, to allow fork restart or mitotic DNA synthesis. Since BLM is able to cause fork regression, it is logical that there is a possibility that Sgs1 could play a similar role.

Sgs1 also has multiple roles in homologous recombination, making it influential in this pathway. This manifests itself in Sgs1 being involved in the recombination-based repair process in response to double stranded breaks and other replication impediments. Sgs1, in conjunction with Exo1 and Dna2, is able to resect from a double stranded break in order to prepare the sequence for homologous recombination (Zhu et al. 2008). Sgs1 also forms a complex with Top3

and Rmi1 (called the RMT complex) which is able to dissolve Holliday junctions, meaning that the structures are unwound without crossing over (Ashton & Hickson 2010, Figure 8). In the absence of Sgs1, cells have increased rates of recombination (Gangloff et al. 2000). Additionally, *sgs1*Δ *srs2*Δ strains are not viable, but are able to survive if *rad51* is knocked out as well (Gangloff et al. 2000). This implies that the issue that causes lethality is unresolved recombination intermediates, and that Sgs1 has a role in unwinding them. Strains lacking both *sgs1* and *mus81* also are synthetic lethal, and are both proteins that are required in recombination-based repair (Bastin-Shanower et al. 2003). This role in recombination could also be related to reversed forks, as recombination-based repair can be a form of restarting both stalled and reversed forks.

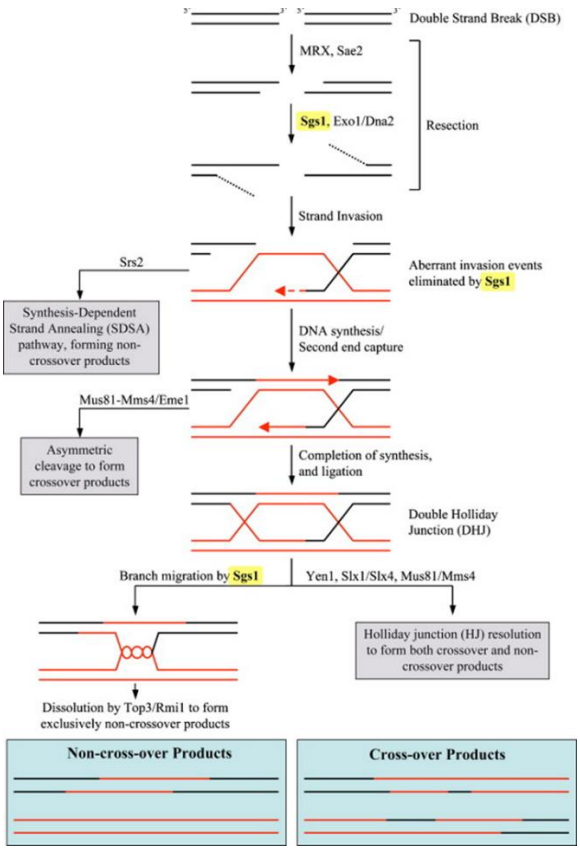


Figure 8. Sgs1 has a role in resection at the initiation of the double strand break, and in dissolution of double Holliday junctions after they form (Ashton & Hickson 2010).

Mph1

Mph1 is a 3’ to 5’ helicase that is known to have roles in unwinding D loops, which has an effect on stabilizing reversed forks and template switching. Mph1 stands for Mutator Phenotype, since deletion of *mph1* causes an increase in rate of spontaneous mutations (Scheller et al. 2000), possibly due to its role in homologous recombination-based repair. Its human homolog is FANCM, which is part of the Fanconi Anemia pathway. Fanconi Anemia is an inherited human disorder that usually manifests itself as an increased risk of cancer, an inability to produce sufficient blood cells, developmental defects, and other physical problems. It is characterized on a molecular level by a less effective response to DNA damage, especially cross-linking agents that stall replication forks. Mph1 fulfills many of the same functions in *Saccharomyces cerevisiae* as FANCM does in the DNA repair pathways in humans.

Mph1 has a role in multiple steps throughout process of fork reversal in DNA replication. Like its human homolog FANCM, Mph1 catalyzes regression of the replication fork in order to begin the formation of the reversed fork (Zheng et al 2011). It also has a role in facilitating branch migration of the reversed fork (Xue et al. 2016). The Smc5-6 complex downregulates the fork reversal and branch migration roles of Mph1 through direct interactions between the two proteins, likely acting as a regulating force to prevent the formation of excessive and toxic joint molecules (Xue et al. 2014). Conversely, the MHF complex upregulates the actions of Mph1 by minimizing how much Smc5 was able to interact with Mph1, since MHF proteins and Smc5 interact with the same domain of Mph1 (Xue et al. 2014). Mte1 is another protein that upregulates Mph1 activity in fork regression and migration, although it has a negative effect on the role of Mph1 in recombination-based repair (Xue et al. 2016). All of these proteins that

regulate Mph1 strive for an equilibrium between creating reversed forks to allow for repair and effective replication and avoiding the creation of toxic joint molecules (Xue et al. 2016, Figure 9A).

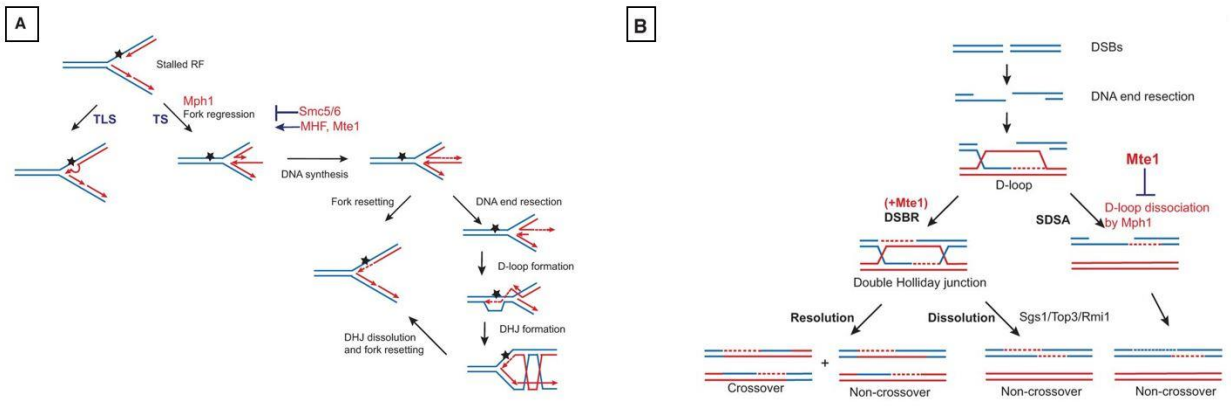


Figure 9. A) Mph1 has a role in fork regression, and is regulated by Smc5-6, MHF, and Mte1. B) Mph1 has a role in the dissolution of the D loop during homologous recombination, leading to noncrossover events. Sgs1, in a complex with Top3 and Rmi1, have the ability to dissolve double Holliday junctions to produce noncrossover events (Xue et al. 2016).

Mph1 also has a role in regulating and controlling homologous recombination, especially recombination that restarts reversed forks. Mph1, using its helicase domain, is able to bind to, unwind and disassociate the D loops that form through the actions of Rad51 (Figure 9B). In doing so, Mph1 reduces the amount of crossover events during recombination. Mph1 can dissociate D loops with 3' overhangs, 5' overhangs, and no overhangs, all with around equal efficiency in vitro (Prakash Satory et al. 2009). Unlike Srs2, Mph1 is able to dissociate the D loop even after it has formed completely. The function of Mph1 in dissociating D loops seems to be linked to its helicase abilities, and it does not have the ability to remove Rad51 from the single stranded DNA. Mph1 is recruited to double stranded breaks, which taken with the role of Mph1 in recombination could imply a role in recombination-based double strand break repair. Although Mph1 does significantly decrease levels of crossing over, it does not decrease overall

efficiency of repair (Prakash Satory et al. 2009), and does not dramatically decrease the frequency of spontaneous recombination (Schurer et al. 2004). Therefore it seems to have a role in reducing crossing over, not as an antirecombinase.

METHODS

Amplification of Knockout Fragment from Genomic DNA

The *rrm3::KANMX4, sgs1::KANMX4, mph1::KANMX4,* and *srs2::TRP1* knockout PCR fragments were created using 20 base pair primers with around 200 base pairs of homology on either side of the gene of interest. For all knockouts, each reaction had a total volume of 12.5 µl and the Sib enzyme was used. The mix included, for each reaction, 7.55 µl ddH_2O, 1.25 µl Empire Genomics 10x reaction buffer, 0.75 µl 20 mM $MgSO_4$, 1.25 µl of each 10 µm primer dilution, 0.25 µl dNTPs, 0.2 µl Sib enzyme, and a small portion of a *S. cerevisiae* patch. The reactions were run on a modified TaqCol program, with an annealing temperature equal to the melting temperature of the primer with the lowest melting temperature for each reaction. Once the knockout fragments were obtained from the PCR reaction, they were concentrated using the ZYMO "DNA Clean and Concentrator" kit.

Transformation of Knockout Fragments

The *rrm3::KANMX4* knockout fragment was transformed into CFY #2525, which has a cassette on chromosome II with a Flex1 (AT)34 sequence. The *srs2::TRP1, sgs1::KANMX4,* and *mph1::KANMX4* knockout fragments were transformed into CFY #3457, which has a Yeast Artificial Chromosome (YAC) with a Flex1 S5'(AT)34S3' o1 sequence. The transformation was performed according to the Freudenreich laboratory "Lithium Acetate Transformation for *S. cerevisiae* for gene disruption by PCR" protocol, with some modifications. CFY # 2525 cells were grown in YEPD media, and CFY#3457 cells were grown in YC-Leu-Ura media. Once cells were made competent through the lithium acetate protocol, they were kept in 20% glycerol stocks in the -80˚C freezer until needed. Once the cells were defrosted, the protocol was

followed, but the addition of DMSO step was skipped, as was the recovery step. After the steps of the protocol were taken, the supernatant removed and the cells were resuspended in 100 µl H_2O. The strains knocked out with *KANMX4* were plated onto YEPD as a recovery plate and then replica plated onto YEPD+G418 the next day. The *srs2::TRP1* strains were plated directly onto YC-Trp.

Checking PCRs: ORF Absent and Junction Checks

Two different PCR checks were performed for all transformants to make sure that the potential transformants were knocked out at the correct locus in the genome. For all checking PCRs, Sib enzyme was used and the volume of the reaction was 12.5 µl. The mix included, for each reaction, 7.55 µl ddH_2O, 1.25 µl Empire Genomics 10x reaction buffer, 0.75 µl 20 mM $MgSO_4$, 1.25 µl of each 10 µm primer dilution, 0.25 µl dNTPs, 0.2 µl Sib enzyme, and a small amount of yeast from the patched-out transformation colonies. ORF absent PCR used internal primers, meaning that both bind to sequences within where the gene of interest would be if the gene was in the genome. The presence of a band means there was an improper transformation. For the junction check, one primer is used that that binds outside the point of knockout fragment insertion and one is used that binds to the marker. For both the ORF absence check and the junction check, the PCR products were run out on 1.5% agarose gels, and 9.0-15.5 µl of a mix of 12.5 µl PCR product and 3 µl loading dye were loaded into the wells. Less liquid (closer to the 9 µl end of the range) were loaded once it became apparent that there might be a problem with overflow from the wells into other lanes.

Direct Duplication Recombination Assay

The direct duplication recombination assay (DDRA) was performed according to the Freudenreich Laboratory "Direct Duplication Recombination Assay" protocol. For Flex1 *rrm3*Δ CFY# 4106 and CFY #4119, 1 µl of suspended cells was plated. For Control *rrm3*Δ CFY#4590 and #4591, 10 µl were plated. A 10^{-5} dilution was plated onto two YEPD plates in order to get the total cell count. All plates were allowed to grow for five days, when colonies are counted and FOA resistance (and recombination rate) were calculated using the method of the median from FALCOR fluctuation analysis calculator. The first assay done on each new mutant was a five colony pilot assay, which was not counted in the final data or rate averages (one for Flex1 *rrm3*Δ and one for Control *rrm3*Δ).

End Loss Assay

The end loss assay was performed according to the "Freudenreich Lab Fluctuation Analysis Procedure (Culture method) For Strains Containing YAC CF" protocol. This YAC is slightly different from the one that was used in these experiments (it has AT repeats instead of CAG repeats), but the protocol is easily modified and applied to Flex1 strains. The one colony inoculation protocol was used, but a small part of a patch was inoculated into 1.5 mL of YC-leu instead of a colony. The initial OD, before the YC-Leu overnight, was 0.03. The protocol was followed, and once plated on FOA-Leu and YC-Leu the cells were allowed to grow for five days. After five days, the colonies were counted and then replica plated onto YC-His to account for point mutations. Anything that grew on YC-His did not have a true breakage and resection event, because the *HIS* gene should be lost when the chromosome is resected. In that situation, it is

possible that there was a point mutation in the *URA* gene that allowed it to grow on FOA. The YC-His plates were counted and compared to their respective FOA-Leu plates. Any colonies that line up with colonies on the FOA-Leu plates were subtracted from the total number of colonies for that plate, because they did not have true breakage events. The rates of fragility were calculated using the method of the median from FALCOR fluctuation analysis calculator. The first assay done on each new mutant was a five colony pilot assay, which was not counted in the final data or rate averages.

RESULTS

Creation of *rrm3*Δ strains containing Flex1 on chromosome II and evaluation of their rates of recombination

The helicase Rrm3 has a role in removing bound proteins from DNA, as well as unwinding R-loops (Tran et al. 2017). In order to determine if Rrm3 has a role in preventing fragility at the Flex1 sequence, the *RRM3* gene was replaced with *KANMX4* in a strain that, because of its background and the cassette that it contained with Flex1, could be used to evaluate rate of recombination. If Rrm3 had a role in preventing fragility, an increase in recombination would be expected in the Flex1 *rrm3*Δ strain as compared to Flex1 wildtype.

Two strains of *rrm3::KANMX4* with Flex1 in the chromosome II DDRA cassette needed to be made so that it was clear that any changes in rate of recombination were because of the lack of *rrm3* and not because of any individual mutations that one strain may have. To make these strains, *rrm3::KANMX4* knockout fragments were created and successfully transformed into a Flex1 *S. cerevisiae* strain, and saved as a yeast stock. This was the first strain of Flex1 strains that had been knocked out for *rrm3*. The *rrm3::KANMX4* knockout fragment for the first transformation was very close in size to the expected 2,080 base pairs, around 2,000 base pairs (Figure 10A). There was one colony on the YEPD+G418 replica plate, which grew when patched out onto another YPED+G418 plate.

The method of ORF absent checking PCR is usually performed in order to determine if a possible transformant is knocked out for the correct gene. If the transformation was successful, no band is expected on the gel. If the transformation was not successful, the band will be seen. In the ORF absent check PCR for the Flex1 *rrm3*Δ strain, the transformant did not have a band, which implies a successful transformation (Figure 10B). If the ORF was present, and the gene

had not been knocked out correctly, a band of 610 base pairs would have been visible. In a junction check PCR, the transformant had PCR fragments that were very close in size to the expected size, 688 base pairs (Figure 10C). The expected size would only be seen if the KANMX4 marker was incorporated in the correct place in the genome. That transformant was saved as a yeast stock with the number CFY #4106 and the name Flex1 *rrm3::KANMX4*.

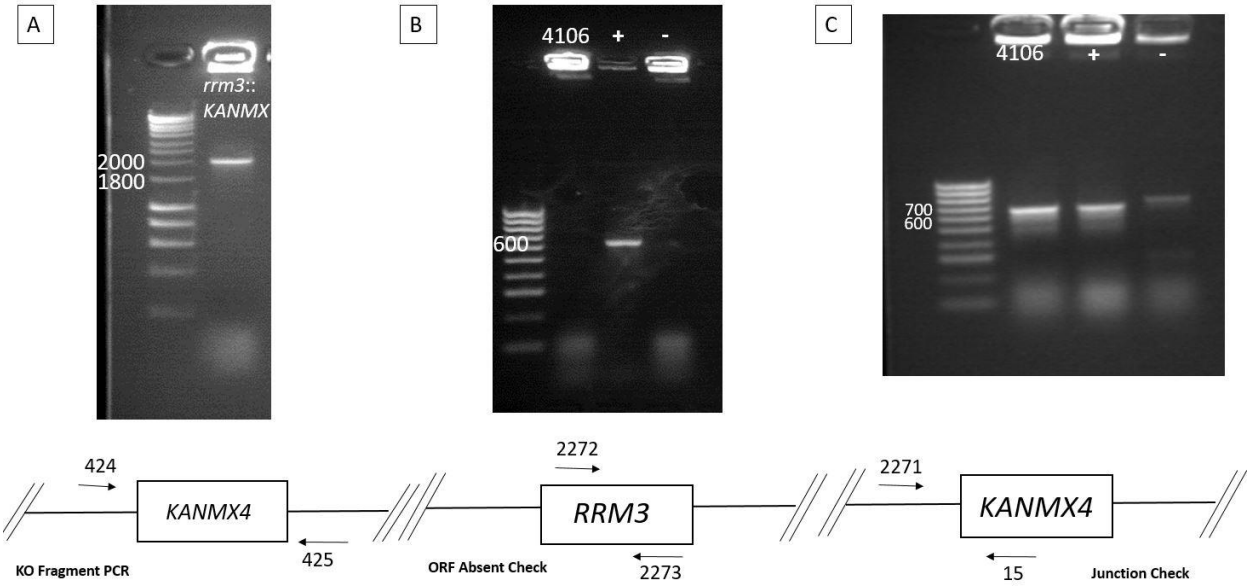


Figure 10. Creation of CFY #4106 ch.II Flex1 *rrm3::KANMX4* A) Amplification of the knockout fragment for *rrm3::KANMX4*. The expected band size was 2,080 base pairs, and the actual size is very close to that. The knockout fragment diagram shows where the primers in relation to the gene. B) ORF Absent PCR for the singular possible transformant. The strain that would become CFY #4106 had no band, as is expected of a proper transformant. C) 5' junction check for potential transformant. The transformant had a band at the proper size, around 688 base pairs, so it passed the junction check and was saved as a glycerol stock.

To get another Flex1 *rrm3::KANMX4* strain, the transformation was done again. The knockout PCR was repeated, and once again the band was the correct size, close to 2,080 base pairs (Figure 11A). The knockout fragment was concentrated and 0.93 µg of DNA was used, so that there would be a greater chance of the cells being transformed. The transformants that grew were patched out on a plate with G418, and then an ORF absent check was done on 21 of them.

Of these 21 patched out colonies, one did not have a band and was carried forward as a potential correct transformant (Figure 11B). That one transformant passed the 5' junction test by having a band at the correct size of 688 base pairs, (Figure 11C). A 3' junction check was also performed on the potential transformant. The band may be slightly larger than expected (it is close to 1,000 base pairs when it should have been 902 base pairs), but it is the same as other strains made earlier, so it was moved forward and was saved as a yeast stock with the number CFY #4119 (Figure 11D).

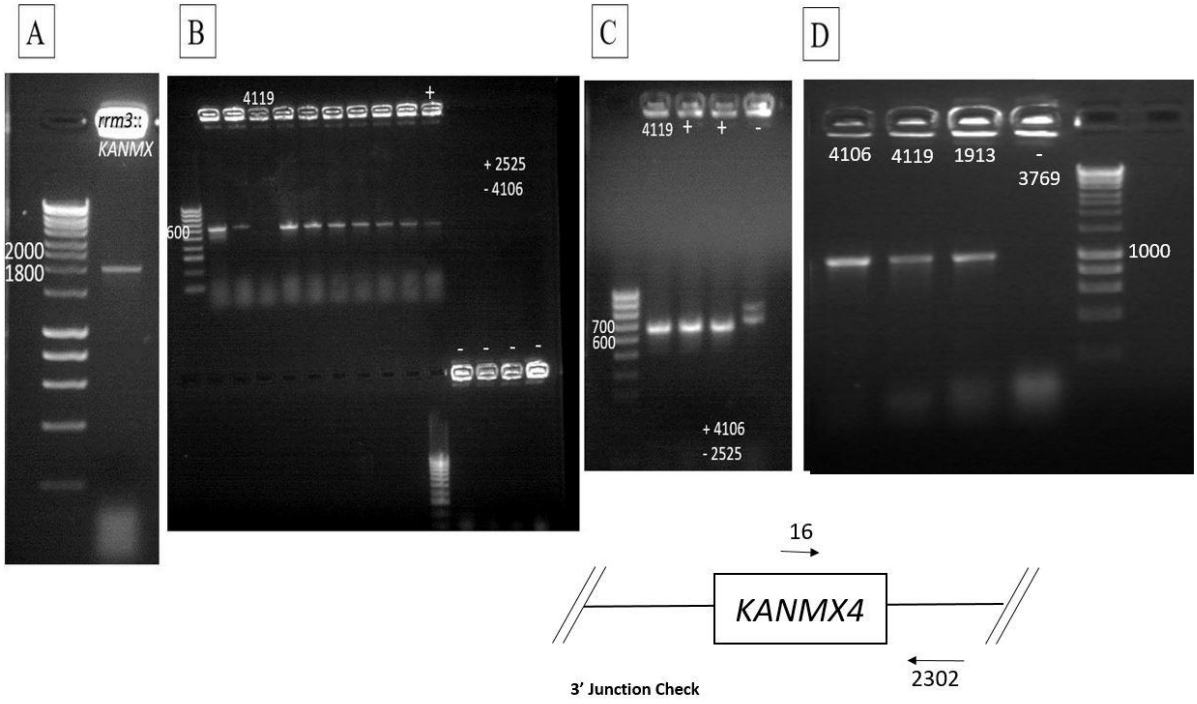


Figure 11. Creation of CFY #4119 ch.II Flex1 *rrm3::KANMX* A) Amplification of the knockout fragment for *rrm3::KANMX*. The expected band size was 2,080 base pairs, and the actual size is very close to that. The band is somewhat feint, so the DNA was concentrated so that 0.93 µg of DNA were used for the transformation. B) ORF Absent PCR for the possible transformants. The strain that would become CFY #4119 had no band, as is expected of a proper transformant. C) 5' junction check for potential transformant. The transformant had a band at the proper size, around 688 base pairs, so it passed the junction check. D) 3' junction check for potential transformant. The band was around the expected band of 900, below 1,000 base pairs. The band size might be slightly larger than expected, but it is the same as other strains made earlier, so it is moved forward. The 3' junction diagram shows where the primers bind.

The direct duplication recombination assay (DDRA) uses a cassette incorporated onto chromosome II that contains two separated parts of the *ade2* gene, with homologous sequences. If these parts were next to each other they would make a full functional *ADE2* gene, but apart they are nonfunctional. Between the two nonfunctional *ade2* genes, there is a *URA*3 gene and a Flex1 sequence. The Flex1 sequence contains 34 non-interrupted AT repeats with flanking sequences of human DNA on either side. If a double stranded break occurs at Flex1, recombination is stimulated, and the two nonfunctional parts of the *ade2* gene to recombine and become a fully functioning *ADE2* gene through single strand annealing. Through this process, the Flex1 sequence and *URA3* will be lost. When the cells are plated on FOA-Ade, only cells that have undergone breakage and recombination will grow. Only those cells that have a functional *ADE2* (occurring through recombination) and lack a *ura3 gene* (lacking a *ura3* gene gives cells FOA resistance) will survive on these plates (Figure 12). Deletions at the FRA16D locus are very common (Bignell et al. 2010), and the majority of these deletions include the Flex1 region (Ried et al. 2000). Our assay mimics events that occur naturally in cancer cells or cells under replicative stress, just in a way that we can measure.

Direct duplication recombination assays were performed on the Flex1 *rrm3*Δ strains to determine the rate of recombination, which implies the rate of DNA breakage. Six assays were performed in total, three from each strain. Average recombination rates were calculated for all assay using the Lea-Coulson Method of the Median in the Fluctuation Analysis Calculator (FALCOR) program (Hall et al. 2009). The average rate of recombination was determined to be 26.1×10^{-5} $FOA^{R}Ade^{+}$. The Flex1 *rrm3*Δ rate showed a significant decrease in average rate of recombination compared to the rate of recombination for wildtype Flex1 strains, which was 39.3

x 10^{-5} $FOA^{R}Ade^{+}$ (Figure 13). The next step was to delete *RRM3* in a control strain to see if a similar effect was seen.

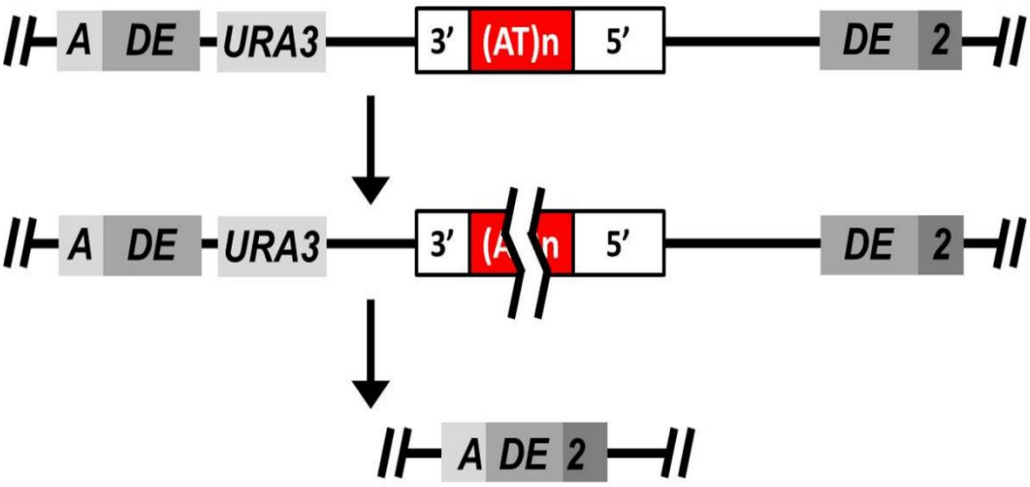
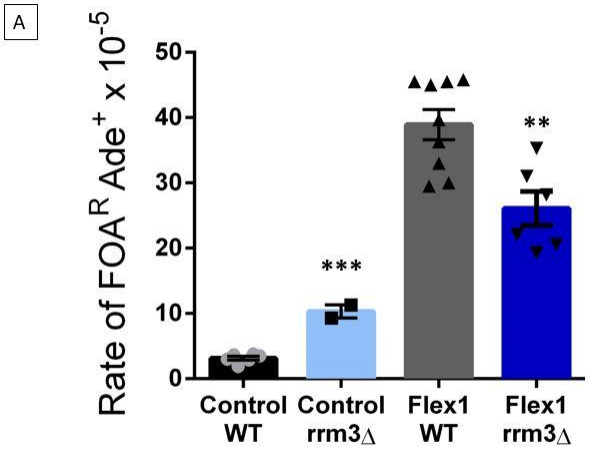


Figure 12. The chromosome II cassette used for direct duplication recombination assays (DDRA) is made of two nonfunctional parts of an *ade2* gene, with a *URA3* gene and Flex1 AT repeat sequence in between them. When breakage occurs in Flex1, resection causes the nonfunctional *ade2* parts to recombine and become fully functioning, and the *URA3* gene and F1 sequence are lost. Because of this, recombination can be selected for by plating on FOA-Ade. FOA selects for the loss of the *URA3* gene and the lack of adenine selects for non-separated *ADE2* gene (SK & CHF).



B

Control WT	Control *rrm3Δ*	Flex1 WT	Flex1 *rrm3Δ*
2.9	9.3	45.0	20.6
3.8	11.3	39.7	19.3
1.9		29.5	28.2
3.7		45.5	35.3
3.0		45.5	22.1
3.5		30.0	31.0
		33.0	
		45.8	
		36.3	
Average: 3.1	**Average: 10.3**	**Average: 38.9**	**Average: 26.1**

Figure 13. A) The rate of recombination of Flex1 *rrm3*Δ is significantly higher than the wildtype with the same cassette of DNA (** =p=0.0084 compared to Flex1 WT, 1-way ANOVA), but the rate of recombination of Control *rrm3*Δ is significantly higher than the wildtype with the control cassette (**** = p<0.0001 compared to WT Control, t test). B) A table of all assays on Flex1 *rrm3*Δ and Control *rrm3*Δ strains thus far. All numbers are rate of FOA^R Ade^+ x 10^{-5}.

Creation of *rrm3*Δ strains containing a control sequence on chromosome II and evaluation of their rates of recombination

The recombination assays proved that the absence of Rrm3 seems to cause a decrease in rate of recombination, and therefore rate of fragility, at Flex1. However, these results do not show whether or not the decrease in fragility was sequence specific. There was no way of knowing whether the decrease in fragility was due to the way that Rrm3 was unable to interact with something at the Flex1 repeats, or if it was a more universal effect. Before testing the control strain, it was unknown whether the fragility decrease was due to an artifact of the assay or a more general feature of Rrm3 that was not specific to its role at the Flex1 repeat. In order to determine if these results were specific to Flex1, *rrm3* was knocked out in a control strain. Specifically, Rrm3 could have an uncharacterized role in single strand annealing, either in

resection or annealing. This strain contained a cassette on chromosome II that could be evaluated with the same recombination assay. Instead of having Flex1 on the cassette, there was a control sequence that contained a sequence from FRA16D that did not have any repeats predicted to form secondary structures. If the Control *rrm3*Δ strain had a decreased rate of recombination compared to that of the wildtype Control strain, the implication would be that the lowered rate of fragility was due to a more universal role of Rrm3, and was not due to any interactions with Flex1.

To make these strains, *rrm3::KANMX4* knockout fragments were created and successfully transformed into a control *S. cerevisiae* strain, which contained a chromosome II cassette with a sequence from FRA16D which was not predicted to form secondary structures. The *rrm3::KANMX4* knockout fragment for the first transformation was very close in size to the expected 2,080 base pairs, around 2,000 base pairs (Figure 14A). Around 1.2 µg of DNA knockout fragment were transformed into the Control strain. Eleven colonies grew on the YEPD+G418 replica plate, and all also grew when patched out onto another YPED+G418 plate.

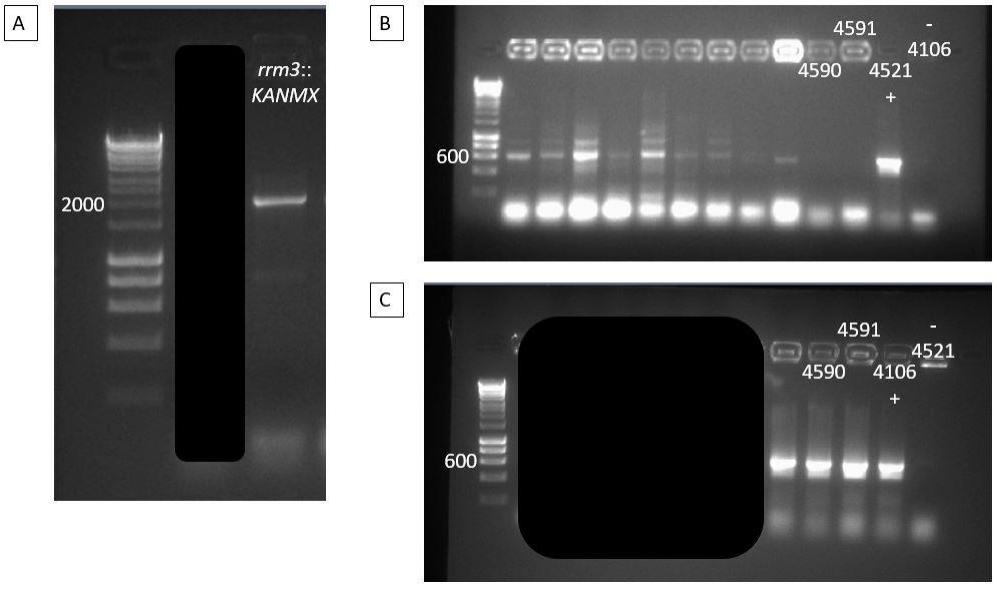


Figure 14. Creation of CFY# 4590 and #4591, chromosome II Control *rrm3::KANMX*. The same primers were used for knockout fragment creation, ORF absent check, and junction check as were used for the Flex1 *rrm3*Δ, and the diagrams are the same. A) Amplification of the knockout fragment for *rrm3::KANMX*. The expected band size was 2,080 base pairs, and the actual size is very close to that. 1.2 µg of DNA was used in the transformation. B) An ORF Absent PCR for the possible transformants was performed. These strains had no band, as is expected of a proper transformant. If they contained an intact ORF, the band size would have been around 610 base pairs. C) A junction check for potential transformant. The transformants had a band at the proper size, around 688 base pairs, so they passed the junction check.

An ORF absent check PCR was performed for the potential Control *rrm3*Δ transformants, and three of the transformants did not have a band, which implies a successful transformation (Figure 14B). If the ORF was present, and the gene had not been knocked out correctly, a band of 610 base pairs would have been visible. A 5' junction check PCR was performed, and all three potential transformants had PCR fragments that were very close in size to the expected size, 688 base pairs (Figure 14C). Two of the transformants were saved as yeast glycerol stocks with the numbers CFY #4590 and CFY #4591, and the name Control *rrm3::KANMX4*.

Direct Duplication Recombination Assays were performed on these strains to determine the rate of recombination. At this point in time, two ten-plate assays have been performed, with an average of 10.3 FOA^R Ade^+ x 10^{-5}. This was an increase in recombination when compared to the wildtype Control strain, which has a rate of recombination of 3.1 x 10^{-5} FOA^RAde^+. (Figure 13).

Rrm3 does have a specific role at the Flex1 AT repeat. However, the results are not consistent with what would be expected if Rrm3 was unwinding cruciforms in order to prevent fragility. The effect may relate to the role of Rrm3 in removing R-loops, due to the similarity of the effect of Rnh1 at the Flex1 repeat.

Creation of *srs2*Δ strains containing Flex1 on a Yeast Artificial Chromosome and evaluation of their rates of fragility

The helicase Srs2 has many roles in promoting functional replication. It can unwind secondary structures such as hairpins, and also serves both antirecombinase and recombinase roles through its strippase ability of disassembling Rad51-ssDNA nucleoprotein filament. In this experiment the Yeast Artificial Chromosome system was used to test the rate of fragility of strains in which *srs2* was replaced with a *TRP1* marker gene.

For the construction of these strains, *srs2::TRP1* knockout fragments were created and successfully transformed into a *S. cerevisiae* strain which contained a Yeast Artificial Chromosome (YAC) with the human Flex1 region on it. The *srs2::TRP1* knockout fragment for the transformation was very close in size to the expected 1357 base pairs, as it was between 1,000 and 1,500 base pairs (Figure 15A). Around 1.0 µg of DNA knockout fragment was transformed into the Flex1 strain. Many colonies grew on the selective YC-TRP transformation plates, and twenty six colonies were patched onto an additional YC-TRP plate.

An ORF absent check PCR was performed for the potential Flex1 *srs2*Δ transformants, and nineteen of the twenty six potential transformants did not have a band, implying successful transformation (Figure 15B). If the ORF had present, and the gene had not been knocked out correctly, a band of 1,387 base pairs would have been visible. A 3' junction check PCR was performed on those nineteen potential transformants, and eighteen of them had PCR fragments that were very close in size to the expected size, around 938 base pairs (Figure 15C). This means that they passed both checks and were confirmed as correct *srs2::TRP1* transformants. Two of the transformants were saved as yeast glycerol stocks with the numbers CFY #4558 and CFY #4559. Later on, when there were some doubts about the integrity of one of those strains, two

more transformants were saved as yeast glycerol stocks with the numbers CFY #4588 and CFY#4589.

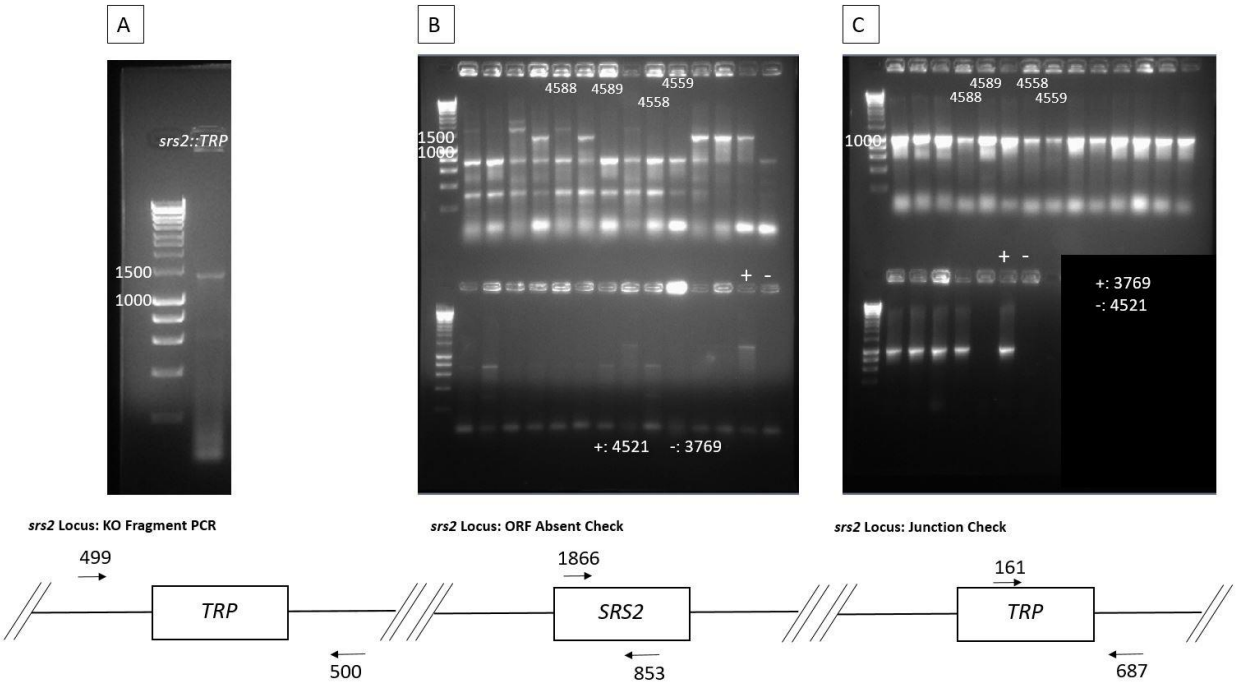


Figure 15. Creation of CFY #4558, #4559, #4588, and #4559 YAC Flex1 *srs2::TRP1* A) Amplification of the knockout fragment for *srs2::TRP1*. The expected band size was 1,387 base pairs, and the actual size is very close to that. B) ORF Absent PCR for the possible transformants. These strains had no band, as is expected of a proper transformant. C) Junction check for potential transformant. The transformants had a band at the proper size, around 938 base pairs, so they passed the junction check. The diagrams show where the primers bind.

Srs2 is known to have a role in recombination as both a recombinase and antirecombinase, and specifically has been shown by the Haber laboratory to have a role in single strand annealing (Sugawara et al. 2000). Because of this, the rates of fragility of these strains were tested using a YAC system in an end loss assay, which is not as dependent on recombination for the assay to function (Figure 16). The end loss assay is fairly similar to the direct duplication recombination assay, but it requires end loss and telomere addition instead of recombination to recover the broken chromosome. If breakage occurs at the region of interest, in this case AT repeats from the Flex1 sequence, resection can occur from the break to the telomere

seed sequence. Yeast telomerase can then add on a telomere to the G4T4 sequence, so that the left arm of the YAC continues to work as a functional chromosome. When this occurs, everything that was to the right of the telomere seed sequence (G4T4 sequence) is lost. This includes the Flex1 sequence and the *URA3* and *HIS3* genes. When the cells are plated onto FOA-Leu, only the cells that have gone through breakage should survive, since only cells without a functional *URA3* gene can grow on FOA. The media without Leu selects for cells that still have the right arm of the YAC (selecting against cells that lost the whole YAC), and the FOA in the plates selects for cells that have lost the *URA* gene.

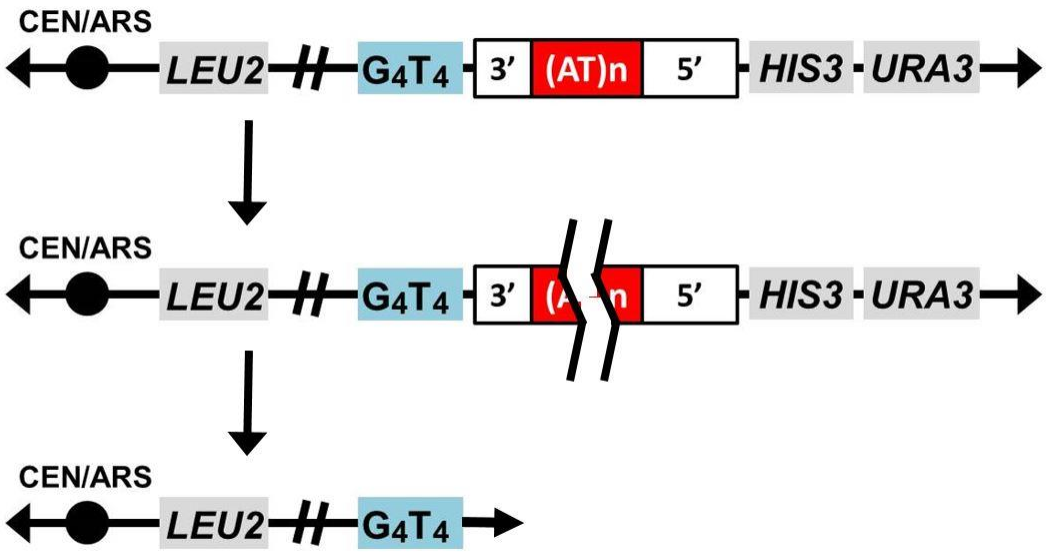


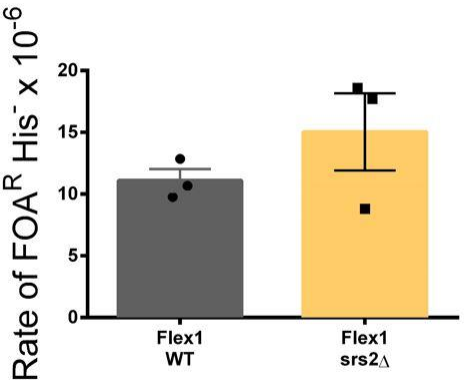
Figure 16. The YAC has the *LEU* gene on one arm of the YAC, and a telomere seed sequence, F1 AT repeat, *HIS* gene, and *URA* gene on the other. If breakage occurs at the AT repeat, there can be resection from F1 to the telomere seed sequence. Everything to the right of the G4T4 sequence is lost, and telomerase adds the telomere to the seed sequence. Because of this, fragility can be selected for using FOA-Leu media, which selects for the remaining left arm of the YAC because of the Leu and selects for the loss of *ura* with the FOA.

The media without His (onto which the FOA-leu colonies are replica plated) selects for cells that have an intact *HIS3* gene. Since the *HIS3* gene should be lost during the breakage event (and therefore no cells that grew on FOA-Leu should be able to grow on YC-His), any colonies

that grow on FOA-Leu and also on YC-His should be discounted from the total count of that FOA-Leu plate. If a colony grows on both FOA-Leu and YC-His, it is implied that there was not an actual breakage event, and instead a point mutation in the *URA3* gene could have occurred.

In addition to a pilot assay, three end loss assays were performed on the *srs2::TRP1* strains, CFY #4558 and #4559. Then an additional assay was performed on #4558. However, shortly after these assays were done it became clear that #4559 was *his* negative, meaning that it did not have an intact *HIS3* gene on the YAC. This could be indicative of a broken YAC, or a point mutation within the *HIS3* gene. In any case, without a functional *HIS3* gene, it is impossible to check for point mutations in *URA3* gene, so the assay done on strain #4559 was removed from the data set. That is the reason that two additional *srs2*Δ strains were saved. An end loss assay was performed on one of those strains (CFY#4588). Of the usable assays, the average rate of end loss was 15.0 FOA^R His^- x 10^{-6}. Based this rate of fragility, which is implied by the rate of FOA^R HIS^-, the absence of *srs2* does not seem to significantly affect the rate of breakage at Flex1 (Figure 17).

A



B

Flex1 WT (Rate of FOA^R His^- x 10^{-6})	**Flex1 *srs2*Δ (Rate of FOA^R His^- x 10^{-6})**
10.67	17.7
12.86	18.6
9.78	8.8

Average: 11.1	Average: 15.0

Figure 17. A) The rate of end loss of Flex1 *srs2*Δ is not significantly different from the rate of end loss of wildtype Flex1 (p=0.2943, t test). B) A table of all assays on wildtype Flex1 and Flex1 *srs2*Δ.

Creation of *sgs1*Δ strains containing Flex1 on a Yeast Artificial Chromosome and evaluation of their rates of fragility

The helicase Sgs1 has a role in dissolving double Holliday junctions and promoting noncrossing over in recombination, potentially in recombination-based repair to restart reversed forks. It has also been implicated in unwinding secondary structures in vitro, and its absence increases fragility of a CAG repeat (Kerrest et al. 2009). This helicase was tested by creating strains that were deleted for *sgs1*. Because Sgs1 has a role in long-range resection, the recombination assay could not be used. Instead, the YAC end loss assay is better suited to evaluating the role of Sgs1 in fragility at the Flex1 repeat. In the recombination assay, resection of more than 2,000 base pairs must occur in both directions from the Flex1 region (the location of the break) to the partial *ADE2* genes. In the YAC end loss system, around 300 base pairs of resection are required. Short range resection is considered to be around 100-200 base pairs, so the YAC assay technically contains "short" long range resection, and is better suited for evaluating Sgs1 than the recombination assay. This small amount of long range resection is not a serious concern because Sgs1 works with Exo1 in its long-range resection role, so both would need to be knocked out to eliminate long range resection. Additionally, if there was a lack of long range resection occurring with the lack of Sgs1, one would expect levels of fragility to be underrepresented, because without resection from the break to the telomeric seed sequence,

telomere addition and healing could not occur. Therefore the YAC end loss assay is the best choice for examining the role of Sgs1.

For the construction of these strains, *sgs1::KANMX4* knockout fragments were created and successfully transformed into a *S. cerevisiae* strain which contained a YAC with the human Flex1 region on it. The transformation needed to be repeated because the first transformation only yielded one correct transformant. The *sgs1::KANMX4* knockout fragment for the first transformation had a band very close in size to the expected 1,990 base pairs, around 2,000 base pairs. However, it also had another equally bright band around 1,400 base pairs (Figure 18A). The larger band (the correct band) was gel purified to make sure that the DNA that was being transformed into the Flex1 YAC strain was only the correct knockout fragment. Four colonies grew on the selective YEPD + G418 transformation plates, and were patched out onto another YEPD + G418 plate.

An ORF absent check PCR was performed on the four potential Flex1 *sgs1*Δ transformants. Two of the potential transformants had an obvious band around 859 base pairs, implying that they contained the ORF and that *SGS1* was not correctly knocked out (Figure 18B). However, one of the potential transformants and the negative control both had light bands, so the ORF absent check was repeated with new primers on the two transformants that were not obviously wrong. From that ORF absent check, it was clearer that only one of the potential transformants was lacking the ORF, as the other one had a band at 743 base pairs (Figure 18C). A 5' junction check PCR was performed on those two potential transformants, and both passed the check. They had PCR fragments that were very close in size to the expected size, around 452 base pairs (Figure 18D). The one transformant *sgs1*Δ that passed all checks was saved as yeast glycerol stocks with the numbers CFY #4521.

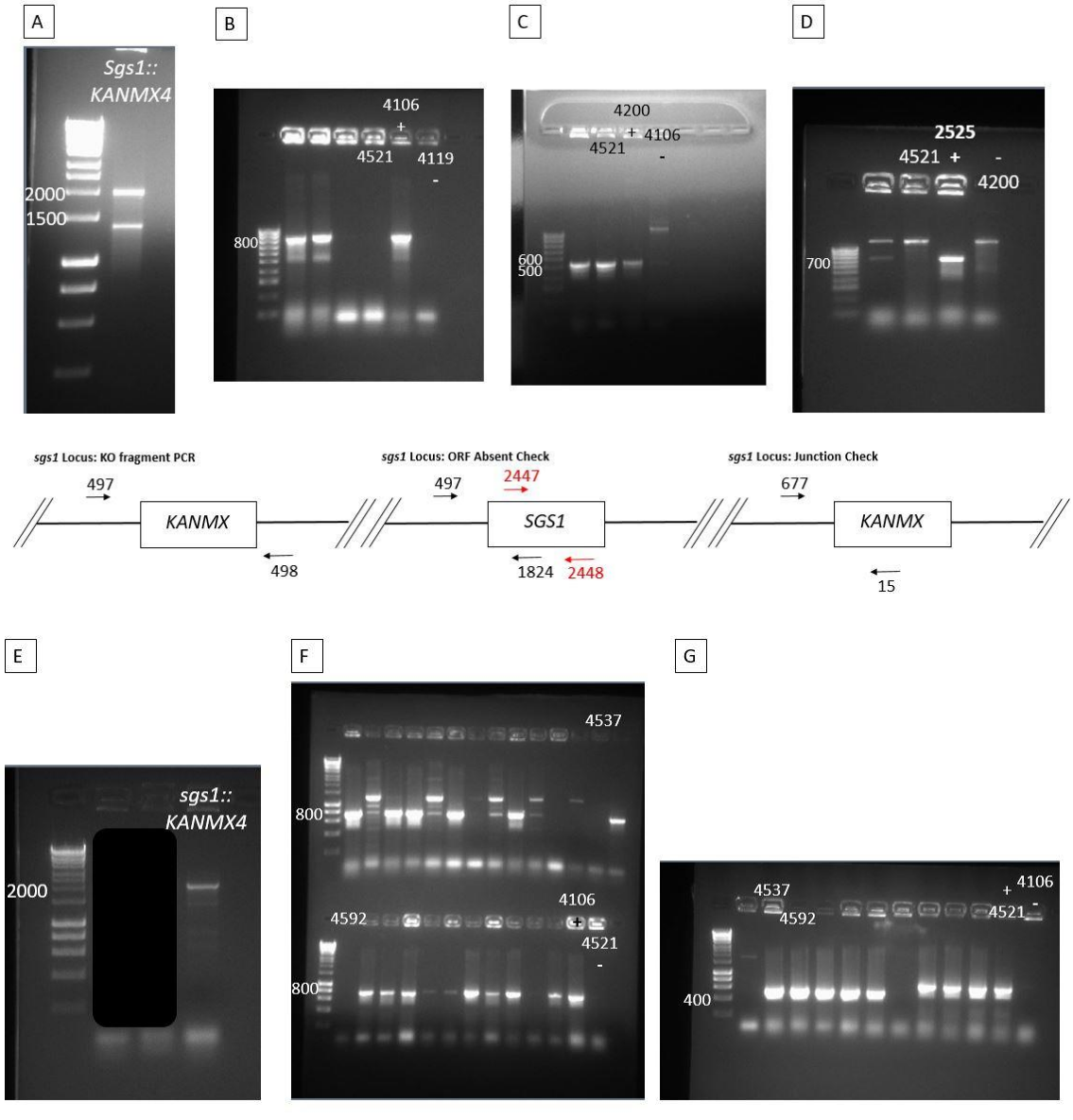


Figure 18. Creation of CFY #4521 YAC Flex1 *sgs1::KANMX* A) Amplification of the knockout fragment for *sgs1::KANMX*. The correct band around 1,990 base pairs was gel purified. B) An ORF absent PCR check was performed, but there was a light band in the negative control. The primers for the first ORF absent check are in black on the diagram. C) A junction check was performed for the potential transformants that passed the ORF absent check. Both of the potential transformants had a band near the proper size, around 524 base pairs. D) A second ORF absent PCR with new primers was performed on both potential transformants that passed the junction check, since the first ORF absent check was not conclusive. Only one transformant passed the test by having no bands. The expected band size for incorrect transformants was 743 base pairs, as is seen in the positive control and the other potential transformant. The primers for the second ORF absent check are in red on the diagram. Creation of CFY #4538 and #4592 YAC Flex1 *sgs1::KANMX4* E) Amplification of the knockout fragment for *sgs1::KANMX*. The band is at the right size, around 1,990 base pairs. F) An ORF absent check was performed, using the

same primers as the second ORF absent check for CFY #4119. The strains that were carried forward did not have a band at 743 base pairs, meaning that they passed the ORF absent check. G) A junction check was performed for the potential transformants that passed the ORF absent check. CFY #4537 and #4592 both had bands at 452 base pairs, and so passed the check and were saved as glycerol stocks.

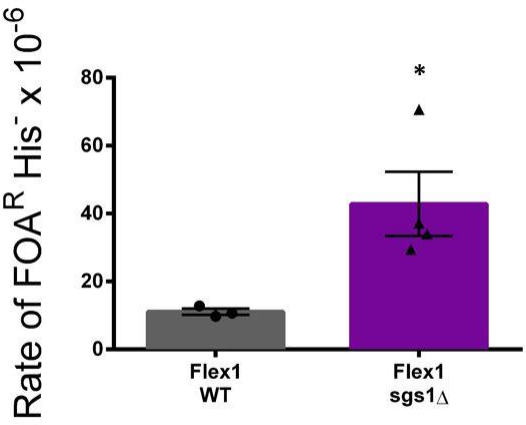
Another transformation was done in order to get the second Flex1 *sgs1*Δ transformant. More *sgs1::KANMX4* knockout fragment needed to be created, and around 1.0 μg of DNA was used (Figure 18E). Many colonies grew on the selective YEPD + G418 plates, and 26 of those were checked to determine if *sgs1* was correctly replaced with the *KANMX4* marker.

An ORF absent check PCR was performed for the potential Flex1 *sgs1*Δ transformants, and four of the twenty six potential transformants did not have a band, which implies that they did not have the ORF (Figure 18F). If the ORF had been present, and the gene had not been knocked out correctly, a band of 743 base pairs would have been visible. A 5' junction check PCR was performed on those four potential transformants, and all of them had PCR fragments that were very close in size to the expected size, 452 base pairs (Figure 18G). This means that they passed both checks and were confirmed as correct Flex1 *sgs1::KANMX4* transformants. Two of the *sgs1*Δ transformants were saved as yeast glycerol stocks with the numbers CFY #4537 and CFY #4592.

End loss assays were performed on these strains in order to determine the rate of fragility in Flex1 *sgs1*Δ strains, and compare it to the fragility of wildtype Flex1 strains (Figure 19). The two assays performed on strain #4521 had rates that were very close to one another, with an average between the two of them of 31.8×10^{-6} FOA^R HIS^-. However, one of the assays that has thus far been performed on CFY #4537 had a rate that was more than double that, 70.7×10^{-6} FOA^R HIS^- (the other within the range of what would have been expected based on the assays from CFY#4521. In the future, an assay using CFY #4592 will be done to see if it also has a

large range of rates of end loss. Overall, the rate of end loss of Flex1 *sgs1*Δ (which implies rate of fragility) is significantly increased compared to that rate of end loss for wildtype Flex1, signifying that Sgs1 does seem to have a role in preventing fragility at Flex1. Among potential explanations include a role in recombination-based repair and restarting reversed forks that could form upon encountering secondary structures at Flex1, or in unwinding those same secondary structures. However, more tests are required to determine exactly which of the roles of Sgs1 are affecting fragility at Flex1.

A



B

Flex1 WT (Rate of FOA^R His$^-$ x 10^{-6})	**Flex1 *sgs1*Δ (Rate of FOA^R His$^-$ x 10^{-6})**
10.67	29.5
12.86	34.0
9.78	70.7
	37.2
Average: 11.1	**Average:** 42.9

Figure 19. A) The rate of end loss, and therefore fragility, of Flex1 *sgs1*Δ is significantly greater than the rate of end loss of wildtype Flex1 (*=p=0.0361 compared to Flex1 WT, t test). B) A table of all assays on Flex1 *sgs1*Δ and wildtype Flex1.

Creation of *sgs1*Δ *rad51*Δ strains containing Flex1 on a Yeast Artificial Chromosome and evaluation of their rates of fragility

Since Sgs1 has so many roles, it is difficult to determine which of its functions caused the increase in fragility in Flex1 *sgs1*Δ strains. The Rad51 protein has an important role in recombination, as it binds to single stranded DNA and the Rad51 filament is instrumental in the search for homology. Without Rad51 there are issues with recombination, and therefore problems with the efficiency and effectiveness of recombination-based repair, which could be used to heal breaks that occur within Flex1. Therefore *sgs1*Δ *rad51*Δ Flex1 strains were created in order to test if the role of Sgs1 in recombination-based repair was the reason for the increase in fragility at Flex1.

A transformation was done to create Flex1 *sgs1*Δ *rad51*Δ strains. PCR was used to make *rad51::NATMX* knockout fragment, which had a size of around 1,748 base pairs (Figure 20). About 0.94 μg of DNA was used in the transformation. Many colonies grew on the selective YEPD + NAT plates, but checking PCR reactions have not yet been done to determine whether they are correct transformants.

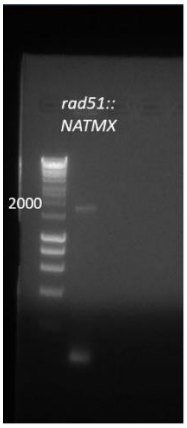


Figure 20. A *rad51::NATMX* knockout fragment was created. The band is at the correct size, around 1,748 base pairs. This knockout fragment (and more made using the same PCR protocol) was used to create *rad51::NATMX* knockout fragment to transform into both Flex1 *sgs1*Δ and Flex1 *mph1*Δ strains.

Creation of *mph1*Δ strains containing Flex1 on a Yeast Artificial Chromosome and evaluation of their rates of fragility

The helicase Mph1 has a role in recombination, since it has a role in the dissociation of D loops. We decided to test this protein because of its role in fork reversal, and recovery from fork stalling from crosslinks. Since it has a somewhat similar role to Sgs1 and Srs2 (in that they are all involved in recombination, with potential roles in fork regression), the role of Mph1 in fragility was also evaluated using the end loss YAC assay system, so that the effects of the proteins could be accurately compared to one another.

For the construction of these strains, *mph1::KANMX4* knockout fragments were created and successfully transformed into a *S. cerevisiae* strain which contained a YAC with the human Flex1 region on it. The *mph1::KANMX4* knockout fragment for the first transformation had a band very close in size to the expected 1,854 base pairs, between 1,500 and 2,000 base pairs (Figure 21A). Around 0.66 μl of DNA were used in the transformation. Many colonies grew on the selective YEPD + G418 transformation plates, and twenty two were patched out onto another YEPD + G418 plate.

Twenty one of those patches grew, and were tested with an ORF absent PCR to determine if they were true transformants that were actually missing the *MPH1* gene. In this PCR reaction, five of the transformants did not have a band, implying a successful transformation (Figure 21B). The potential transformants for which the gene had not been knocked out correctly had a band of around 1,598 base pairs. Then a 3' junction check PCR was performed, and all five of the transformants had PCR fragments that were very close in size to the expected size, 776 base pairs (Figure 21C). However, some of them had large, bright smudges, and the negative

control had a band that was smaller than 776 bases but still present, around 700 base pairs. Because of this, the 3' junction check was repeated. The result was the same with less smudged PCR bands, and confirmed that that those transformants were correct and were knocked out for *mph1*. Two of the transformants were saved as a yeast stocks with the number CFY #4538 and #4539, and the name Flex1 *mph1::KANMX4*.

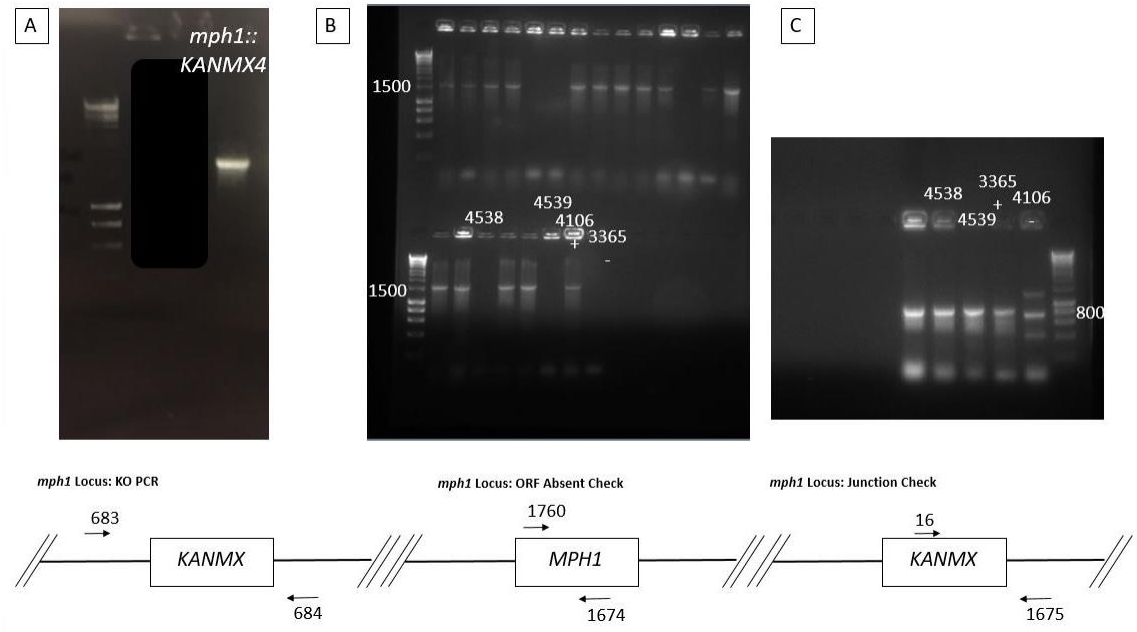
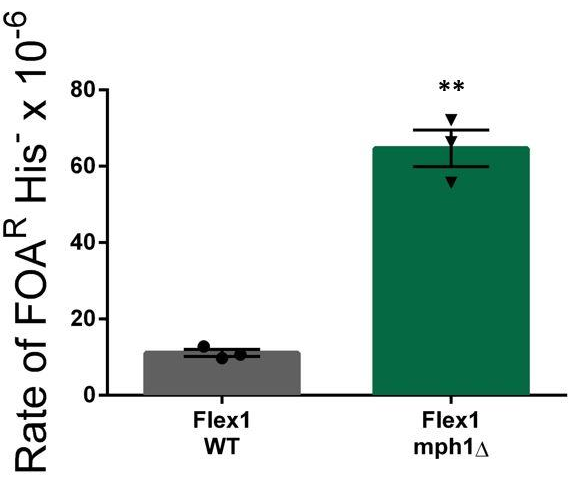


Figure 21. Creation of CFY #4538 and #4539 YAC Flex1 *mph1::KANMX* A) Amplification of the knockout fragment for *mph1::KANMX*. The expected band size was 1,854 base pairs. Although it is difficult to see due to bad picture quality, the actual size is very close to that. B) ORF Absent PCR for the possible transformants. These strains had no band, as is expected of a proper transformant. C) Junction check for potential transformant. The transformants had a band at the proper size, around 776 base pairs, so they passed the junction check. The diagrams show where the primers bind.

End loss assays were performed on these strains in order to determine the rate of fragility in Flex1 *mph1*Δ strains, and compare it to the fragility of wildtype Flex1 strains. The Flex1 *mph1*Δ strains had a higher rate of end loss, 64.7 x 10^{-6} FOA^R HIS^-, (implying a higher rate of fragility) compared to the wildtype Flex1 strains (Figure 22). This implies that Mph1 has a role in preventing fragility at Flex1, perhaps through the formation of reversed forks.

A



B

Flex1 WT (Rate of FOA^R His$^-$ x 10^{-6})	Flex1 *mph1*Δ (Rate of FOA^R His$^-$ x 10^{-6})
10.67	55.7
12.86	72.1
9.78	66.3
Average: 11.1	**Average: 64.7**

Figure 22. A) The rate of end loss, and therefore fragility, of Flex1 *mph1*Δ is significantly greater than the rate of end loss of wildtype Flex1 (***=p=0.0004 compared to Flex1 WT, t test). B) A table of all assays on Flex1 *mph1*Δ and wildtype Flex1.

Creation of *mph1*Δ *rad51*Δ strains containing Flex1 on a Yeast Artificial Chromosome and evaluation of their rates of fragility

Because Mph1 is so multifunctional, it is not possible without further testing to determine which of its roles, when absent, was the cause of this fragility at Flex1. Rad51 can be used to determine if the cause of the fragility is related to recombination, in a similar way to how it is being used with Sgs1. In order to investigate the role of Mph1 at Flex1, *mph1*Δ *rad51*Δ strains containing a YAC with Flex1 are in the process of being created.

A transformation was done to create Flex1 *mph1*Δ *rad51*Δ strains. PCR was used to create *rad51::NATMX* knockout fragment, which had a size of around 1,748 base pairs (Figure 20). Around 0.85 μg of DNA were used in the transformation. Many colonies grew on the selective YEPD + NAT plates, but checking PCR reactions have not yet been done to determine whether they are correct transformants.

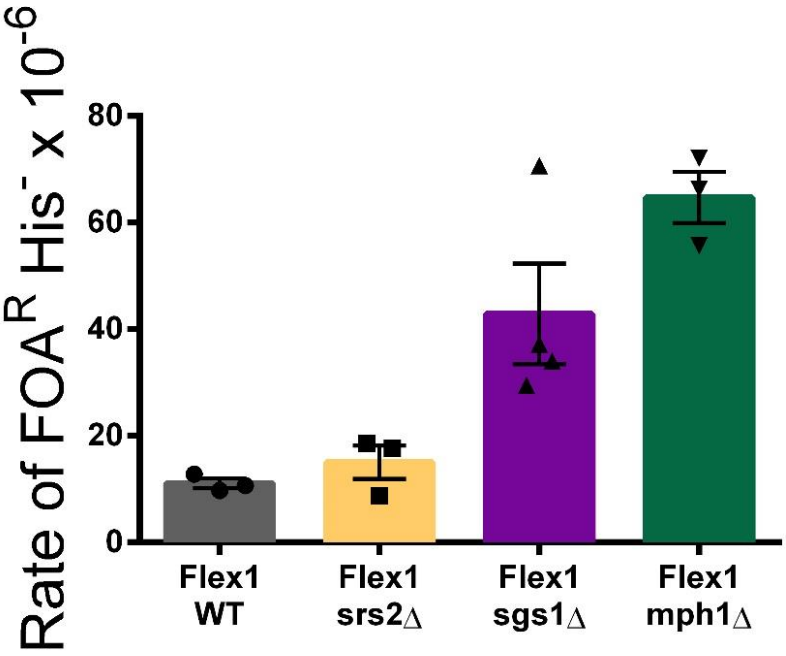


Figure 23. There is no significant difference in rate of end loss between wildtype Flex1 and Flex1 *srs2*Δ. There is however, a significant increase in end loss (compared to wildtype Flex1) for both Flex1 *sgs1*Δ and Flex1 *mph1*Δ. That implies that both Sgs1 and Mph1 have a role in preventing breakage at Flex1.

DISCUSSION

Rrm3

The recombination assays from the Flex1 *rrm3*Δ strains imply that a lack of *rrm3* causes a decrease in the rate of fragility at Flex1. According to these results, the presence of *RRM3* could actually increase fragility. These results were unexpected, and surprising. If there was a bound protein or a secondary structure at Flex1 that Rrm3 was able to remove, an increase in fragility would have been expected in its absence. If there was no impediments to replication at Flex1 that would be affected by Rrm3, no change in fragility would have been expected compared to wildtype. Rrm3 binding sites in *S. cerevisiae* are associated with loss of heterozygosity and breakpoints in recombination events (Song et al 2014), so there is record of Rrm3 being connected to breakage events. However, its role at these fragile sites is not known with absolute certainty.

One potential explanation for these results has to do with the idea that Rrm3 has a role in forming reversed forks in *rad53*Δ checkpoint-deficient cells, because Rad53 is not present to phosphorylate Rrm3. Rad53 phosphorylation of Rrm3 keeps Rrm3 in check, and prevents it from forming reversed forks at an undesired rate (Rossi et al 2015). Freudenreich laboratory member Simran Kaushal has shown that Flex1 *sml1*Δ *mec1*Δ strains do not significantly differ in rate of recombination from wildtype Flex1 (Kaushal et al. unpublished). Since Mec1 is a checkpoint protein (ATR in humans) that activates Rad53, this shows that the ATR checkpoint does not seem to be important for the prevention of Flex1 fragility. Cells with Flex1 may act similarly to checkpoint-deficient cells, since the checkpoint is not activated in either situation. If Rrm3 is present but not phosphorylated by Rad53, it could be overactive in forming reversed forks. These reversed forks could be substrates for Mus81 cleavage (Kaushal et al. in prep), which would

cause fragility. Based on this model, cells that are *rrm3*Δ would form fewer reversed forks, because Rrm3 would not be present to cause their formation. That would mean there would be less substrates available for Mus81 to cleave, causing the decrease in fragility seen in our results. One way to investigate this theory further would be to create phospho-defective (unable to be phosphorylated) and phospho-mimicking (acts as if it is phosphorylated) mutants of *RRM3*, and expose all to a replication stressor such as HU, similar to the experiments of the Giannattasio group (Rossi et al 2015). It would be possible to see if the phospho-defective mutant strains reacted to replication stress in the same way as *rrm3*Δ strains, and better understand if the phosphorylation of Rrm3 is important.

Another hypothesis is based on the known role of Rrm3 in unwinding DNA:RNA hybrids, or R loops. It was considered that an R loop could have the capability of forming at Flex1, and preventing the DNA from popping out into a cruciform. This would operate under the assumption that R loops are less deleterious than cruciforms (which has not been proven). If Rrm3 unwound this R loop, a cruciform would be able to form, and the region would be more susceptible to breakage. If Rrm3 was absent, the R loop would stay in place, and no cruciform would form. Another Freudenreich laboratory member Ruby Ye evaluated the rate of recombination of strains that were Flex1 *rnh1*Δ, since Rnh1 is a protein that has a role in degrading the RNA of RNA-DNA hybrids, or R loops. The rate of recombination was decreased compared to wildtype, very similar to the level of Flex1 *rrm3*Δ rates (Figure 24). The Flex1 *rnh1*Δ rate of recombination were not significantly lower than the wildtype Flex1 rate, but not significantly higher than the *rrm3*Δ Flex1 rate. Of the *rnh1*Δ Flex1 assays, there was one assay that had a high rate of recombination, while the other values were very close to the rates of *rrm3*Δ Flex1. This could imply that the ability of Rrm3 to unwind R loops is the reason for the

decrease in recombination with *rrm3*Δ. Or, Rrm3 could have some other role that is not currently known.

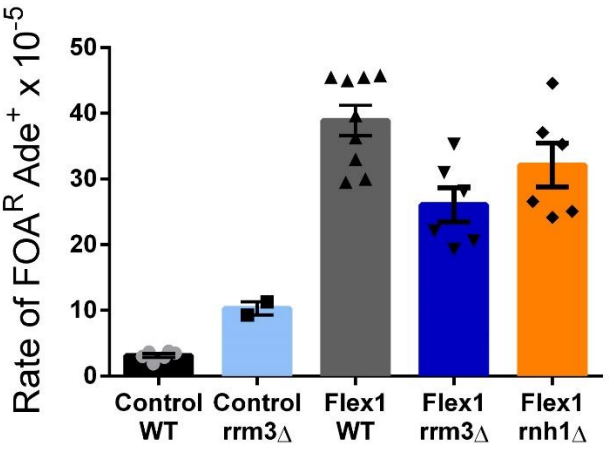


Figure 24. This graph contains the same data as was portrayed in the Rrm3 section of the results section, but it has added Flex1 *rnh1*Δ recombination rates, which were collected by Freudenreich lab member Ruby Ye. *rnh1* has an intermediate rate of recombination, between the rate for wildtype Flex1 and the rate for Flex1 *rrm3*Δ. Although the Flex1 *rrm3*Δ rate of recombination is significantly lower than the wildtype Flex1 rate, the Flex1 *rnh1*Δ is not significantly different from either of them (p=0.1988 compared to Flex1 WT, p=0.3284 compared to Flex1 *rrm3*Δ, Analyzed using 1-Way ANOVA, multiple comparisons).

The results of the recombination assays from Control *rrm3*Δ imply that genome-wide, the lack of *rrm3* causes an increase in the rate of fragility. This could mean that that the presence of *RRM3* has a protective role against breakage in the genome as a whole, but not specifically at Flex1. Or, the results could signify that *RRM3* has some previously uncharacterized effect on the assay, such as a role in inhibiting single strand annealing. Another possibility is that the control strain may have a binding site for some protein that is removed by Rrm3, and in the absence of Rrm3 it is possible that there are increased collisions with this protein, causing increased fragility. One such protein that could be removed by Rrm3 (and therefore not removed when *rrm3* is absent) is RNA polymerase II, which binds to the DNA (for example, within *URA3)* in the chromosome II cassette during transcription (Kaushal et al. unpublished). There is increased

fork pausing at RNA polymerase II transcribed genes in *rrm3*Δ cells (Azvolinksy et al 2009), so it is possible that in the absence of *RRM3* the fork is unable to continue. Overall, the Control *rrm3*Δ results were somewhat puzzling in light of the Flex1 *rrm3*Δ results. It seems counter-intuitive that Rrm3 would have a protective role in the genome in general (as implied by this control strain), but at Flex1 would be implicated in causing even more breakage. It is not entirely clear why there would be such a large difference between the effects of Rrm3 on the Flex1 AT repeat compared to the effects of Rrm3 on the rest of the genome, or at least the control strain. However, this result could imply that since *rrm3*Δ seems to cause an increase in fragility in general, the significant but somewhat modest decrease in fragility seen at Flex1 is actually representative of a large decrease in fragility.

Since Rrm3 and Pif1 are both Pif1 helicases and have many overlapping functions, it could be interesting to knock them both out together, as one might be acting as a backup to the other. For example, Rrm3 can unwind G4 forming sequence in *pif1*Δ cells (Paeschke et al 2013), and Pif1 could unwind R loops and remove bound proteins at tDNA when *rrm3* is absent (Tran 2017). Since *pif1*Δ strains produce many petites, this would likely be a strain with *rrm3*Δ and a mutated *PIF1* gene, such as *pif1-m2*.

Srs2

The end loss YAC assays performed on the Flex1 *srs2*Δ strains imply that the lack of Srs2 does not have an effect on fragility at the Flex1 AT repeat. This was somewhat unexpected, because Srs2 is known to have a role in recombination-based repair, and because it has the ability to unwind some secondary structures such as CAG and CGG repeats (Anand et al. 2011). It is possible that Flex1, if breaking, could be repaired by some other repair pathway that does

not involve any of the putative roles of Srs2, neither secondary structure unwinding nor recombination nor fork regression. Additionally, Srs2 has been implicated in unwinding hairpins, but not necessarily cruciforms. Since it is not positive that a cruciform is forming at Flex1 and not just two hairpins across from one another, the lack of effect of Srs2 could lend some support to the idea that a cruciform, not two hairpins, is forming at Flex1.

Sgs1

The end loss assays performed on the Flex1 *sgs1*Δ strains imply that the rate of breakage at Flex1 is increased when Sgs1 is not present. This could have occurred for a variety of reasons, because Sgs1 has many roles. Some likely causes that had been hypothesized were that Sgs1 either had a role in unwinding secondary structures at Flex1, or in the stabilization of reversed forks. Perhaps Sgs1 is able to form a reversed fork, like its homolog BLM. Such a structure could form in order to allow fork restart or mitotic DNA synthesis in order to continue replication past the secondary structure at Flex1. In the absence of Sgs1, these pathways would have been less effective, and breakage could have occurred.

Alternatively, it is also possible that Sgs1 has a role in recombination-based repair which could have a role in healing a double stranded break at Flex1. In order to determine whether the increase in fragility with the absence of Sgs1 was due to problems with recombination-based repair, double knockout strains are currently being created that are *sgs1*Δ *rad51*Δ. Additionally, in the future, it will be necessary to create Flex1 *rad51* strains to interpret the double knockout results. The rate of end loss of *rad51*Δ strains with a YAC with CAG repeats is published, and shows a minimal increase in rate of FOA^R (Sundararajan et al 2010).

If Sgs1 and Rad51 are working through two different pathways (for example, if Sgs1 unwinds secondary structures and Rad51 promotes effective recombination-based repair), then the double knockout would be expected to have an additive effect, the rate of end loss being similar to the rate of end loss for *sgs1*Δ strains and the rate of end loss for *rad51*Δ strains combined. If Sgs1 and Rad51 were working in the same pathway, one would expect the end loss rate of the double knockout to be similar to the rate of *sgs1*Δ strains, and of *rad51*Δ strains. Also, if *rad51*Δ has little effect on rate of end loss, the implication would be that Rad51 and Mph1 are working in two different pathways, since the rate of Flex1 *mph1*Δ would be much larger than the rate of Flex1 *rad51*Δ. Additionally, if the role of Sgs1 at Flex1 is due to recombination, this recombination could take place after fork regression as a way to restart a reversed fork.

Since Sgs1 forms complexes with quite a few proteins in order to carry out its many functions, it might be interesting to knock out some of these proteins in the future to gain a better understanding of the roles of Sgs1 at Flex1. For example, the dissolvase role of Sgs1 is carried out in a complex with Top3 and Rmi1. It could be of interest to test the rate of fragility of a Flex1 *top3*Δ strain, and how it compares to *sgs1*Δ strains. Additionally, it would likely be fruitful to create a Control *sgs1*Δ strain in order to determine if this increase in fragility in the absence of Sgs1 is a genome-wide effect, or if it is specific to Flex1.

Mph1

The end loss YAC assays performed on the Flex1 *mph1*Δ strains show that the rate of fragility increases at Flex1 when *mph1* is absent, implying that Mph1 has a protective role against breakage at Flex1. This could be for a multitude of different reasons, including the characterized role of Mph1 in forming reversed forks. For example, fork reversal can occur in

order for replication machinery to be able to recover, and remove an impediment. It is possible that a reversed fork could be required to allow a secondary structure at Flex1 to be removed. Or, perhaps a reversed fork must form, and then be cleaved by Mus81 in order to allow fork restart or MiDAS repair synthesis to fill in across the AT repeat (proposed model of Kaushal et al. in prep). If this is true, then the process of fork reversal could be crucial to avoiding fragility at Flex1.

However, it is also possible that in the absence of *mph1*, recombination-based repair is less efficient. This recombination-based repair could occur in response to a double stranded break, or as a way to restart a reversed fork. In order to determine if the cause of the increase in fragility at Flex1 was due to the recombination role of Mph1, a double knockout is in the process of being made that is Flex1 *mph1*Δ *rad51*Δ. Rad51 was chosen because it is involved in mitotic recombination-based repair, and is part of the D loop that Mph1 is able to disassociate.

If Mph1 and Rad51 are working through two different pathways, then the double knockout would be expected to have an additive effect. If Mph1 and Rad51 were working in the same pathway, one would expect the end loss rate of the double knockout to be similar to the rate of *mph1*Δ strains, and of *rad51*Δ strains. Also, if *rad51*Δ has little effect on rate of end loss, it would imply that Rad51 and Mph1 are working in two different pathways, since the rate of *mph1*Δ strains would be much larger than that of *rad51* strains.

Since Mph1 is involved in the formation reversed forks, it is possible that Mph1 contributes to the creation of a reversed fork at Flex1 when the replication machinery encounters a secondary structure. If this is the case, the regressed fork could form and be cleaved by Mus81 to allow fork restart or MiDAS repair synthesis. Based on this possible model, in the absence of *mph1* no reversed fork would form at Flex1 (or, if it did form, would have less efficient branch

migration and therefore not form into the best possible structure). Without this way to repair itself the cells could experience breakage, possibly because of irreparably stalled forks or collisions between the replisome and the secondary structures. In the future, it would be relevant to create Control *mph1*Δ strains. It could also be interesting to evaluate the rate of fragility of a strain that was Flex1 *mph1*Δ *mus81*Δ, due to the potential role of Mus81 in cleaving reversed forks for survival of the cell and continuation of replication. Since Sgs1 and Mph1 both have characterized roles in the process of fork reversal, it could also be interesting to make Flex1 *sgs1*Δ *mph1*Δ strains.

Conclusion

Although all these proteins are helicases, it is obvious that they all affect the Flex1 AT repeat in different ways. These differences are due to the vast array of functions that each helicase has, and the different pathways they are a part of. Because of this, makes sense that the knockout of each helicase would not cause the same amount of fragility at Flex1. Even those with overlapping roles, such as formation of stalled forks or dissolution of the D loop in homologous recombination, each have different substrates or mechanisms. By determining how these helicases function and prevent (or do not prevent) breakage at the Flex1 AT repeat, it becomes possible to gain an even deeper understanding of the mechanisms of breakage at Flex1. This information could potentially be applied to the larger FRA16D common fragile site, or to other common fragile sites in general. It could be even, sometime in the future, be applied to attempting to minimize breakage at FRA16D in order to mitigate its effects on cancer proliferation.

REFERENCES

Anand, R. P., Shah, K. A., Niu, H., Sung, P., Mirkin, S. M., & Freudenreich, C. H. (2011). Overcoming natural replication barriers: differential helicase requirements. *Nucleic acids research*, *40*(3), 1091-1105.

Antony, E., Tomko, E. J., Xiao, Q., Krejci, L., Lohman, T. M., & Ellenberger, T. (2009). Srs2 disassembles Rad51 filaments by a protein-protein interaction triggering ATP turnover and dissociation of Rad51 from DNA. *Molecular cell*, *35*(1), 105-115.

Aqeilan, R. I., Donati, V., Gaudio, E., Nicoloso, M. S., Sundvall, M., Korhonen, A., ... & Croce, C. M. (2007). Association of Wwox with ErbB4 in breast cancer. *Cancer research*, *67*(19), 9330-9336.

Ashton, T. M., & Hickson, I. D. (2010). Yeast as a model system to study RecQ helicase function. *DNA repair*, *9*(3), 303-314.

Azvolinsky, A., Dunaway, S., Torres, J. Z., Bessler, J. B., & Zakian, V. A. (2006). The S. cerevisiae Rrm3p DNA helicase moves with the replication fork and affects replication of all yeast chromosomes. *Genes & development*, *20*(22), 3104-3116.

Azvolinsky, A., Giresi, P. G., Lieb, J. D., & Zakian, V. A. (2009). Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Molecular cell*, *34*(6), 722-734.

Bastin-Shanower, S. A., Fricke, W. M., Mullen, J. R., & Brill, S. J. (2003). The mechanism of Mus81-Mms4 cleavage site selection distinguishes it from the homologous endonuclease Rad1-Rad10. *Molecular and cellular biology*, *23*(10), 3487-3496.

Bednarek, A. K., Keck-Waggoner, C. L., Daniel, R. L., Laflin, K. J., Bergsagel, P. L., Kiguchi, K., ... & Aldaz, C. M. (2001). WWOX, the FRA16D gene, behaves as a suppressor of tumor growth. *Cancer research*, *61*(22), 8068-8073.

Bignell, G. R., Greenman, C. D., Davies, H., Butler, A. P., Edkins, S., Andrews, J. M., ... & Widaa, S. (2010). Signatures of mutation and selection in the cancer genome. *Nature*, *463*(7283), 893.

Carter, S. D., Vigašová, D., Chen, J., Chovanec, M., & Åström, S. U. (2009). Nej1 recruits the Srs2 helicase to DNA double-strand breaks and supports repair by a single-strand annealing-like mechanism. *Proceedings of the National Academy of Sciences*, *106*(29), 12037-12042.

Chavdarova, M., Marini, V., Sisakova, A., Sedlackova, H., Vigasova, D., Brill, S. J., ... & Krejci, L. (2015). Srs2 promotes Mus81–Mms4-mediated resolution of recombination intermediates. *Nucleic acids research*, *43*(7), 3626-3642.

Cimprich, K. A. (2003). Fragile sites: breaking up over a slowdown. *Current biology*, *13*(6), R231-R233.

Cobb, J. A., Bjergbaek, L., Shimada, K., Frei, C., & Gasser, S. M. (2003). DNA polymerase stabilization at stalled replication forks requires Mec1 and the RecQ helicase Sgs1. *The EMBO journal*, *22*(16), 4325-4336.

Dayn, A., Malkhosyan, S., Duzhy, D., Lyamichev, V., Panchenko, Y., & Mirkin, S. (1991). Formation of (dA-dT) n cruciforms in Escherichia coli cells under different environmental conditions. *Journal of bacteriology*, *173*(8), 2658-2664.

Durkin, S. G., & Glover, T. W. (2007). Chromosome fragile sites. *Annu. Rev. Genet.*, *41*, 169-192.

Elango, R., Sheng, Z., Jackson, J., DeCata, J., Ibrahim, Y., Pham, N. T., ... & Ira, G. (2017). Break-induced replication promotes formation of lethal joint molecules dissolved by Srs2. *Nature communications*, *8*(1), 1790.

Finnis, M., Dayan, S., Hobson, L., Chenevix-Trench, G., Friend, K., Ried, K., ... & Richards, R. I. (2005). Common chromosomal fragile site FRA16D mutation in cancer cells. *Human molecular genetics*, *14*(10), 1341-1349.

Franchitto, A. (2013). Genome instability at common fragile sites: searching for the cause of their instability. *BioMed research international*, *2013*.

Gangloff, S., Soustelle, C., & Fabre, F. (2000). Homologous recombination is responsible for cell death in the absence of the Sgs1 and Srs2 helicases. *Nature genetics*, *25*(2), 192.

Glover, T. W. (2006). Common fragile sites. *Cancer letters*, *232*(1), 4-12.

Glover, T. W., Arlt, M. F., Casper, A. M., & Durkin, S. G. (2005). Mechanisms of common fragile site instability. *Human molecular genetics*, *14*(suppl_2), R197-R205.

Glover, T. W., Wilson, T. E., & Arlt, M. F. (2017). Fragile sites in cancer: more than meets the eye. *Nature Reviews Cancer*, *17*(8), 489.

Güven, G. S., Hacihanefioğlu, S., & Cenani, A. (1999). Expression of aphidicolin, FUdR and caffeine-induced fragile sites in lymphocytes of healthy Turkish individuals. *Genetica*, *105*(1), 109-116.

Hall, B. M., Ma, C. X., Liang, P., & Singh, K. K. (2009). Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria–Delbrück fluctuation analysis. *Bioinformatics*, *25*(12), 1564-1565.

Hashash, N., Johnson, A. L., & Cha, R. S. (2011). Regulation of fragile sites expression in budding yeast by MEC1, RRM3 and hydroxyurea. *J Cell Sci*, *124*(2), 181-185.

Helmrich, A., Ballarino, M., & Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular cell*, *44*(6), 966-977.

Ivessa, A. S., Zhou, J. Q., Schulz, V. P., Monson, E. K., & Zakian, V. A. (2002). Saccharomyces Rrm3p, a 5′ to 3′ DNA helicase that promotes replication fork progression through telomeric and subtelomeric DNA. *Genes & development*, *16*(11), 1383-1396.

Keil, R. L., & McWilliams, A. D. (1993). A gene with specific and global effects on recombination of sequences from tandemly repeated genes in Saccharomyces cerevisiae. *Genetics*, *135*(3), 711-718.

Kerrest, A., Anand, R. P., Sundararajan, R., Bermejo, R., Liberi, G., Dujon, B., ... & Richard, G. F. (2009). SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. *Nature structural & molecular biology*, *16*(2), 159.

Le Beau, M. M., Rassool, F. V., Neilly, M. E., Espinosa III, R., Glover, T. W., Smith, D. I., & McKeithan, T. W. (1998). Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Human molecular genetics*, *7*(4), 755-761.

Lemoine, F. J., Degtyareva, N. P., Lobachev, K., & Petes, T. D. (2005). Chromosomal translocations in yeast induced by low levels of DNA polymerase: a model for chromosome fragile sites. *Cell*, *120*(5), 587-598.

Letessier, A., Millot, G. A., Koundrioukoff, S., Lachagès, A. M., Vogt, N., Hansen, R. S., ... & Debatisse, M. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature*, *470*(7332), 120.

Macheret, M., & Halazonetis, T. D. (2015). DNA replication stress as a hallmark of cancer. *Annual Review of Pathology: Mechanisms of Disease*, *10*, 425-448.

Macris, M. A., & Sung, P. (2005). Multifaceted role of the Saccharomyces cerevisiae Srs2 helicase in homologous recombination regulation. *Biochem Soc Trans*. 33(6), 1447-50.

Madireddy, A., Kosiyatrakul, S. T., Boisvert, R. A., Herrera-Moyano, E., García-Rubio, M. L., Gerhardt, J., ... & Aguilera, A. (2016). FANCD2 facilitates replication through common fragile sites. *Molecular cell*, *64*(2), 388-404.

Marini, V., & Krejci, L. (2010). Srs2: the “odd-job man” in DNA repair. *DNA repair*, *9*(3), 268-275.

Marini, V., & Krejci, L. (2012). Unwinding of synthetic replication and recombination substrates by Srs2. *DNA repair*, *11*(10), 789-798.

Minocherhomji, S., Ying, S., Bjerregaard, V. A., Bursomanno, S., Aleliunaite, A., Wu, W., ... & Hickson, I. D. (2015). Replication stress activates DNA repair synthesis in mitosis. *Nature*, *528*(7581), 286.

Muñoz-Galván, S., García-Rubio, M., Ortega, P., Ruiz, J. F., Jimeno, S., Pardo, B., ... & Aguilera, A. (2017). A new role for Rrm3 in repair of replication-born DNA breakage by sister chromatid recombination. *PLoS genetics*, *13*(5), e1006781.

Niu, H., & Klein, H. L. (2017). Multifunctional roles of Saccharomyces cerevisiae Srs2 protein in replication, recombination and repair. *FEMS yeast research*, *17*(2).

Nunez, M. I., Rosen, D. G., Ludes-Meyers, J. H., Abba, M. C., Kil, H., Page, R., ... & Aldaz, C. M. (2005). WWOX protein expression varies among ovarian carcinoma histotypes and correlates with less favorable outcome. *BMC cancer*, *5*(1), 64.

O'keefe, L. V., Colella, A., Dayan, S., Chen, Q., Choo, A., Jacob, R., ... & Richards, R. I. (2010). Drosophila orthologue of WWOX, the chromosomal fragile site FRA16D tumour suppressor gene, functions in aerobic metabolism and regulates reactive oxygen species. *Human molecular genetics*, *20*(3), 497-509.

Paeschke, K., Bochman, M. L., Garcia, P. D., Cejka, P., Friedman, K. L., Kowalczykowski, S. C., & Zakian, V. A. (2013). Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, *497*(7450), 458.

Prado, F., & Aguilera, A. (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *The EMBO journal*, *24*(6), 1267-1276.

Prakash, R., Satory, D., Dray, E., Papusha, A., Scheller, J., Kramer, W., ... & Ira, G. (2009). Yeast Mph1 helicase dissociates Rad51-made D-loops: implications for crossover control in mitotic recombination. *Genes & development*, *23*(1), 67-79.

Ralf, C., Hickson, I. D., & Wu, L. (2006). The Bloom's syndrome helicase can promote the regression of a model replication fork. *Journal of biological chemistry*, *281*(32), 22839-22846.

Richards, R. I., Choo, A., Lee, C. S., Dayan, S., & O'Keefe, L. (2015). WWOX, the chromosomal fragile site FRA16D spanning gene: its role in metabolism and contribution to cancer. *Experimental Biology and Medicine*, *240*(3), 338-344.

Ried, K., Finnis, M., Hobson, L., Mangelsdorf, M., Dayan, S., Nancarrow, J. K., ... & Baker, E. (2000). Common chromosomal fragile site FRA16D sequence: identification of the FOR gene spanning FRA16D and homozygous deletions and translocation breakpoints in cancer cells. *Human molecular genetics*, *9*(11), 1651-1663.

Sáez, M. E., González-Pérez, A., Martínez-Larrad, M. T., Gayán, J., Real, L. M., Serrano-Ríos, M., & Ruiz, A. (2010). WWOX gene is associated with HDL cholesterol and triglyceride levels. *BMC medical genetics*, *11*(1), 148.

Scheller, J., Schürer, A., Rudolph, C., Hettwer, S., & Kramer, W. (2000). MPH1, a yeast gene encoding a DEAH protein, plays a role in protection of the genome from spontaneous and chemically induced damage. *Genetics*, *155*(3), 1069-1081.

Schmidt, K. H., & Kolodner, R. D. (2004). Requirement of Rrm3 helicase for repair of spontaneous DNA lesions in cells lacking Srs2 or Sgs1 helicase. *Molecular and cellular biology*, *24*(8), 3213-3226.

Schürer, K. A., Rudolph, C., Ulrich, H. D., & Kramer, W. (2004). Yeast MPH1 gene functions in an error-free DNA damage bypass pathway that requires genes from homologous recombination, but not from postreplicative repair. *Genetics*, *166*(4), 1673-1686.

Seong, C., Colavito, S., Kwon, Y., Sung, P., & Krejci, L. (2009). Regulation of Rad51 recombinase presynaptic filament assembly via interactions with the Rad52 mediator and the Srs2 anti-recombinase. *Journal of Biological Chemistry*, *284*(36), 24363-24371.

Song, W., Dominska, M., Greenwell, P. W., & Petes, T. D. (2014). Genome-wide high-resolution mapping of chromosome fragile sites in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, *111*(21), E2210-E2218.

Sugawara, N., Ira, G., & Haber, J. E. (2000). DNA length dependence of the single-strand annealing pathway and the role of Saccharomyces cerevisiae RAD59 in double-strand break repair. *Molecular and cellular biology*, *20*(14), 5300-5309.

Sun, H., Karow, J. K., Hickson, I. D., & Maizels, N. (1998). The Bloom's syndrome helicase unwinds G4 DNA. *Journal of Biological Chemistry*, *273*(42), 27587-27592.

Sung, P., & Klein, H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature reviews Molecular cell biology*, *7*(10), 739.

Torres, J. Z., Schnakenberg, S. L., & Zakian, V. A. (2004). Saccharomyces cerevisiae Rrm3p DNA helicase promotes genome integrity by preventing replication fork stalling: viability of rrm3 cells requires the intra-S-phase checkpoint and fork restart activities. *Molecular and cellular biology*, *24*(8), 3198-3212.

Tran, P. L. T., Pohl, T. J., Chen, C. F., Chan, A., Pott, S., & Zakian, V. A. (2017). PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. *Nature communications*, *8*, 15025.

Vaze, M. B., Pellicioli, A., Lee, S. E., Ira, G., Liberi, G., Arbel-Eden, A., ... & Haber, J. E. (2002). Recovery from checkpoint-mediated arrest after repair of a double-strand break requires Srs2 helicase. *Molecular cell*, *10*(2), 373-385.

Dillon, L.W, Burrow, A.A., & Wang, Y. H. (2010). DNA instability at chromosomal fragile sites in cancer. *Current genomics*, *11*(5), 326-337.

Xue, X., Choi, K., Bonner, J., Chiba, T., Kwon, Y., Xu, Y., ... & Sung, P. (2014). Restriction of replication fork regression activities by a conserved SMC complex. *Molecular cell*, *56*(3), 436-445.

Xue, X., Papusha, A., Choi, K., Bonner, J. N., Kumar, S., Niu, H., ... & Lichten, M. (2016). Differential regulation of the anti-crossover and replication fork regression activities of Mph1 by Mte1. *Genes & development*, *30*(6), 687-699.

Yang, L., Liu, B., Huang, B., Deng, J., Li, H., Yu, B., ... & Yang, X. (2013). A functional copy number variation in the WWOX gene is associated with lung cancer risk in Chinese. *Human molecular genetics*, *22*(9), 1886-1894.

Yu, K., Fan, J., Ding, X., Li, C., Wang, J., Xiang, Y., & Wang, Q. S. (2014). Association study of a functional copy number variation in the WWOX gene with risk of gliomas among Chinese people. *International journal of cancer*, *135*(7), 1687-1691.

Zhang, H., & Freudenreich, C. H. (2007). An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in S. cerevisiae. *Molecular cell*, *27*(3), 367-379.

Zheng, X. F., Prakash, R., Saro, D., Longerich, S., Niu, H., & Sung, P. (2011). Processing of DNA structures via DNA unwinding and branch migration by the S. cerevisiae Mph1 protein. *DNA repair*, *10*(10), 1034-1043.

Zhu, Z., Chung, W. H., Shim, E. Y., Lee, S. E., & Ira, G. (2008). Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell*, *134*(6), 981-994.

APPENDIX

Strain Number	Strain Name	Background, Genotype	Parent Strain	Knockout Fragment Primers	Knockout Fragment Template Strain	ORF Absence Check Primers	Junction Check Primers
4106, 4119	F1-S5'(AT)34S3' o1 *rrm3::KANMX*	yPH499 MATa, ura3-52, trp1-Δ63, his3-Δ200, leu2-Δ1; ade2Δ::hisG (salmonella), LYS2::ADE2::URA3-F1, rrm3::KANMX	2525	424 & 425	1913	2272 & 2273	2271 & 15 2302 & 16
4590, 4591	No F1 control *rrm3::KANMX*	yPH499, MATa, ura3-52, trp1-Δ63, his3-Δ200, leu2-Δ1; ade2Δ::hisG (salmonella), LYS2::ADE2::URA3-no_repeat_control, rrm3::KANMX	2863	424 & 425	4106	2272 & 2273	2271 & 15
4558, 4559, 4588, 4589	F1-S5'(AT)34S3' o1 *srs2::TRP1*	BY4705, MATa ade2Δ::hisG his3Δ200 leu2Δ0 lys2Δ0 Met15Δ0 trp1Δ63 ura3Δ0 YAC:Ura+ Leu+His+ srs2::TRP	3457	499 & 500	3769	1866 & 853	687 & 161
4521, 4537, 4592	F1-S5'(AT)34S3' o1 *sgs1::KANMX*	BY4705, MATa ade2Δ::hisG his3Δ200 leu2Δ0 lys2Δ0 Met15Δ0 trp1Δ63 ura3Δ0 YAC:Ura+ Leu+His+ sgs1::KANMX	3457	497 & 498	3836, 4521	2447 & 2448	677 & 15
4538, 4539	F1-S5'(AT)34S3' o1 *mph1::KANMX*	MATa ade2Δ::hisG his3Δ200 leu2Δ0 lys2Δ0 Met15Δ0 trp1Δ63 ura3Δ0 YAC:Ura+ Leu+His+ mph1::KANMX	3457	683 & 684	3365	1760 & 1674	1675 & 16

Table 1. Yeast strains created for this project.