Assessing the Effects of Cis and Trans Factors on Fragility at the Common Fragile Site FRA16D

Alice Haouzi
Senior Honors Thesis 2015
Advisor: Catherine Freudenreich
Mentor: Simran Kaushal

Acknowledgements

I would like to extend a warm and sincere thank you to Dr. Catherine Freudenreich for giving me the chance to work on this fascinating project for the past two years, as well as Dr. Joshua Kritzer for his guidance both as my thesis committee member and as my academic advisor.

I owe Simran Kaushal a great deal of appreciation for being a patient and caring mentor and for guiding me from through this project from the very start. This project has also been made all the better thanks to the Freudenreich lab members (past and present), including Melissa Kosh, Erica Polleys, Jennifer Nguyen, Allex Xianfeng Su, Elliot Philips, Stephen Walsh, Christelle Salomon, Katherine Wu, John Volpe, Sarah Reed, Michaela Gold, Austin Leck, and Michael Sigouros.

Last but not least, I give a sincere thanks to my father for first introducing me to the world of research, and to the rest of my family and friends for their unrelenting support and kindness.

Table of Contents

Abstract

Common Fragile Sites (CFSs) are regions within chromosomes prone to breakage and characterized by late replication. FRA16D, located within the tumor suppressor gene *WWOX*, is the second most highly expressed CFS in humans and has been associated with a variety of cancer cell lines. It is our hypothesis that multiple sequences within FRA16D work in a concerted effort to cause expression of the CFS. The first goal of this project was to assess these cis factors through a subtractive approach, using a YAC breakage assay designed to correlate the number of FOA^R colonies to the rate of breakage. Deleting F1 and F5, two sequences containing AT repeats, caused a significant decrease in breakage, while deleting P5P5b, a region expected to form two large cruciforms (one of which is AT-rich), did not yield a change in fragility. In contrast, an additive approach was taken to measure the fragility of the F5 region via a direct duplication recombination assay (DDRA). The introduction of F5 with 24 interrupted AT repeats did not yield an increase in breakage, while the introduction of F1 with 23 perfect AT repeats had previously been shown to cause fragility. Taking findings from both approaches suggests that F5 might only cause breakage in the context of FRA16D, acting as a kind of "enabler" in the region. Trans factors affecting breakage were also tested via the DDRA in yeast strains containing a specific cassette with the 34 (AT) repeats of the F1 sequence. The deletion of *MUS81* showed a very significant decrease in the rate of FOA^R compared to the wild type, a result congruent with Mus81's role as an endonuclease that causes breaks at CFSs. The deletion of *MEC1*, an ATR homolog, yielded an increase in FOA^R that was milder than expected, based on the well-documented effect of ATR kinases in preventing fragility. This result was explained by finding that *MEC1* was not actually deleted and was perhaps being preserved as an extrachromosomal element.

Introduction

The maintenance of genomic stability is vital for proper cellular function. In order for cells to divide, they must be able to replicate their genome correctly and completely. Nonetheless, replication is difficult, since our chromosomes are under constant exogenous and endogenous stresses. For instance, replication machinery must unwind the DNA strands in order to copy them, creating torsional strain ahead of the replication fork and displacing DNA-bound proteins such as histones, all in a tightly regulated manner. Our chromosomes are condensed tremendously in prophase and are then pulled apart to either side of the cell at the end of mitosis. In these conditions, genomic instability can occur and may result in breakage of the chromosome, which in turn may lead to deleterious rearrangements such as translocations, mutations, or even cell death (Freudenreich, 2007). Research in the past few decades has shed some light on the possible mechanisms that contribute to this chromosome fragility.

Common Fragile Sites

Common fragile sites (CFSs) are regions within human chromosomes that are prone to breakage. Under normal cellular conditions these sites are usually stable, but during conditions of replication impairment, called replication stress, large gaps can form on metaphase chromosomes (Figure 1). Replication stress can be induced by natural processes and in response to certain drugs, such as low doses of aphidicolin (a specific inhibitor of eukaryotic DNA polymerase α).

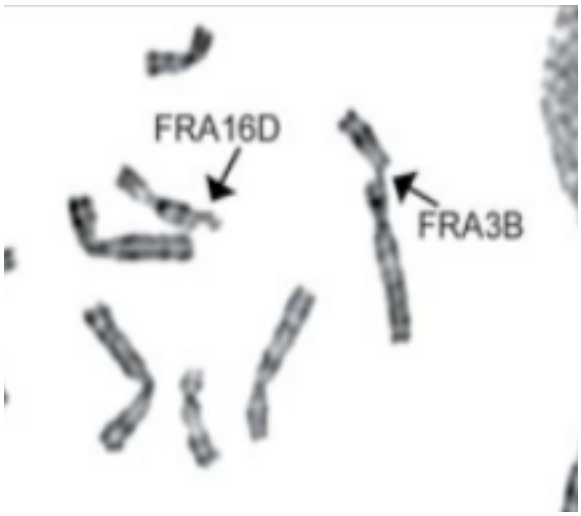


Figure 1. Breaks at common fragile sites 1 FRA3B and FRA16D in metaphase chromosomes (Glover et. al., 2005).

These breaks have been observed to occur over large regions of the genome, up to a megabase or more (Arlt et. al., 2006). Replication stress can be induced by compounds such as caffeine and ethanol and can also be induced by drugs that affect DNA replication, such as aphidicolin (Richards et al., 2008). Over 80 CFSs have been identified thus far, and recent studies have shown that CFS breakage occurs at different frequencies (different "expression") for different sites. Furthermore, CFS expression varies in different cell types. (Le Tallec et. al., 2013; Helmrich, 2011). Common fragile sites regions are of particular importance since they are normal components of chromosome structure and are common to all individuals (Freudenreich, 2007). In fact, they are highly conserved across species, suggesting their existence may serve an important yet still unknown cellular function (Kerem et. al., 2012).

Cancer and Common Fragile Sites

Replication stress occurs during natural circumstances, such as cancer. Cancer cells rapidly divide, therefore they must quickly replicate their genomes, which causes replication stress in the cells. Due to misregulated replication resultant from mutations, cancer cells are prone to a great deal of DNA damage, rearrangement, deletions, and amplifications due to deleterious modifications to the replication and metabolic machineries (Negrini et. al., 2010). In fact, cancer cell DNA is preferentially altered at

CFSs over the rest of the genome. Some CFSs undergo breakage-fusion-bridge cycles, which can alter gene expression (Huebner and Croce 2001; Coquelle et. al., 1997). Very large deletions of up to hundreds of kilobases may occur, which can be deleterious for the cell and can result in tumorigenesis if important regulatory genes are disrupted by CFS breakage (Arlt et. al., 2006; Bignell, 2010). Some of the most highly expressed CFSs are located within tumor suppressor genes, so it is possible that breakage at CFSs is a stepping-stone in the progression of cancer (Georgakilas et. al., 2014). Understanding the mechanism of fragility at CFSs is an important stride to understand cancer progression.

The two most commonly expressed CFSs are FRA3B and FRA16D, which both lie within large tumor suppressor genes. FRA3B is located within the *FHIT* gene and homozygous deletions within this region have been linked to lung, kidney, stomach, and cervical carcinomas (Huebner et. al., 1998). Restoration of the *FHIT* gene, through overexpression or via introduction of a functional allele, could in fact reduce susceptibility to cancer development (Arlt et. al., 2006). Likewise, deletions within FRA16D, a CFS found in the tumor suppressor *WWOX*, have been detected in cell lines derived from carcinomas of the colon, breast, lung, stomach, and ovary (O'Keefe & Richards, 2006). It seems that breakage events at CFSs can in fact play an important role in cancer development, at least in the common cases of FRA3B and FRA16D.

However, not all common fragile sites are associated with tumor suppressor genes. Some fragile sites are not located within any genes at all; FRAXB also shows frequent deletions in tumor cells, but none of the genes associated with it seem to contribute to tumor progression (Arlt et. al., 2002). It is very possible that those CFSs not

associated with particular genes might actually encode important regulatory RNAs, many of which are yet to be identified within our genome.

Le Tallec. et. al. put forth interesting evidence to link cancer and CFSs; in a study mapping CFSs in epithelial and erythroid cells, an extensive overlap was found between CFSs and cancer deletions in large genes over 300kb long. It was also shown that late replication, a documented feature of CFSs and one that is enriched in cancer deletions, is characteristic of large genes. These results strongly suggest that recurrent cancer deletions overlapping large genes originate from CFSs, which could in fact explain over half of all the recurrent deletions seen in tumors (Le Tallec et. al., 2013).

Mechanism of Breakage at Common Fragile Sites

Current theories to explain the source of breakage at CFSs incorporate a number of factors, from intrinsic characteristics of the regions themselves to direct interferences with the replication process, although the specific features that render regions susceptible to breakage remain under investigation.

In an effort to understand the mechanism of CFS expression, a sensible first step was to look into the sequences of CFSs themselves; computational analysis performed on a subset of fragile sites showed that these regions were AT-rich (Zlotorynski et. al., 2003). Further studies indicated that these AT-rich sequences had high "flexibility" and low stability, and it was first hypothesized that these flexible sites might result in the breakage seen at CFSs. However, a number of studies were done since this hypothesis was put forth to show that flexibility did not always correlate to CFS expression; for instance, *Helmrich et al*. actually disputed the notion that CFSs are regions with abnormally high flexibility peaks (*Helmrich et al*., 2007).

An alternative model has now gained grounds, suggesting that breakage is favored by the presence of AT repeats that may promote nucleation of DNA secondary structures, which are noncanonical DNA structures (Zhang & Freudenreich, 2007). Indeed, structure-forming sequences in DNA have been shown to stall replication fork progression in bacterial and yeast cells (Mirkin, 2006), which can result in genomic instability like chromosomal breaks and rearrangements (Voineague et al., 2009). Formation of secondary structures at these AT-rich regions could thereby account for the characteristic late-replication seen at CFSs such as FRA16D (Palakodeti et al., 2004).

The current model that supports this evidence suggests that replication is initiated normally, but the fork stalls within CFSs when it reaches a region difficult to replicate due to the presence of secondary structures, leading to partially unreplicated DNA and ultimately resulting in the double stranded breaks (DSBs) that are the source of chromosomal rearrangements (Zhang & Freudenreich, 2007). How exactly these secondary structures might interact with the replication fork remains unknown; they might act as direct replication barriers or perhaps induce action of nucleases or other proteins that provoke these DSBs. They may also impair transcription and stall RNAPII, which could also impair replication. Genome-wide analyses have not confirmed that CFSs are significantly enriched in sequences prone to form secondary structures as compared with non-fragile regions of similar base composition (Le Tallec et. al., 2014). In addition to the theory of secondary structure formation at CFSs, novel findings have put forth the idea that replication origin density might determine fragility in CFSs. One study mapped origins of replication in FRA3B versus that of non-fragile regions, showing that while still active, the origins at FRA3B seemed to be less efficient, and that

treatment with aphidicolin slowed replication to a greater extent and led to genetic instability that was characteristic of CFSs (Palakodeti et al., 2010). Another study led by the Debatisse group suggested that the fragility of FRA3B in lymphoblastoid cells but not fibroblasts was due to a scarcity of origins, thus forcing replication forks to travel much longer distances to finish replication (Debatisse et. al., 2012). The Debatisse group's observations are especially pertinent in supporting the idea that CFS expression is epigenetically defined; the availability of replication origins will vary between cell types, which is hypothesized to vary expression of the same CFS in different cell types (Le Tallec et. al., 2013). Origin density is thereby an interesting alternative to previous models; CFS expression might be the result of faulty origins or too few origins. It must also be stressed that none of the models discussed thus far are mutually exclusive with one another.

Proteins involved in Regulating Fragility: ATR kinases and endonucleases

A number of cellular pathways are responsible for regulating genomic stability and ensuring that damage is dealt with prior to entry into mitosis. One pathway especially pertinent to CFSs is the ATR-dependent checkpoint, which in humans is activated in the presence of fork stalling caused by conditions such as exposure to ultraviolet light, hydroxyurea, aphidicolin, and hypoxia (Glover et. al., 2005). These conditions block the firing of new replication origins, prevent entry into mitosis, and promote DNA repair (Franchitto, 2013). A clear connection between CFSs and ATR proteins was established by the Glover group, which showed that disturbing ATR induces CFS expression even in the absence of aphidicolin treatment, thereby suggesting that ATR is necessary for stability of DNA under replication stress but also during normal cell division (Casper et.

al., 2002). Since this discovery, a number of proteins that interact with the ATR pathways have also been shown to influence CFS expression, including BRCA1, the Fanconi anemia (FA) pathway proteins, and SMC1, thereby making this an active area of study to further understand mammalian checkpoints and repair pathways (Glover et. al., 2005).

The second protein of interest to our project is Mus81, a DNA structure-specific 3' endonuclease. Mus81 forms a complex with Eme1 that localizes to CFS loci in early mitotic cells in order to create gaps in metaphase chromosomes. The Mus81-Eme1 complex is only activated via phosphorylation in late G2/early M (Gallo-Fernandez et. al., 2012), presumably remaining inactive in S phase in order to prevent inadvertent cleavage of replication forks. The Hickson group proposed a model to account for the role of Mus81-Eme1 in preserving genome integrity at CFSs, whereby the cleavage of CFSs by the Mus81-Eme1 complex actually promotes faithful sister chromatid disjunction in cases of incomplete replication (Ying et. al., 2013) (Figure 2). Naim et. al. 2013 further confirmed that depletion of the Mus81-Eme1 complex affected accurate processing of replication intermediates or under-replicated DNA that persists at CFSs until mitosis. Similarly to the Hickson group, they observed that depletion of the endonuclease also led to an increase in the frequency of chromosome bridges during anaphase that, in turn, favored accumulation of DNA damage in the following G1 phase.

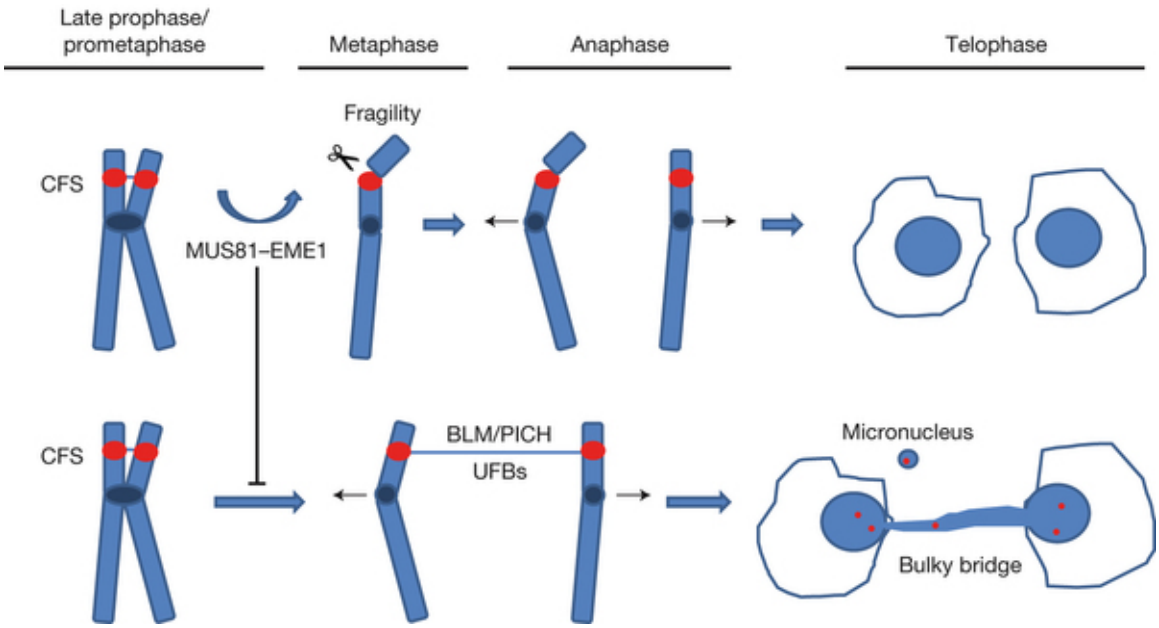


Figure 2. Schematic from Ying et. al. 2013. Model for the role of Mus81–Eme1 in preserving genome integrity at CFSs. In this diagram, sister chromatids are shown in light blue and centromere as dark blue ovals. The red ovals denote the location of a CFS. In the upper panel, the MUS81-EME1 complex is recruited to a site of incomplete replication and cleaves the CFS. This allows sister chromatids to separate and for the DNA damage to be repaired in the next round of replication. If the MUS81-EME1 complex is deficient (lower panel), however, no cleavage occurs at the difficult-to-replicate CFS. This leaves intertwined DNA regions between the unreplicated and replicated DNA strands, thereby preventing proper sister chromatid separation and creating bulky anaphase bridges or ultra-fine anaphase bridges. While some bridges can be resolved, others may trigger chromosome mis-segregation in the form of an uneven distribution of DNA between the daughter cells and/or micronuclei. As shown by the red foci, these structures will have expressed CFSs. (Ying et. al., 2013)

The Common Fragile Site FRA16D

The common fragile site FRA16D is the focus of this project. FRA16D is a 270-350kb region of DNA that lies between the eighth and ninth exons of the tumor suppressor gene *WWOX* on chromosome 16q23.2. As previously mentioned, it is one of the most highly expressed CFSs in humans. It is fragile in a number of different cell types (Tallec et. al., 2013) and has been linked to a number of carcinomas. In a 2007 study by Zhang and Freudenreich, an assay was developed to assess the fragility of a yeast artificial chromosome containing the FRA16D region and surrounding sequence from the *WWOX* gene as compared to a YAC lacking FRA16D. Results indicated a significant

difference in breakage frequency, with YAC801B6 showing a breakage frequency of ~18% compared to that of YAC972D3 at ~4% (Zhang & Freudenreich, 2007) (Figure 3). This assay allows for a quantitative measurement of breakage by correlating it to a rate of FOA resistance. The YACs used in their experiments contained the URA3 marker, which encodes for Orotidine 5'-phosphate decarboxylase (ODCase), an enzyme that catalyzes one reaction in the synthesis of pyrimidine ribonucleotides (a component of RNA). If the drug 5-FOA (5-Fluoroorotic acid) is added to the media, the active ODCase will convert 5-FOA into the toxic compound 5-fluorouracil, causing cell death. Growing cells in media containing 5-FOA thereby allows for selection against yeast carrying the URA3 gene. If breaks occur within the FRA16D region, then a portion of the intermediates will be resected by exonucleases to expose the telomere seed sequence, allowing for rescuing of the broken YAC via *de novo* telomere addition. It is in this process that the YAC loses the URA3 gene, thereby making the cells resistant to 5-FOA, allowing a correlation of the rate of FOA^R to the rate of breakage. Resection to the telomere seed sequence is only one of the outcomes of breakage at the FRA16D region; as shown in Figure 4, breakage is most likely to result in deletion of a region around the lesion via nonhomologous end joining, which might preserve sensitivity to 5-FOA if the URA3 gene is undisturbed, thereby leading to an understatement of the actual rate of breakage events. A third potential outcome is the formation of breakage intermediates for which resection does not go all the way to the G4T4 sequence, which is ~890kb from the FRA16D site (Figure 4).

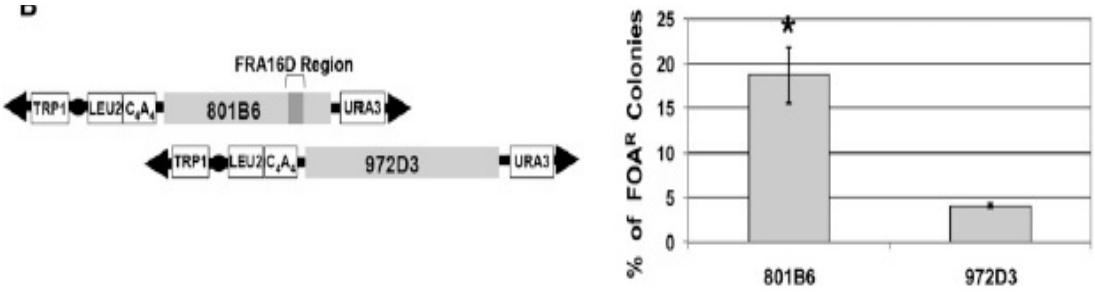


Figure 3. Schematic from the Zhang and Freudenreich 2007 study, illustrating YACs from the CEPH YAC library that were modified by adding a telomere seed sequence and a LEU2 marker. Human sequences are represented by grey boxes (not to scale). The dark grey box represents the 270 kb FRA16D region defined as most fragile by (Reid et al., 2000). The breakage assay was performed for YAC 801B6 and YAC 972D. Results showed a statistically significant increase in frequency of FOA resistance for the YAC containing FRA16D (* $p < 0.05$).

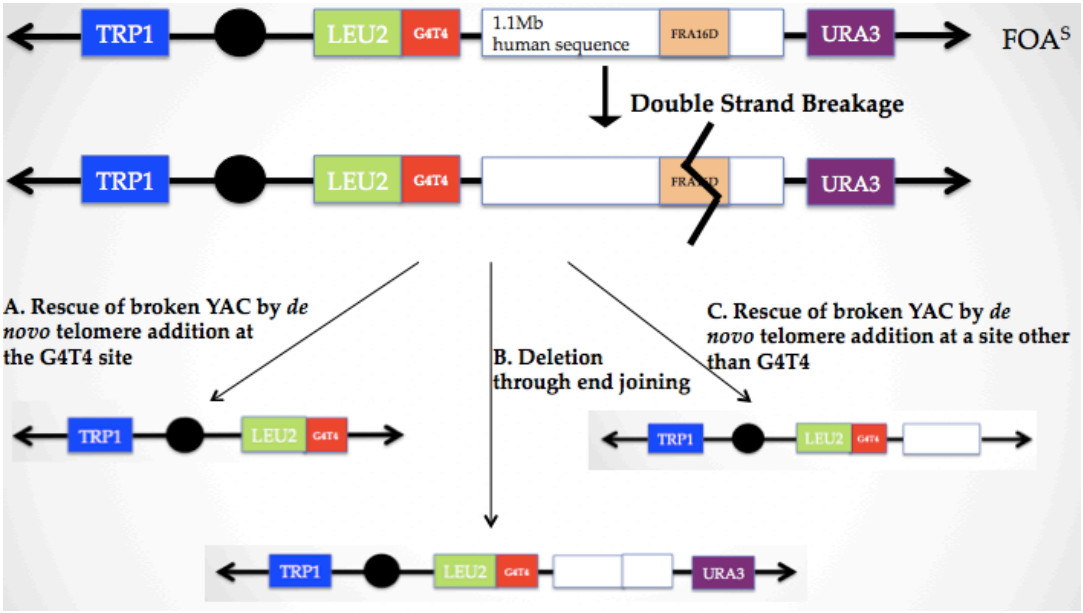


Figure 4. Schematic diagram of the YAC Breakage Assay.

Next, Zhang sought to understand which regions specifically might be responsible for the fragility within FRA16D. F1 is a ~500bp AT-rich region within FRA16D that contains a polymorphic AT repeat that varies from 11-88 perfect ATs in humans (Finnis et al, 2005). The AT repeats are expected to form a stable cruciform structure that contains sequences flanking a frequently deleted region in multiple tumor cell lines. Zhang and Freudenreich demonstrated in 2007 that F1 did in fact increase chromosome fragility as

compared to a control with 386bp of DNA from FRA16D not predicted to form a secondary structure. This effect was even more pronounced in the absence of the double stranded break repair protein, Rad52, as well as in the presence of hydroxyurea, an inhibitor of DNA synthesis. F1's fragility was measured using a small YAC that was specifically designed with a G4T4 telomere seed sequence and a URA3 marker gene as shown in Figure 5. Because the number of AT repeats at F1 actually varies within the human population, the fragility of F1 with different numbers of AT repeats was also tested. Results from the Zhang and Freudenreich study showed an upward trend of chromosomal breakage with increasing (AT) dinucleotide length. 2D gels also showed increased fork stalling with greater AT length (Zhang & Freudenreich, 2007).

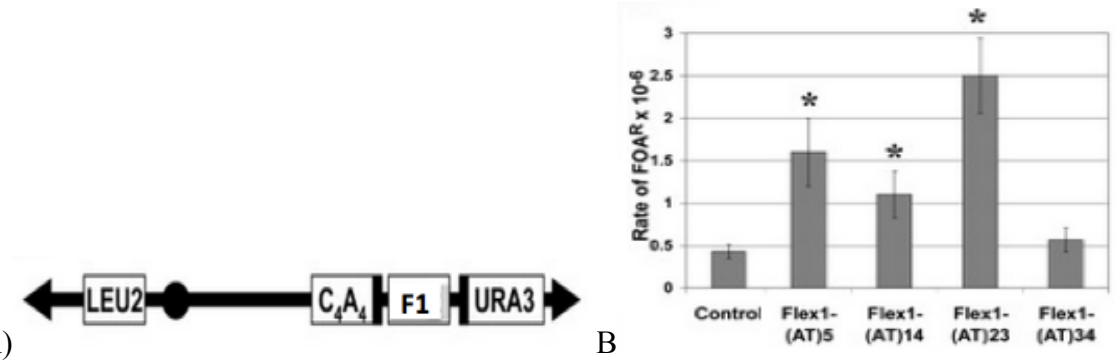


Figure 5. Schematic and results from the Zhang and Freudenreich 2007 study, showing that the F1 sequence increases chromosome breakage. A) Structure of the small YAC containing the F1 subregion of FRA16D with varying number of AT repeats (labeled "F1"). B) Results of the YAC breakage assay in the rad52Δ background. The control is 1320 bp of FRA16D sequence not predicted to form secondary structures. $*p < 0.05$ compared to the control by a pooled variant t test.

Results obtained from the large YAC801B6 assay and the small F1 YAC assay were therefore the foundation of this project, which looked to the cis and trans factors that affect fragility of the common fragile site FRA16D. Cis factors refer to the DNA sequences themselves within FRA16D; indeed it is our hypothesis that multiple regions within FRA16D, in addition to F1, work in a concerted manner to induce fragility at the fragile site. An additive approach was taken to assess the individual fragility of the

sequences, while a subtractive approach was done to assess fragility of the FRA16D region in the absence of those sequences. Trans factors refer to the proteins Mec1 and Mus81, both of which are involved in different pathways thought to regulate CFS expression.

Methods

Cis Factors: Additive Approach

Yeast and Bacterial Strains

Freudenreich lab bacterial stock #223 was used for plasmid pBL007 vector and #457 was used to obtain the F5 sequence. The F5 cloning constructs were transformed into yeast strain #2268.

Preparation of F5 fragment for cloning

The bacterial plasmid #457 (provided by the Eckert group) was used as it contained the F5 fragment with BamHI cut sites flanking both sides of the sequence (See Appendix). The plasmid stock was streaked out for singled colonies on an LB+Amp plate at 37°C overnight. A colony was picked out and grown in 3mL of LB broth +0.15mg ampicillin overnight at 30°C. From this culture, the plasmid DNA was isolated using a Zyppy^{TM} plasmid miniprep kit (Zymo Research) according to the manufacturer's instructions. The plasmid DNA was then digested with the BamHI restriction enzyme for 3 hours at 37°C in a 50 µL total volume (5µL NEB Buffer4, 5µL 10xBSA, 1µL NEB BamHI, 1.095 µg DNA). The digest was heat inactivated at 65°C for 20 minutes. The product was run on a 1.5% agarose gel (50µL of the digest with 10µL of 6X loading dye) at 100V for 45 minutes. Expected size of the product was 172bp; the band at this size was cut out and gel purified using the Axygen Gel Clean-Up Kit according to the manufacturer's instructions. Final concentration was recorded via Nanodrop reading.

Preparation of the pBL007 vector for cloning

The plasmid stock #223 (pBL007, see Figure 4) was streaked out for singled colonies on an LB+Amp plate at 37°C overnight. A colony was picked out and grown in 3mL of LB

broth +0.15mg ampicillin overnight at 37°C. From this culture, the plasmid DNA was isolated using a Zyppy™ plasmid miniprep kit (Zymo Research) according to the manufacturer's instructions. The plasmid was then digested with BamHI-HF in a 60uL volume (6µL CutSmart Buffer, 2µL BamHI-HF, and 22.79mg pBL007) overnight. The next day, the mixture was incubated at 65°C for 20 minutes to inactivate the enzyme. The digested vector was dephosphorylated for 15 minutes at 37°C in a 70µL volume (all 60µL of the digest, 7µL 10X Antarctic Buffer, 2µL Antarctic phosphatase, 1µL distilled water). The Antarctic phosphatase was heat inactivated for 5 minutes at 70°C. The dephosphorylated plasmid was then run on a 1.5% gel at 100V for one hour (60µL plasmid DNA with 12µL 6x loading dye). Expected size of the product was 5344bp; the band at this size was cut out and gel purified using Axygen Gel Clean-Up Kit according to the manufacturer's instructions. Final concentration was recorded via Nanodrop reading.

Cloning of F5 insert into pBL007

Ligations of the inserts with the pBL007 vector were done in a 35uL volume at 16°C overnight with a 1:3 vector to insert ratio (53.12ng of F5 insert, 550ng pBL007, 3.5µL 10x Buffer T4, and 1uL T4 ligase). The following day, the mixtures were incubated at 65°C for 20 minutes to inactivate the ligase. Prior to performing these ligations, two control ligations were done, one with the vector only (substituted F5 DNA with water) and the other with the vector-only and no ligase (substituted both F5 DNA and ligase volumes with water). The ligated products were then transformed into JM109 heat shock competent *E. coli* cells. The bacterial transformation protocol for heat shock competent cells was the following: 10uL of the ligated products were added to the thawed

competent cells and set on ice for 30 minutes. Cells were heat shocked for 45 seconds in a 42°C water bath and transferred back on ice for 2 minutes. 0.8 mL of LB broth was added per tube and the mixtures were incubated at 200rpm and 37°C under shaking for 40 minutes. The cultures were spun down at 10000 rpm for one minute, the supernatant was discarded and the pellet was resuspended in 100µL of distilled water. 100µL of the undiluted mixture was plated on LB+Amp, and 100µL of a 1:10 dilution was plated on a second LB+Amp plate. Both were grown at 37°C for 24 hours.

Checking for proper insertion of F5 into pBL007

Colony PCR was performed to check for successful clones using the Fasttaq program and primers that amplified the multiple cloning site of pBL007 (48 reactions were made with 458.4µL, 60µL Sib Buffer, 30µL 1:10 primer #679, 30µL 1:10 primer #680, 12µL dNTPs, 9.6µL Sib Taq enzyme, DNA picked from Transformant colony). The PCR products were run out on 1.5% gel at 100V for 45 minutes. Colonies that showed bands at the expected 1172bp size were selected and their DNA was isolated using a Zyppy^{TM} plasmid miniprep kit (Zymo Research) according to the manufacturer's instructions. The plasmid DNA was sent to EtonBio for sequencing with primers #1032/1033, which were located on either side of the MCS. Since the cloning was non-directional, verified successful transformants in both orientations were saved as glycerol stocks #520/523 (orientation 1) and #521/522 (orientation 2).

Transformation of F5 construct into yeast

The F5 vectors (strains #520/521) were linearized via digestion with the XBaI restriction enzyme at 37°C overnight in a 20µL total volume (2µL CutSmart, 1µL XBaI, 2µg plasmid DNA). Digests were heat inactivated at 65°C for 20 minutes. The

transformation of the linearized product into yeast strain #2268 was performed using the Freudenreich lab's modified Lithium Acetate Transformation Protocol. Cells from were inoculated in 2mL of YC-Ura to select for overnight at 30°C. The next day, 10mL of YC-Ura mixed with 400µL of the overnight culture was inoculated to a 0.2 OD, which was then incubated at 220rpm at 30°C for 4-5 hours to 1.0 OD. The culture was centrifuged for 5 minutes at 3000rpm. The pellet was resuspended in 1 mL H2O and transferred to an eppendorf tube and centrifuged for 30 seconds at 10000rpm. The pellet was resuspended in 900 uL of freshly prepared solution A (110 uL LiAc 10x, 110 uL TE 10x, 880 uL H2O), centrifuged for 30 seconds at 10000rpm, and the pellet was resuspended again in 100uL of solution A. 100 uL of the cells just prepared, 10 uL of single stranded sperm DNA denatured at 100°C for 10 minutes, 20 uL of the appropriate PCR product, and 600 uL of solution B (480 uL PEG 50%, 60 uL LiAc 10x, 60uL TE 10x) were mixed and incubated for 30 minutes at 200rpm and 30°C under shaking. The mixture was heat shocked at 42°C for 15 minutes, centrifuged at 10000rpm for 30 seconds and the supernatant was poured off. The pellet was resuspended in 3mL YEPD and recovered at 30°C for 4 hours (allowing cells to express marker needed for selection). The mixture was centrifuged for 5 minutes at 3000rpm, the supernatant was poured off and resuspended in 100 uL H2O, and plated on YC-Ura plates at 30°C for 3 days to verify their +URA selection.

Checking for successful insertion of pBL007+F5 into yeast ch II LYS2 locus, to create an F5 DDRA construct

Three different PCRs were performed to check for insertion of the F5 construct into the right locus. The first two PCRs were performed to check for the 5' and 3' ends of

the F5 construct by using the Fasttaq program. One primer annealed to the internal F5 construct and the other annealed outside of it on the yeast chromosome. Each PCR reaction had a total volume of 12.5µL (9.55µL distilled water, 1.25µL Sib Buffer, 0.625µL primer#1028 (for 5'end check) or #1030 (for 3'end check), 0.625µL primer#1029 (for 5'end check) or #1047 (for 3'end check), 0.25µL dNTPs, 0.2µL Sib enzyme, and 1µL yeast gDNA). The third PCR was done to amplify the AT repeats of F5 using the special PCR program (ASK SIM FOR SPECIFIC PROTOCOL) with a total volume of 200µL for 16 reactions (152.4µL distilled water, 60.0µL Q5 Buffer, 30.0µL primer#1032, 30.0µL primer#1033, 6.0µL dNTPs, 4.8µL Q5 enzyme, and 2µL yeast gDNA). The PCR products were purified via ethanol precipitation by incubating 43µL of the PCR product with 1/10 volume of 3M KAc and 2 volumes 100% ethanol overnight at -20°C. The next day the product was spun down at 13500rpm at 4°C for 30 minutes, 100µL 70% ethanol was added to the pellet and spun down for 5 minutes at room temperature. The pellet was allowed to dry for 10 minutes and was resuspended in 15µL 1xTE. The purified PCR product was sent for sequencing at EtonBio with primers #1032/1033. Verified successful transformants were saved as glycerol stocks #3526/3527 (F5 orientation 2) and #3528/3529 (F5 orientation 1).

Direct Duplication Recombination Assays

Cells were taken from the patch of the strain grown on the YC-Ura plate and were mixed with 1000µL water. 50µL of a 10^{-4} dilution were plated on a YPD plate at 30°C for 3 days. A single colony was selected and mixed in 400µL water (done ten separate times for a 10-colony assay; all colonies taken have an O.D. between 3 and 5). From the 400µL colony suspension tube, a total of 10µL of cells was plated in a 100µL volume on FOA-

Ade plates for strains with low recombination rates, while 2μL of cells was plated in a 100 μL volume for strains with high recombination rates. For the total cell count, 100μL was taken from each of the ten 400μL colony suspensions and mixed together. 100μL of a 10^{-5} dilution was then plated on 2 YPD plates. The colonies were grown at 30°C for 5 days. Rate of FOA resistance was determined using the Method of the Median in the FALCOR program with a mutation frequency of 10^{5}.

Cis Factors: Subtractive Approach

Yeast Strain

YAC 801B6 (strain #1086) in the AB103 strain background was used to make all the deletions.

Amplifying Marker with 40bp homology to regions flanking sequence of interest

The plasmids used were isolated from bacteria using a Zyppy™ plasmid miniprep kit (Zymo Research) according to the manufacturer's instructions. The marker was then amplified from the plasmid with 60bp primers via PCR using Sib taq enzyme and buffer. The forward primer was designed with a 40bp homology to a region upstream of the 5'end of the region of interest, and the reverse primer had a 40bp homology downstream of the 3'end of the region of interest (See Table 1). To check for successful amplification, samples of the PCR products were run through a 1.5% gel at 100V for one hour.

Table 1. Regions deleted within FRA16D of YAC801B6.

Region to be deleted	Marker gene	Primers to amplify marker	Marker template (plasmid stock number)	Amount of DNA deleted	Primers to check successful deletion in YAC801B6	Expected product size with checking primers	Stock number of saved yeast strains
F1	KanMX6	#1236/1237	pFA-KanMX6 (#136)	1560bp	#16/1267	537bp	3076/3077
F5	His3MX6	#1448/1449	pFA-His3MX6 (#138)	1404bp	#375/1545	959bp	3513/3514/3587/3588
P1	KanMX6	#1573/1574	pFA-KanMX6 (#136)	1603bp	#16/1704	688bp	3589/3590
P2P2b	KanMX6	#1571/1572	pFA-KanMX6 (#136)	1726bp	#16/1703	648bp	3591/3592
P5P5b	Hyg	#1555/1556	pAG32 (#241)	1842bp	#396/1481	1157bp	3517/3549

Yeast Transformation of marker gene in YAC 801B6

The marker gene flanked by homology to the sequence to be deleted was transformed into yeast strain #1086 following the Freudenreich Transformation Protocol descrived above in the Additive Approach section. Changes made to the protocol included the firs step, in which cells were inoculated in 2mL of YC-Leu-Ura to select for the YAC overnight at 30°C, and the final step, in which transformants were plated on 2 selective media plates (YPD+G418 for F1/P1/P2P2b, YC-His for F5, and YPD+Hyg for P5P5b) at 30°C for 3 days. Resultant transformant colonies were patched out on their respective media plates at 30°C to confirm the selection.

Verification of successful integration of marker gene

To check for successful transformation and integration of the marker gene, primers were designed such that one annealed internally to the marker gene and the other annealed outside of the region to be deleted on the YAC. Colony PCR was performed with the Fasttaq program by selecting individual colonies from the transformation plates. Each PCR reaction had a total volume of 12.5µL (9.55µL distilled water, 1.25µL Sib Buffer, 0.625µL 1:10 forward primer, 0.625µL 1:10 reverse primer, 0.25µL dNTPs, 0.2µL Sib enzyme, and DNA from the transformation patch). The PCR products were run on a 1.5% gel at 100V for one hour to check for the expected amplicon sizes. Confirmed transformants were saved as glycerol stocks.

YAC Fragility Assay

Cells were patched onto YC-Ura-Leu-Trp plates for 3 days at 30°C to select for both arms of the YAC. From the patch, 10^{-4} dilutions were plated for single colonies on YC-Ura-Leu-Trp for 2 days at 30°C. Part of a colony from this plate was incubated in

5mL of YC-Leu (this was the breakage culture) and grown at 30°C for 6 to 7 divisions (~16 hours) with a starting OD of 0.2-0.4. From the breakage culture, 100µL of 10^{-4} dilutions were plated on FOA-Leu to select for breakage events in which the URA3 gene was lost, and 100µL of the 10 combined colonies at a 10^{-4} dilution were plated on YC-Leu for the total cell count. Both sets of plates were grown at 30°C for 3 days. The number of colonies on the FOA-Leu and YC-Leu plates was counted, and average breakage frequency was calculated.

Preparation of DNA plugs and Pulse Field Gel Electrophoresis

The preparation of the plugs was adapted from a protocol provided by Kirill Lobachev. Cells were taken from a patch and inoculated in 20mL YC-Leu-Ura overnight at 30°C. Cells were spun down at 3000rpm for 10 minutes and washed once with 20mL of distilled water. The cell pellet was resuspended in 600µL of distilled water. Number of cells was counted via hemocytometry. $8x10^8$ cells and were transferred in a microcentrifuge tube, spun down at 10000rpm for 1 minute, and washed with 200µL of plug solution (0.5M EDTA, 10mM Tris, pH7.0). Amount of desired plug solution was added to the pellet, mixed gently with a pipette, and 5mg/mL lyticase was added to the required volume to have 8.24% lyticase final concentration. Cells were placed in a 45°C water bath for 2 minutes, and 2% CleanCut Agarose was added to yield a 0.8% agarose final concentration. The solution was mixed twice using a pipette, quickly transferred to plug molds, and allowed to solidify at 4°C. Each plug was incubated in 600µL plug solution with 10µL 5mg/mL lyticase overnight at 30°C. 200µL of protease K solution (5mg/mL protease K, 5% sarkosyl, 0.5M EDTA, pH7.5) was added to each tube and incubated for 5 hours at 37°C. The plugs as well as a size ladder (BioRad *S. cerevisae*

chromosomal DNA ladder) were then loaded onto a 1.2% agarose gel (24g of BioRad Pulse Field Certified agarose in 200mL 0.5X TAE). The PFGE was run for 24 hours in 2mL 0.25X TAE at 14°C, with a switch time of 60-120seconds and a voltage of 6V.

Trans Factors: Deletion of MEC1 and MUS81

Yeast Strains

Freudenreich lab yeast stocks #657, #1791, #2863 (no repeat 386bp control) and #2525 (F1-S5'$(AT)_{34}$S3' orientation 2) were used.

Amplification of SML1::TRP1, MEC1::KANMX6, and MUS81::KANMX6

The SML1::TRP1, MEC1::KANMX6, and MUS81::KANMX6 regions were amplified via PCR with the Expand program. The primers used for the amplification were the original primers used to make the knockout of the gene with the selectable marker. Consequently these primers contained sequence homology upstream and downstream of the deleted genes. Each PCR tube contained 1.5µL forward/reverse primers (#570/571 for SML1::TRP1, #567/568 for MEC1::KANMX6, and #572/573 for MUS81::KANMX6), 5µL Expand Buffer with $MgCl_2$, 19µL water, 1µL dNTP, 1µL of 1:100 dilution of template gDNA (#657 for SML1::TRP1 and MEC1::KANMX6, #1791 for MUS81::KANMX6) and 1µL of Expand enzyme.

Transformation of amplified regions

The transformation of the PCR amplified product was done in yeast strains #2525 and #2863 for SML1::TRP1 and MUS811::KANMX6. MEC1::KANMX6 transformations were performed in the constructed *sml1* knockout strains (#3300 and 3227). Cells from the strains were inoculated in 2mL of YC-Ura to select for the chromosome II DDRA cassette overnight at 30°C. The next day, 10mL of YC-Ura mixed with 400µL of the overnight culture and the LiAc transformation protocol was followed described in the Additive Approach section. Transformants were plated on appropriate selective plates at 30°C for 3 days.

Verification of successful integration of marker gene

Colonies resultant from the transformation were patched on selective plates and successful transformants were checked via colony PCR using primers #45/140 (SML1::TRP1), #16/103 (MEC1::KANMX6), #16/574 (MUS81::KANMX6) and the Fasttaq program (for 56 reactions each 12.5µL: 573µL distilled water, 70µL Sib Buffer, 35µL primer#16, 35µL primer#574, 14µL dNTPs, and 11.2µL Sib Enzyme). A second colony PCR was performed to confirm the absence of the gene maintained as an extrachromosomal element, using the Fasttaq program and primers #1643/103 for *mec1*check and #1644/574 for *mus81* check. Each PCR reaction had a total volume of 12.5µL (9.55µL distilled water, 1.25µL Sib Buffer, 0.625µL 1:10 forward primer, 0.625µL 1:10 reverse primer, 0.25µL dNTPs, and 0.2µL Sib enzyme). Confirmed transformants were saved as glycerol stocks (See Table 2).

DDRAs were then performed as described above in the "Cis Factors: Additive Approach" section.

Table 2. Constructed Strains

Yeast strain CFY #	Strain Name
#3299/3300	sml1Δno repeat
#3227/3228	sml1ΔF1-S5'AT_{34}S3'
#3346	sml1Δmec1Δno repeat (found later to not be a true mec1Δ)
#3338/3339	sml1Δmec1Δ F1-S5'FAT_{34}S3' (found later to not be a true mec1Δ)
#3375/3375	mus81Δno repeat
#3377/3378	mus81ΔF1-S5'AT_{34}S3'

Results

Cis Factors: Additive Approach

The additive approach assesses the individual fragility of a DNA sequence isolated from the larger FRA16D context, using a different kind of assay: the Direct Duplication Recombination Assay (DDRA). For the purpose of this project, only the fragility of F5, a region with 24 interrupted AT repeats (See F5 sequence in Appendix), was assessed. The F5 sequence was isolated from plasmid #457 by BamHI-HF digestion. The 172bp expected band for the F5 fragment was very faint but present and was gel purified (Figure 6A). The vector plasmid #223 pBL007 was prepared for cloning by digesting with BamHI-HF. The BamHI-HF digested pBL007 showed a clear band at the expected size of 5308bp in comparison to the undigested vector, which showed multiple bands due to supercoiling of the plasmid (Figure 6B). Concentrations of BamHI digested F5 fragment and BamHI digested pBL007 after purifications were 2.4ng/µL and 110ng/µL, respectively.

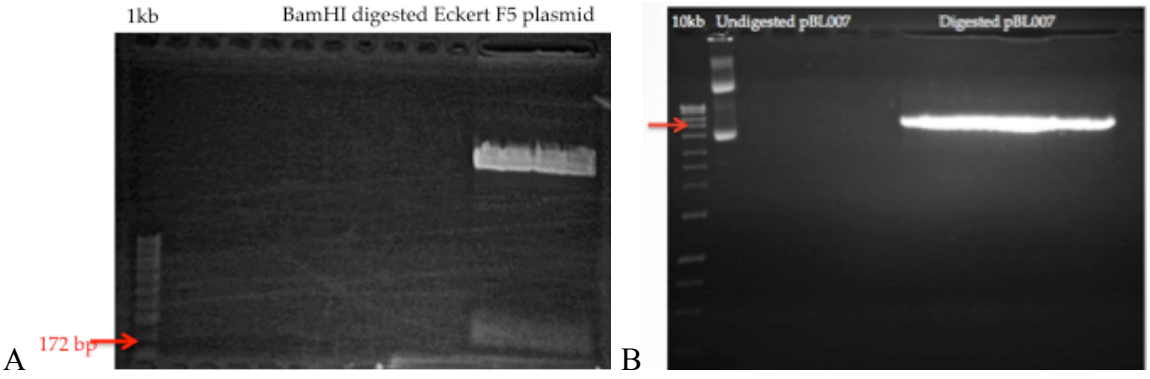


Figure 6. A) Gel electrophoresis of bacterial plasmid #457 digested with BamHI-HF. F5 fragment is at expected size 172bp (red arrow). B) Gel electrophoresis of bacterial plasmid pBL007 digested with BamHI-HF. Expected size is at 5344bp (red arrow). Lane 2 is undigested plasmid. For both A) and B), the bands are the expected size were cut out and gel purified.

Once the F5 fragment and the vector were ligated together, the construct was transformed into heat shock competent JM109 *E. coli* cells and resultant colonies were checked via colony PCR (Figure 7). As shown in the gel electrophoresis, the lanes that showed double bands were expected to be successful clones; indeed the presence of AT repeats within the sequence is known to cause multiple bands on a gel due to various secondary structures forming and traveling through the agarose at different speeds. In contrast, the negative control (colony PCR of the MCS of pBL007 without F5) showed a clear single band at 1000bp (lane 12 on bottom gel, Figure 7).

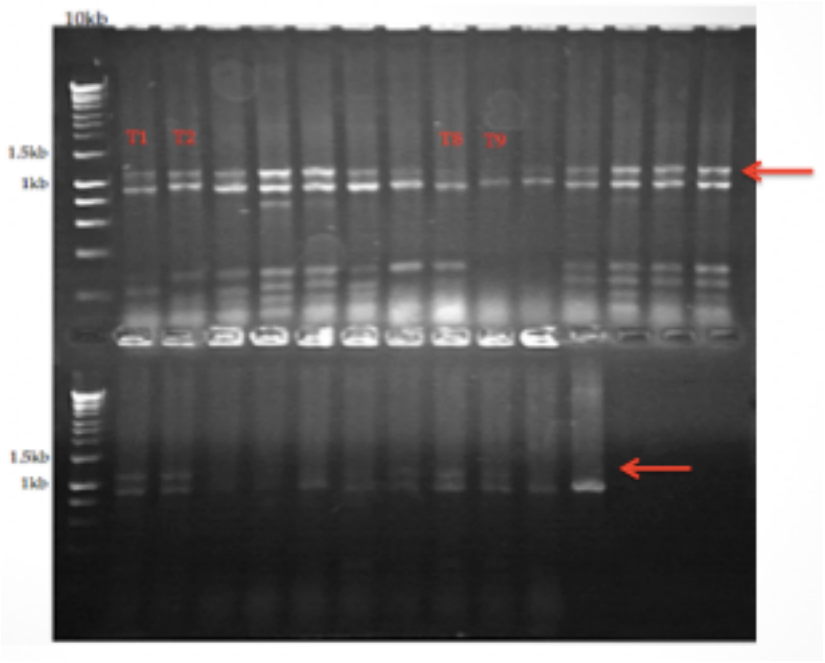


Figure 7. Gel electrophoresis of the colony PCR done to check for successful transformants containing the pBL007-F5 plasmid. Expected size of the amplified MCS of pBL007 without F5 was 1000bp, and with F5 was 1172bp. Lane 12 on the bottom gel was the negative control (colony PCR of the MCS of pBL007 without F5). Lanes labeled T1, T2, T8, and T9 were confirmed as successful transformants via sequencing and saved as glycerol stocks #520/523 (orientation 1, T1/T9) and #521/522 (orientation 2, T2/T8).

The successful recombinant plasmids strains #520 and 521 (one for each orientation of F5) were then linearized via XBaI digestion (Figure 8) and transformed into yeast. Transformant colonies were analyzed via colony PCR in order to check the 5'

and 3’ ends flanking F5 (Figure 9). The AT repeats of F5 and its flanking regions were sequenced to confirm that no mutations, expansions, or contractions had taken place (See Appendix for sequencing results).

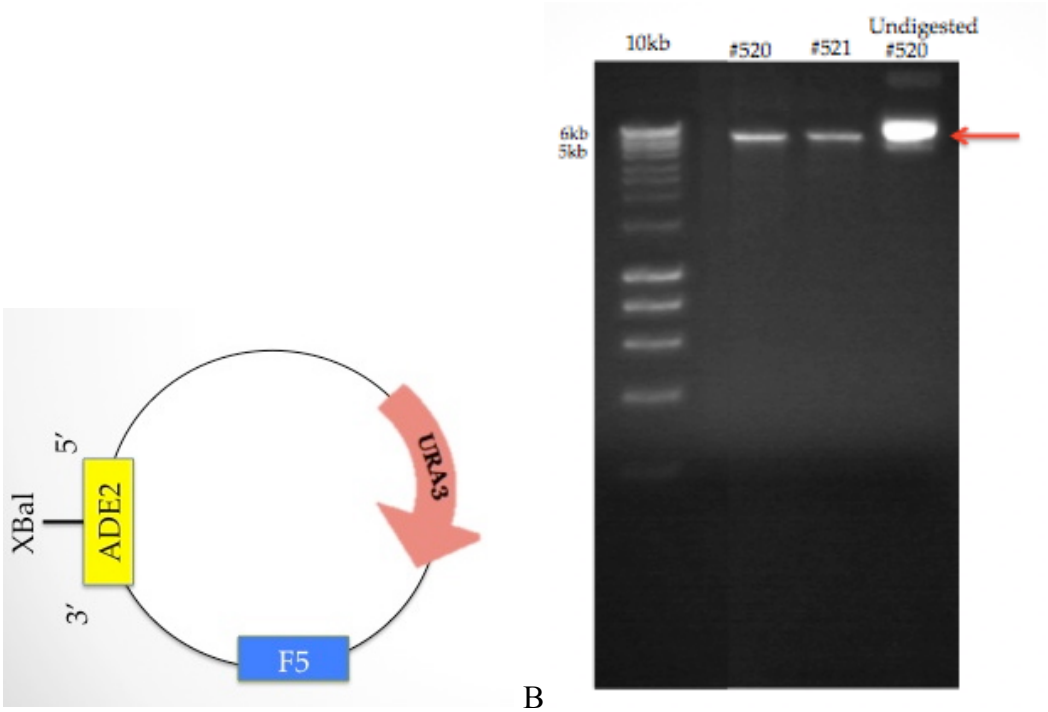
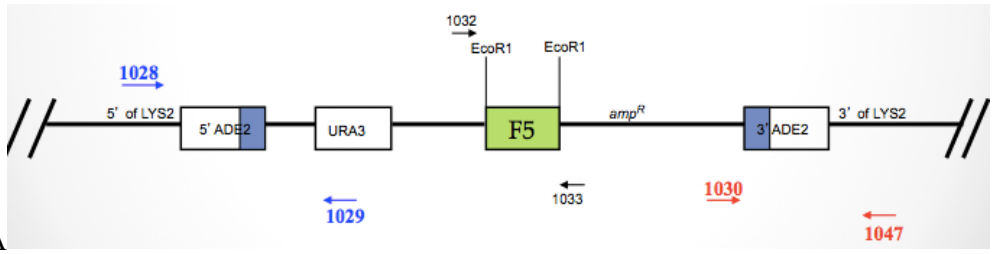


Figure 8. A) Schematic of plasmid linearized via XBaI digestion. B) Gel electrophoresis of XBaI digestion of the F5+pBL007 constructs in both orientations. The undigested control shows supercoiled plasmid DNA, while both digests showed single bands, indicating successful digestion. Size of the plasmid with F5 is 5516bp.

A



B

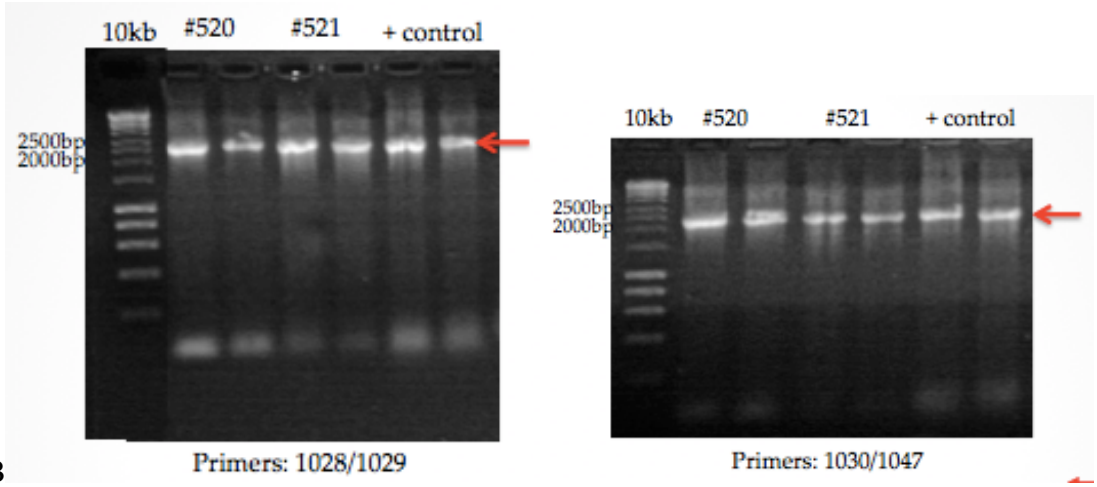


Figure 9. A) Schematic of the F5 construct inserted into the yeast genome with primer locations. B) Gel electrophoresis of colony PCR to check for successful F5 construct integration into yeast. Undiluted (first band of bacterial strains #520/521 lanes) and 1:100 diluted (second band of #520/521 lanes) gDNA were used as template for the two strains. The gel on the left uses primer set #1028/1029 for 5' junction check and gel to the right uses primer set #1030/1047 for 3'junction check. Expected product size indicated by the red arrows.

The confirmed successful transformants were saved as glycerol stocks (two strains for each orientation) and the Direct Duplication Recombination Assays (DDRA) were performed. The confirmed construct made for this assay was inserted at the natural LYS2 locus on chromosome II. It contained the URA3 gene (1.17 kb) and the F5 sequence (170 bp), which were placed between a 1480-bp 5' fragment of the ADE2 gene and a 1204-bp 3' fragment of the ADE2 gene. This created a 968-bp duplicated segments of ADE2, represented by the red box. If breakage occurred in the 4.8-kb region between the duplicated segments of ADE2, resection and recombination could occur between the

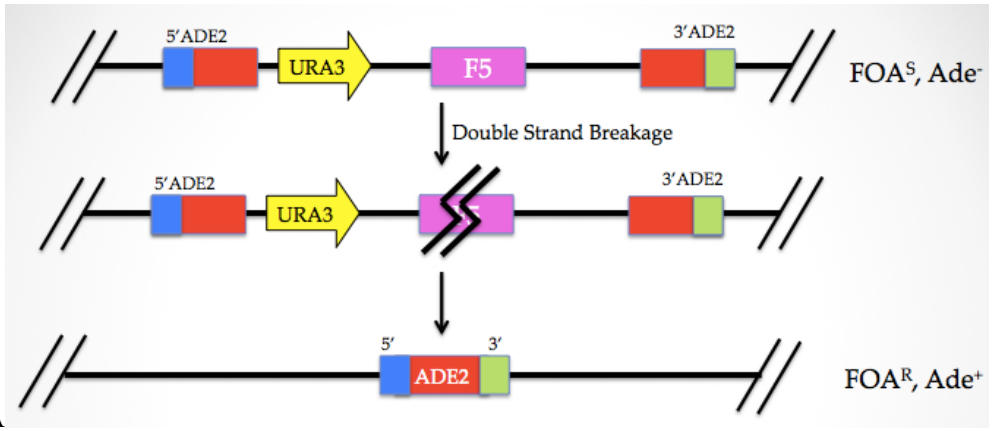
duplicated ADE2 regions, thereby restoring an intact ADE2 gene and eliminating the intervening sequences. The resulting cells will be ADE+ and FOAR (due to loss of the URA3 gene). This allows us to correlate the rate of resistance to FOA with the rate of chromosomal breakage. Results of the DDRAs showed no significant change in FOA resistance in either orientations compared to the wild type (Table 3, Figure 10).

DDRA results of F5 constructs

	F5 orientation 1	**F5 orientation 2**
Assay #1 (10col)	4.70 (#3528)	2.62 (#3527)
Assay #2 (10col)	2.25 (#3528)	2.33 (#3526)
Assay #3 (10col)	2.90 (#3528)	1.26 (#3527)
Assay #4 (10col)	5.55 (#3529)	N/A
Average Rate of FOA^R (x10-5)	3.85 (±1.33)	2.07 (±0.58)

Table 3. Strain number used for each assay is indicated in parenthesis next to the rate of FOA resistance. Difference between the two orientations was not statistically significant.

A



B

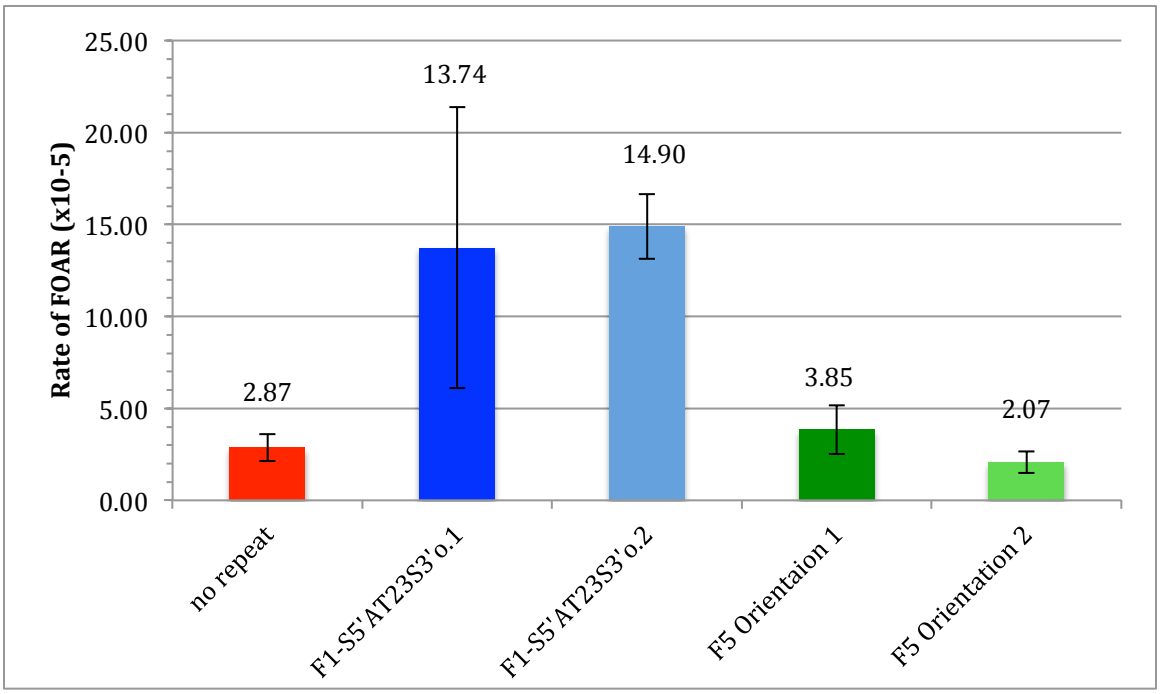


Figure 10. A) Schematic of Direct Duplication Recombination Assay (DDRA). B) DDRA results of F5 constructs in both orientations. First bar represents a collection of 3 assays previously performed in our lab on a no repeat control strain. The second and third bars are the results of assays performed by Simran Kaushal in a similar construct as the F5 ones, where 23 AT repeats of F1 were incorporated instead of the F5 region. F1-S5'AT_{23}S3'o.1 was the result of 5 assays and F1-S5'AT_{23}S3'o.2 of two assays. The mutation rate of F1-S5'AT_{34}S3' was also tested for orientations 1 and 2 and yielded rates of $10.02x10^{-5}$ and $41.13x10^{-5}$, respectively (not shown in graph). There was no statistically significant difference between the F5 orientation 1 and orientation 2 constructs and the no repeat control.

Cis Factors: Subtractive Approach

From the Zhang & Freudenreich 2007 paper, the F1 sequence was discovered to cause an increase in fragility as compared to a control sequence (Figure ??); however, this study was assessing the fragility of F1 outside of the larger FRA16D context. The first goal of this project was therefore to delete the F1 region from FRA16D within our large YAC801B6 YAC and assess fragility. If F1 plays a significant role in contributing to the fragility of FRA16D, then by deleting it we might expect a decrease in rate of breakage.

It is also our hypothesis that several secondary structure-prone sequences close together, not just F1, work in a concerted manner in order to cause the fragility observed within FRA16D. The entire FRA16D region was examined via a palindrome analysis program, which detected a number of regions, including F1, that were expected to form palindromes under the specific parameters outlined in Figure 11. Table 4 shows particular regions of FRA16D that were predicted to form stable secondary structures. Similar to F1, we wanted to delete these sequences of interested from the large FRA16D YAC801B6 to determine their effect on overall FRA16D fragility. F5 (Palindrome 4) is a region of interest due to its similarity to F1; it is composed of interrupted AT repeats but also has a poly-A sequence, and has been shown to pause and slow replication significantly in vitro (Walsh et. al., 2013). P5P5b is intriguing because it is only partially AT-rich but our palindrome analysis program predicts the formation of two very large palindromes close to one another; assessing its fragility could help elucidate whether breakage occurs due to structure formed by AT-rich regions specifically or due to any large secondary structure. P2P2b is a region near the start of the 270kb FRA16D core and

is expected to form two large palindromes that are AT-rich.. Our analysis program also detected P1, a very large cruciform just outside of the FRA16D 270kb start site. This "minimal 270kb core" required to observe expression of FRA16D is an approximate region and may not exactly encompass all of the DNA regions that significantly contribute to the FRA16D fragility. It is consequently worthwhile to investigate P1's contribution to fragility, and perhaps redefine what we take to be the FRA16D "minimal" region needed for expression.

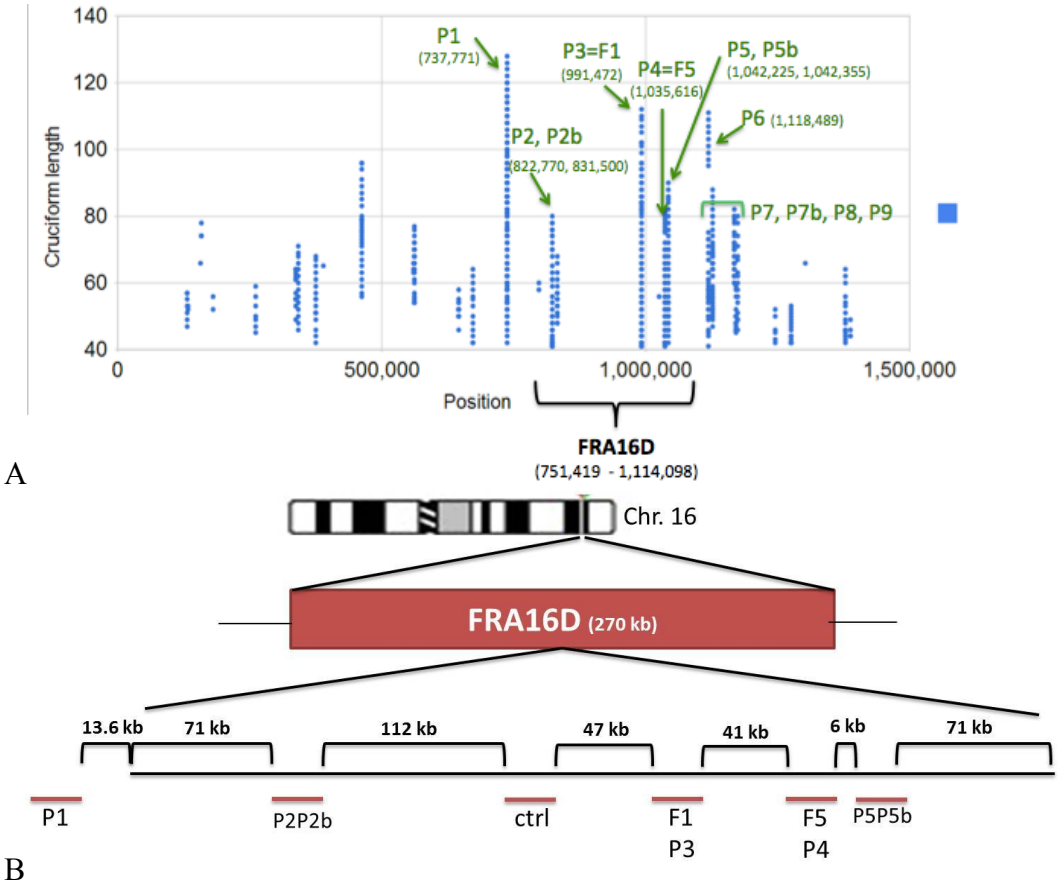


Figure 11. A) Palindrome analysis program detecting potential cruciform formation on chromosome 16. Analysis of Figure by Anoop Kumar and C.H. Freudenreich. Sequences from predicted cruciforms at position 991472 and 1035616 were matched to the known sequences of F1 and F5, respectively, allowing us to define the start and end of FRA16D in the program. Cruciforms were determined by the following criteria: Length of the stem is at least 20 bp, 3 or less mismatches allowed in the stem, 2 or less inserts/gaps allowed in the stem, length of loop is less than 12 bases, length of loop can be up to 30 bases in AT rich regions, and AT rich regions defined as region with more than 80% A or T bases. **B) Schematic of detected palindromes within FRA16D.** Distances between each region is indicated.

Table 4. Features of palindromes within FRA16D as detected by our Palindrome Analysis program.

Palindrome Name	*Position in palindrome program*	*Largest Palindrome length (bp)*	*Palindrome composition*
P1	737,771	127	Many interrupted AT repeats
FRA16D Start (270kb core)	751,419		
P2	822,770	80	26 AT repeats + 12 nt A/Ts on each side
P2b		68	
P3 (= F1)	991,472	112	52 AT repeats with 1 AA, 1 TT, 1 TG mismatch; 2 inserts T,C
P4 (=F5)	1,035,616	82	32 and 3 AT repeats, 2 A inserts, 2 A-A and 1 A-G mismatches (the poly-A flanking the F5 AT repeats is of interest but not included in palindrome)
P5	1,042,225	90	Non-AT hairpin with 3 short T runs, 3 short A runs, interspersed sequence
P5b	1,042,355	70	Mixed AT-rich 21 bp stem with large 28bp loop
FRA16D End (270kb core)	1,114,098		

The deletions were performed by replacing each sequence of interest by a selectable marker that could recombine at the desired locus on the YAC via homologous recombination. Once the markers were amplified (Figure 12) and transformed into the yeast strain #1086, proper integration at the desired locus within the YAC was verified via colony PCR (Figure 13).

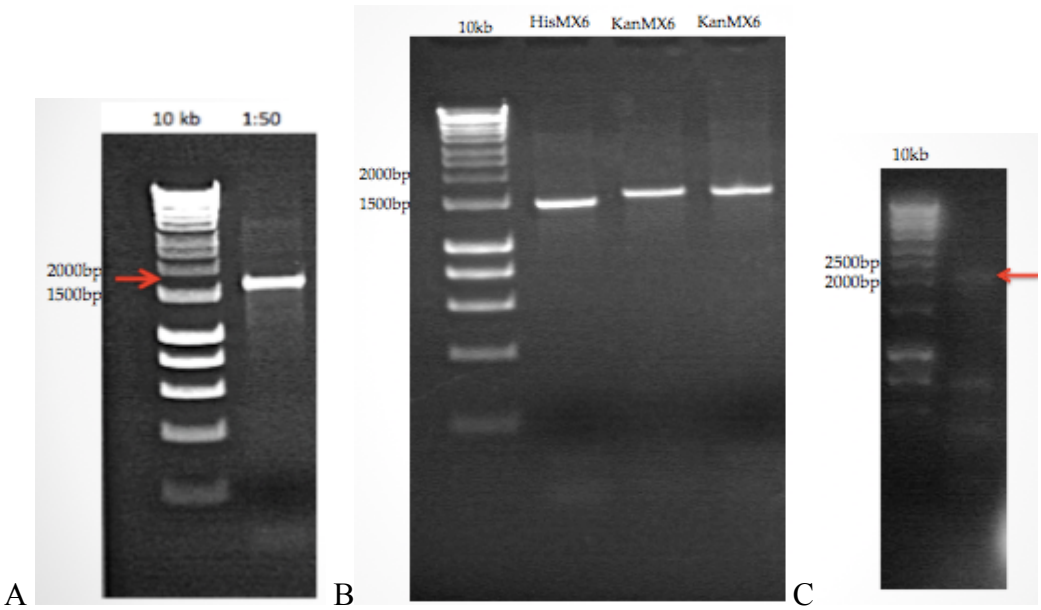
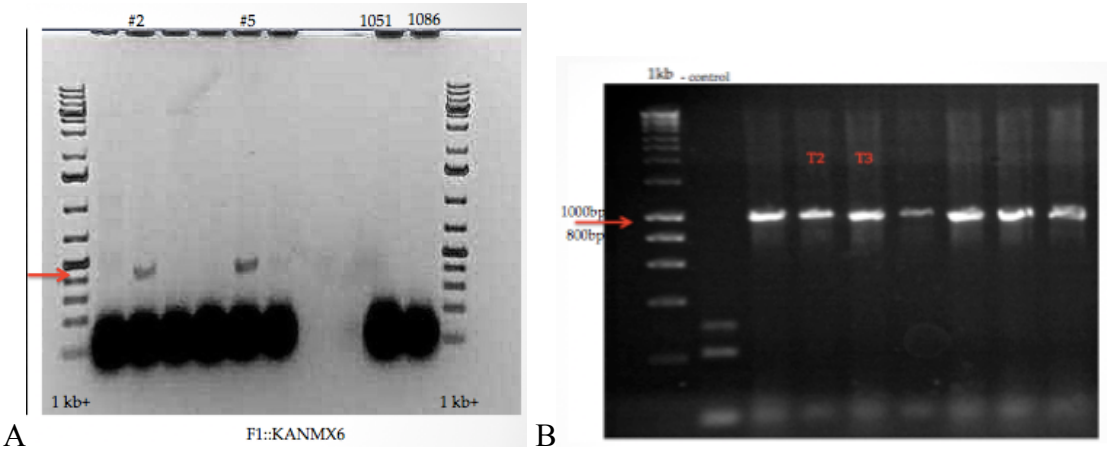
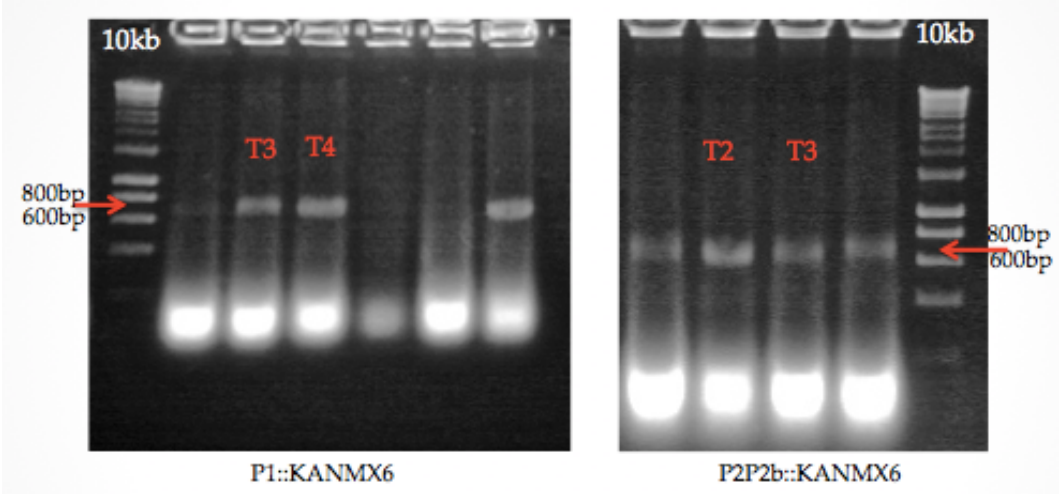


Figure 12. Gel electrophoresis of marker amplifications containing 40bp homology to the upstream and downstream regions of interests. A) KanMX6 with F1 homology. Expected size (red arrow) is 1560bp with primers #1236/1237. B) Lane 2 is His3MX6 with F5 homology. Expected size is 1450bp with primers #1448/1449. Lane 3 is KanMX6 with P1 homology (primers #1573/1574) and lane 4 is KanMX6 with P2P2b homology (primers #1571/1572). Expected size of both is 1560bp. C) Hyg with P5P5b homology, Expected size is 1800bp (red arrow) with primers #1555/1556.



C



D

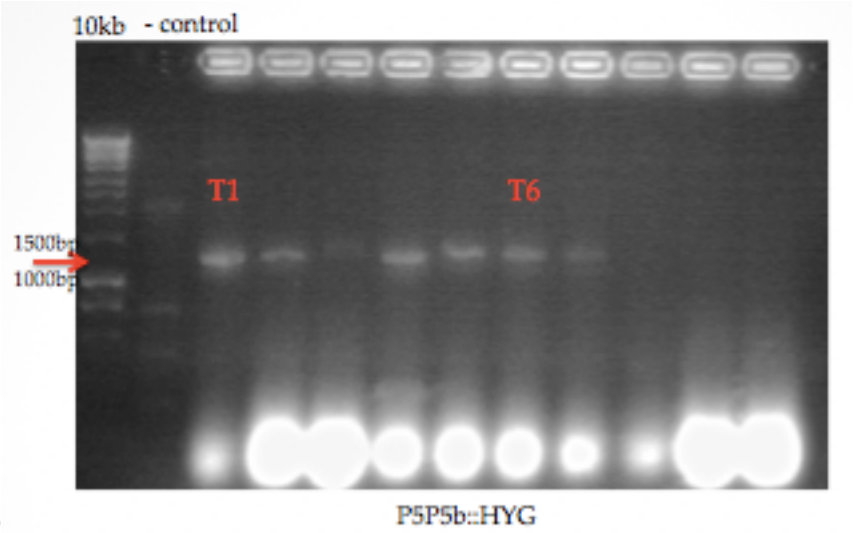


Figure 13. Gel electrophoresis to check for proper integration of marker at target locus. Negative control for each was the starting strain #1086. A) Gel electrophoresis of colony PCR to check for successful integration of KanMX6 marker at F1 locus. Expected Size is 537bp (indicated by red arrow) using primers #16/1267. Lanes labeled 1051 and 1086 were negative controls. Strains from lanes labeled #2 and #5 were saved as glycerol stocks #3076/3077. B) Gel electrophoresis of colony PCR to check for successful integration of His3MX6 marker at F5 locus. Expected Size is 954bp (indicated by red arrow) using primers #1545/374. Strains from lanes labeled T2 and T3 were saved as glycerol stocks #3513/3514. The transformation was repeated to obtain two more successful transformants saved as stocks #3587/3588. C) Gel electrophoresis of colony PCR to check for successful integration of KanMX6 marker at P1 locus (left gel) and at P2P2b locus (right gel). Expected Size is 688bp and 668bp, respectively, (indicated by red arrow) using primers #16/1704 and #16/1704. Strains from lanes T3 and T4 (left gel) were saved as glycerol stocks #3589/3590. Strains from lanes labeled T2 and T3 (right gel) were saved as glycerol stocks #3591/3592. D) Gel electrophoresis of colony PCR to check for successful integration of Hyg marker at P5P5b locus. Expected size is 1157bp (red arrow) with primers #396/1481.

YAC Breakage Assays were then performed and compared to the breakage rate of the control starting strain #1086 (Table 5). The deletion of F1 and F5 showed a significant decrease in FOA resistance compared to the wild type, while deletion of P5P5b showed no significant change (Figure 14).

Assay #	Starting Strain (#1086)	F1 deleted	F5 deleted	P5P5b deleted
1	17.56%	12.55% (#3076)	6.94% (#3513)	16.69% (#3517)
2	17.32%	12.80% (#3076)	8.93% (#3587)	16.03% (#3517)
3	18.36%	13.78% (#3076)	12.07% (#3588)	
4		12.33% (#3077)	13.97% (#3514)	
Average Breakage Rate (±standard deviation)	17.75% (±0.55)	12.86% (±0.64)	10.48% (±2.72)	16.36% (±0.33)

Table 5. Results of YAC breakage assays. Breakage rate was calculated by number of colonies that grew on FOA-Leu plates over the total cell count (number of colonies that grew on YC-Leu). Strain numbers are indicated in parentheses for each assay.

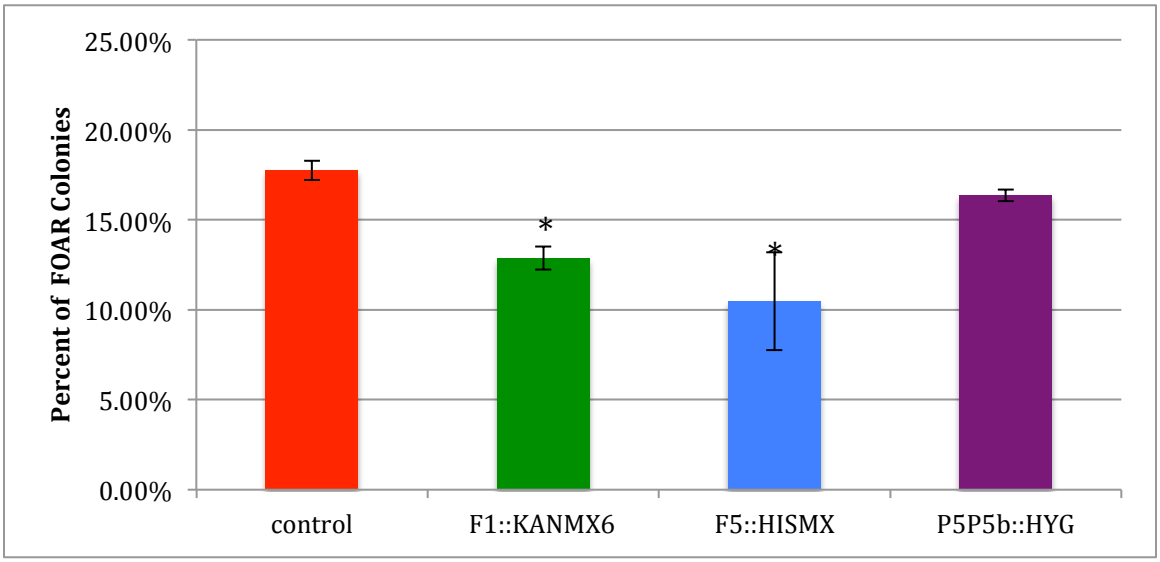


Figure 14. YAC breakage assay results. Control was the starting strain used to make the deletions. Both F1 and F5 deletes showed a statistically significant decrease in average breakage rate compared to the control. p=0.0118 for F5. P=0.0001 for F1. Between F5 and F1 p=0.1872 (not significant).

In order to assess the structural state of the constructed strains and ensure that the starting YAC is intact, a Pulse Field Gel Electrophoresis (PFGE) can first be used to separate the chromosomal DNA by size, followed by a Southern blot using probes that anneal to each end of the YAC. This experiment is currently under way; one successful PFGE was thus far performed on strains #3077 (F1::KANMX6) and #3513 (F5::HISMX) as shown in Figure 15, although the presence of the YAC is difficult to determine from this PFGE.

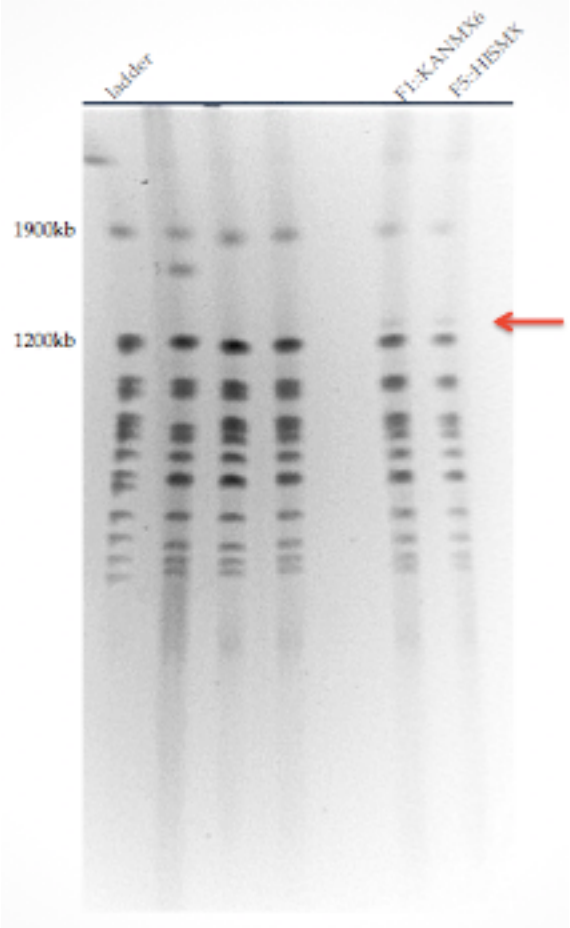


Figure 15. Pulse Field Gel electrophoresis of F1::KANMX6 and F5::HISMX. Lanes 2-5 were yeast strains lacking YAC801B6. Red arrow indicates tentative location of YAC801B6

Trans Factors: Deletion of *MUS81*

Mus81 forms complexes (with Eme1 and Eme2) in order to serve as an endonuclease, which is both necessary for fork restart and also, as a result, for CFS expression since it generates a double stranded break (DSB). By deleting *MUS81*, we could then expect either an increase in fragility because of a disrupted mechanism for proper replication or a decrease in fragility due to the presence of fewer DSBs. To study how fragility of subregions of F1 are affected after the deletion of *MEC1*, we use the chromosome II assay described in Figure 10, this time with yeast strains that contain a specific cassette with the 34 (AT) repeats of the F1 sequence and its flanking regions. This cassette is at the *LYS2* locus on chromosome II.

MUS81 was replaced with KANMX6 from the yeast strain #2525 (consisting of the region F1 with 34 AT repeats at the *LYS2* locus on chr.II), as well as from the control strain #2863 (consisting of a 386bp DNA sequence from FRA16D not predicted to form any secondary structures). Once the *mus81*::KANMX6 region was successfully amplified from strain #1791 and transformed into the yeast strains #2525 (S5' F1-AT_{34}S3') and #2863 (no repeat control), proper integration at the desired locus within the chromosome was verified via colony PCR (Figure 16). Another PCR was also performed as a means to ensure that the deleted gene was not maintained extrachromosomally, a mechanism which can take place via Break-Induced Repair (BIR) when a double stranded break is introduced. One end of the DSB may initiate new DNA synthesis, such that when the newly synthesized DNA is displaced from the template, it rejoins the other end via nonhomologous end-joing, thus creating a circular extrachromosomal element. If an

autonomous replication sequence (ARS) is present as part of this fragment, the extrachromosomal element can be maintained, which might allow expression of the knocked-out gene (Kraus et. al., 2001). We can check for the presence of such an element by designing a primer that anneals just outside the region of homology and another within the gene to be knocked out. If the target gene was properly deleted, then no PCR product should be made from these two primers (Figure 17).

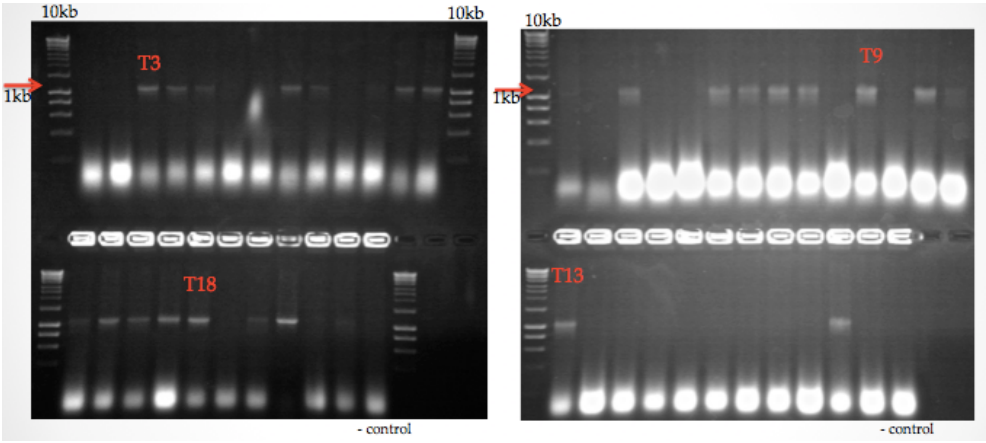


Figure 16. Gel electrophoresis of colony PCR to check for successful integration of KANMX6 at MUS81 locus. Gel on the left shows transformants from strain #2525, and gel on the right shows transformants from strain #2863. Expected size was 1100bp (indicated by red arrow). Lanes labeled in red correspond to the colonies that were saved as glycerol stocks #3375/3376 (no repeat) and #3377/3378 (S5'F1-AT_{34}S3').

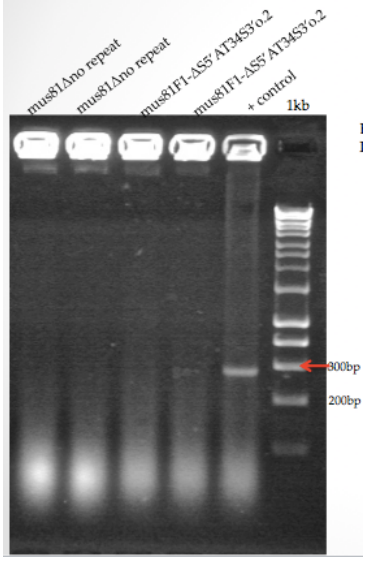


Figure 17. Gel electrophoresis of colony PCR to confirm the lack of the MUS81 ORF. Expected size was 300bp (indicated by red arrow) if mus81 remains present in the cell. No band is expected if mus81 was successfully deleted.

Once the transformant strains were constructed, the Direct Duplication Recombination Assays were performed for each strain and the average mutation rates were calculated using the method of the median of the FALCOR program (Table 6 and Figure 18). The mus81ΔF1-S5'AT_{34}S3' showed a statistically significant decrease in mutation rate compared to the wild type F1-S5'AT_{34}S3', while mus81Δno repeat showed rates equivalent to the wild type no repeat

Table 6. Results of DDRAs for MUS81Δ transformants. All were 10-colony assays.

Assay #	wt F1-S5'AT34S3' o.2 (#2525)	mus81Δ no repeat ctrl (#3375)	mus81Δ F1-S5'AT34S3' o.2 (#3377/3378)
1	47.70	4.43	15.16 (#3377)
2	30.70	3.53	12.91 (#3377)
3	45.00	3.99	13.19 (#3378)
Average Mutation Rate ($x10^{-5}$) (±std deviation)	41.13 (±7.46)	3.98 (±0.45)	13.75 (±1.00)

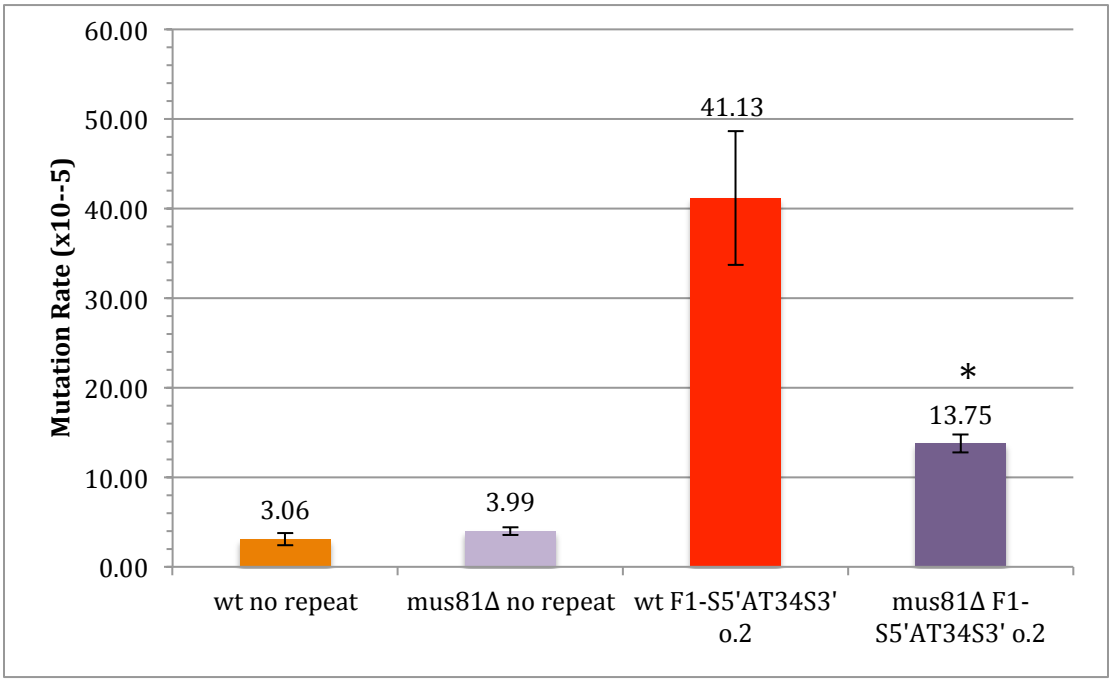


Figure 18. Results of DDRAs for MUS81Δ transformants. The decrease in fragility for mus81Δ F1-S5'AT34S3'o.2 was statistically significant compared to the wild type (*p=0.0284).

Trans Factors: Deletion of MEC1

According to our model, replication polymerase pausing can occur at sites of secondary structure formation, leaving single-stranded DNA regions exposed and stalling the replication fork. In humans, this fork stalling leads to ATR-dependent checkpoint activation, which is responsible for blocking the firing of new replication origins, preventing entry into mitosis, and promoting repair (Franchitto 2013). ATR thus plays an essential role in maintaining stability and preventing fragility at CFS. Mec1 is the budding yeast homolog of ATR and is part of a complex responsible for phosphorylating downstream transducers and effectors such as Rad9 and Rad53 (Paciotti et al. 2000), which induce cell cycle arrest following DNA damage. As mentioned, Mec1 is a well-studied protein and its effects on common fragile fragility are well-documented. Therefore, this experiment was a proof of principle that our assay allows us to show to detect the increase in fragility expected from deleting *MEC1*. Prior to deleting *MEC1*, *SML1* must also be knocked out. This is because Sml1 is an inhibitor of ribonucleotide reductase (RNR) and is inactivated via phosphorylation by Mec1 (Zhao et. al., 2001) during S phase. Consequently, an *sml1Δmec1Δ* strain must be made to maintain nucleotide pool levels high enough during S phase.

SML1 was first replaced from the yeast strain #2525 (consisting of the region F1 with 34 AT repeats at the *LYS2* locus on chr.II), as well as from the control strain #2863 (consisting of a 386bp DNA sequence from FRA16D not predicted to form any secondary structures). Once the *sml1::TRP1* region was successfully amplified from strain #657 (Figure 19) and transformed into the yeast strains #2525 and #2863, proper

integration at the desired locus within the chromosome was verified via colony PCR (Figure 20). Once the *mec1::KANMX6* region was successfully amplified from strain #657 (Figure 21), proper integration at the desired locus was verified via colony PCR (Figure 22). The absence of *MEC1* maintained as an extrachromosomal element was tested and in fact showed that *MEC1* may still be in both the no repeat and the F1-S5'AT_{34}S3' strains (Figure 23), suggesting that the knockout of *MEC1* will have to be repeated.

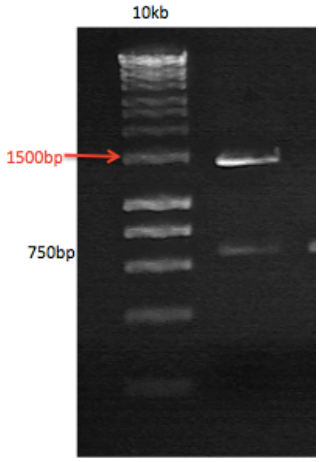


Figure 19. Gel electrophoresis of amplification of *TRP1* with flanking homology to *SML1*. Expected size was 1466bp (red arrow) with primers #570/571.

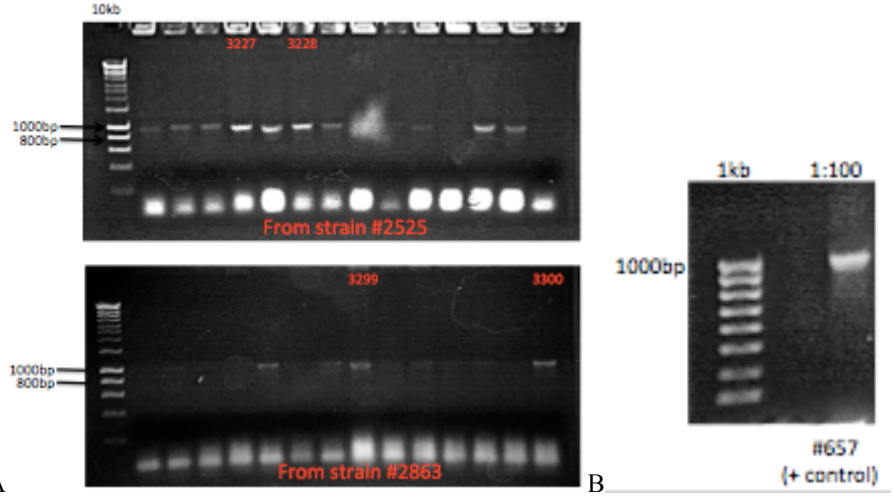


Figure 20. A) Gel electrophoresis of colony PCR to check for successful integration of *TRP1* at *SML1* locus in strains #2525 (top gel) and #2863 (bottom gel). Expected size was 950bp with primers #140/45. Glycerol stocks were saved as indicated on gel as #3227/3228 and #3299/3300. B) Positive control in sml1Δ strain using same primers.

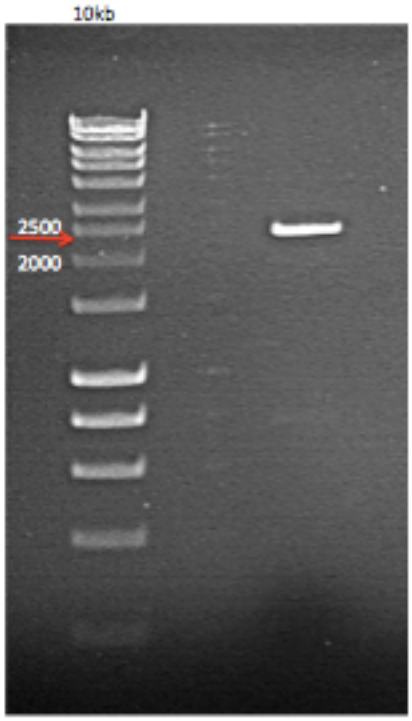


Figure 21. Gel electrophoresis of amplification of *KANMX6* with flanking homology to *MEC1*. Expected size was 2300bp (red arrow) with primers #567/568.

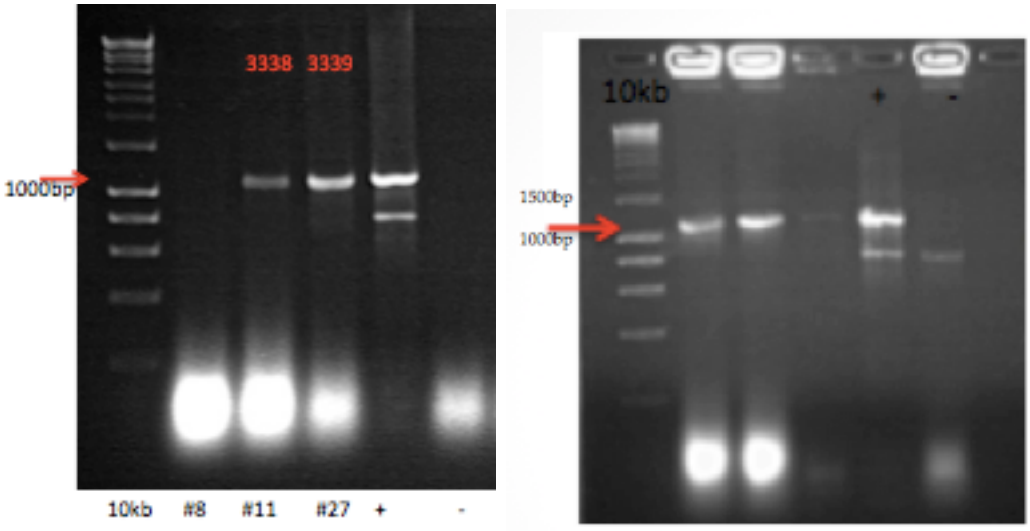


Figure 22. Gel electrophoresis of colony PCR to check for successful integration of *KANMX6* at *MEC1* locus. Gel on left are transformants made in strain #3227 (S5'F1-AT_{34}S3' sml1Δ) and gel on right in strain #3300 (no repeat sml1Δ). Expected size is 1150bp (red arrow). Positive control was mec1Δ strain #657 (Lanes 5) and negative control was starting strain #2525 (Lanes 6).

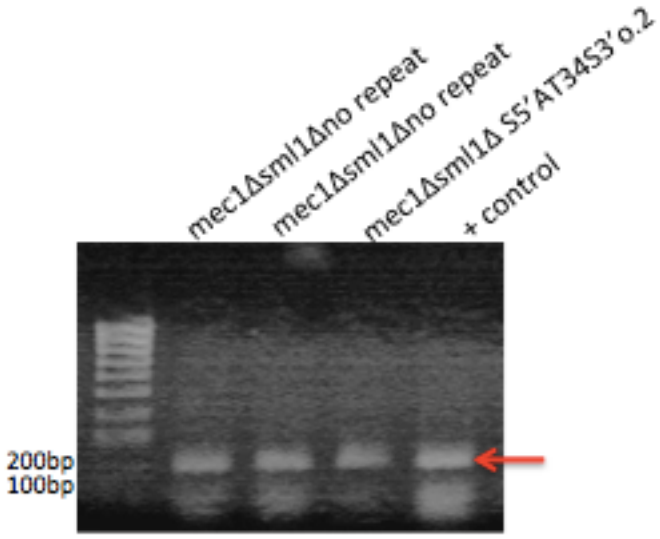


Figure 23. Gel electrophoresis of colony PCR to confirm the lack of the MEC1 ORF. Expected size was 190bp (indicated by red arrow) if mus81 remains present in the cell. No band is expected if mus81 was successfully deleted.

Once the transformant strains were constructed, the Direct Duplication Recombination Assays as previously described were performed for each strain and the average mutation rates were calculated using the method of the median of the FALCOR program (Figure 24).

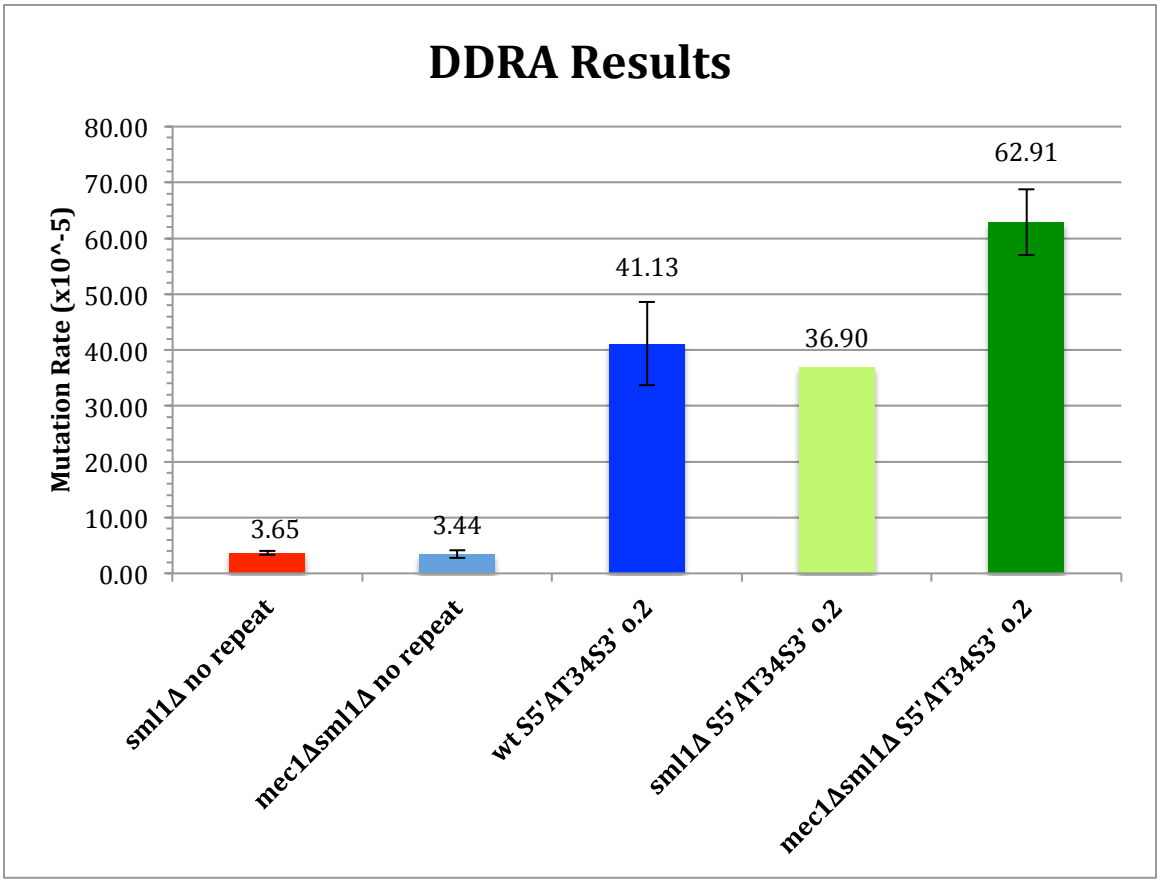


Figure 24. Results of DDRAs for mec1Δsml1Δ transformants. The increase in fragility for mec1Δ sml1ΔF1-S5'AT34S3'o.2 was only just statistically significant compared to the wild type. The sml1Δ no repeat was the result of 3 assays from strain #3300, the mec1Δsml1Δ no repeat was 4 assays from strain #3346, the wtS5'F1-AT_{34}S3'o.2 was pooled from 4 assays (one performed by myself, three by Simran Kaushal), sml1ΔS5'F1-AT_{34}S3'o.2 was 1 assay from strain #3227, and mec1Δsml1ΔS5'F1-AT_{34}S3'o.2 was 4 assays (3 from #3339, 1 from #3338). It should be noted that the mec1Δsml1Δ no repeat and the mec1Δ sml1ΔF1-S5'AT34S3'o.2 may not be true knockouts of *MEC1*.

Discussion

F1 and its effect on fragility in FRA16D

The deletion of F1 from FRA16D in YAC801B6 showed a significant decrease in average breakage frequency from 17.55% in the starting strain to 12.86% in the F1 delete strains. This almost 30% decrease in fragility suggests that F1 contributes significantly to fragility in the FRA16D region. The fact that such a change in breakage frequency was detected from removing such a small sequence was a striking finding. Indeed, the current YAC801B6 used in the experiment is 1.4MB. This stands in contrast to the minimal region of FRA16D that has been shown to be needed for CFS expression of ~270kb, and to the deleted region of F1 itself that was only a few hundred base pairs. The reported findings are also consistent with our lab's past research; the Zhang and Freudenreich 2007 paper reported an average breakage frequency of 18% as compared to the 17.55% reported here using the same strain. The same study had also introduced findings regarding F1 alone, showing that the F1 small YAC construct was capable of inducing breakage in an AT-length dependent manner. The results of this experiment therefore not only compliment the findings mentioned above, but also place them in the larger context of the FRA16D region itself.

F5 and its effect on fragility according to the Additive approach

The results obtained from introducing the F5 construct into yeast yielded low rates of FOA resistance, for orientation 1 construct 3.85×10^{-5} and orientation 2 construct 2.07×10^{-5}, suggesting that the introduced 147bp F5 sequence did not yield a detectable increase in breakage. In fact, these low rates are comparable to rates of the 384bp control of $2.87 \times 10^{-}$

^{5}, as was found by Simran Kaushal. All the more surprising is the sequence similarity between the F5 sequence, which has 24 interrupted AT repeats, compared to that of the F1-S5'AT_{23}S3', which showed higher a mutation rate (particularly orientation 2, at $14.90x10^{-5}$). Different interpretations might account for the disparity observed. First, the F5 AT repeats are interrupted in two places, once by an AAA sequence and then by a TT diclucleotide, while the F1 AT repeats are perfect. It is possible that perfect AT repeats form a more stable secondary structure than interrupted AT repeats, making the uninterrupted repeats more likely to stall replication and damage DNA.

For instance, inputting each sequence within the MFold program provides the expected structure formation with its associated thermodynamic properties, such that the entire F5 region forms three hairpins with a total ΔG of -17.5 kcal/mol, the AT-hairpin having a ΔG of -3.29 kcal/mol. This in contrast to the F1 region with a much more negative total ΔG of -110 kcal.mol (due to the fact that the F1 region alone is larger than the F5 one), but also a much more negative ΔG of -16.06 kcal/mol for the AT-hairpin alone. This suggests that the F1 hairpin might be more likely to form, or perhaps more difficult to unfold than the F5 secondary structure. As a result, the replication machinery may have more difficulty synthesizing through the F1 region. Of course, the MFold program provides only the most likely structure formations under specific cellular conditions and optimizes the hairpin base with the highest possible number of AT base-pairings, which is not necessarily representative of our experimental conditions or of factors such as DNA-protein interactions that might alter the DNA secondary structure. It may be the case that the interrupted AT repeats of F5 actually lead to the formation of multiple small hairpins, which would not contribute to fragility significantly.

Continuing to investigate whether the interrupted versus uninterrupted AT-repeats actually yield significant differences in rates of breakage would require genetic manipulation. For instance, an interruption in the AT repeats of the F1-S5'AT_{23}S3' construct could be introduced to more closely resemble F5, while the F5 AT repeats could be replaced with a set of 24 uninterrupted repeats to mimic the F1 sequence. Following the experiments done with F1-AT length, an assay measuring fragility of F5 with increasing numbers of AT repeats might also shed light on the relationship between AT length and breakage as it pertains to F5 specifically.

A feature unique to F5 as compared to F1 is the poly-A_{28} located just one nucleotide away from the AT-repeats. The Eckert group detected significant stalling by the lagging strand DNA polymerase δ throughout the $[A]_{28}$ repeat of F5, but not the $[T]_{28}$ in an *in vitro* replication system (Walsh et. al, 2013). This suggested that the presence of the poly-A, in addition to the structure-forming AT repeats, might contribute to perturbing fork progression, thereby exacerbating the chance of breakage. The results of our experiment however do not support this conclusion; the two orientations of F5 show similar rates of FOA resistance, suggesting that fragility was unaffected by the direction of replication with respect to the poly-A run in a chromosomal context. Walsh et. al. 2013 also hypothesized mononucleotide $[A]_n$ tracts may inhibit fork progression by promoting the formation of triplex DNA, a DNA structure in which one homopurine or homopyridmidine strand binds to the B-form DNA double helix, specifically forming Hoogsteen pairs with purines of the Watson-Crick basepairs (Frank-Kamenetskii & Mirkin, 1995). Triplex DNA has in fact been shown to promote polymerase stalling *in vitro* (Baran et. al., 1991) and *in vivo* (Krasilnikova and Mirkin, 2004). If the poly-As of

F5 are in fact capable of forming triplex DNA, it is possible that the structure actually inhibits DSB resection, thus resulting in a failure to initiate homologous recombination and accounting for the low detection of breakage in our assay. A study stemming from our lab actually supported a similar hypothesis, showing that the deletion of a 17 bp-hairpin-forming sequence adjacent to the F1 AT_{34} repeats resulted in a much higher rate of FOA resistance, linking the 17 bp-structure to the inhibition of recovery of broken chromosomes in the context of the DDRA (Soo-Mi Alison Lee, Senior Honors Thesis 2011). Removal of the $[A]_{28}$ repeats and testing the fragility of the F5 repeats alone could therefore speak to the contribution of the poly-A in detecting breakage within our assay.

F5 and its effect on fragility in the context of FRA16D according to the Subtractive Approach

In contrast to the additive approach, which did not show any significant change in breakage in the presence of the F5 construct, the YAC breakage assays for F5 delete strains revealed a very low breakage frequency with an average of 10.48%, suggesting that the F5 region does in fact play a significant role in contributing to fragility of the FRA16D region and that its contribution is equivalent to that of F1. To reconcile findings form the additive and subtractive approaches, there are a few things to consider. First, a much larger region of F5 was deleted (1400bp) compared to the one that was added in (147bp). This feature alone is however unlikely to account for the discrepancy just described; the upstream and downstream F5 regions that were removed are not expected to form significant secondary structures according to MFold, and were not detected by our palindrome analysis program as potential candidates for palindrome formation,

suggesting that these regions should not significantly affect breakage. Of greater importance is the fact that the two assays are not actually parallel and do not measure the same features; in one case, the F5 sequence alone is introduced as part of a chrII construct, while in the other, the presence and absence of F5 are measured in the context of the FRA16D region. It may be the case that F5 alone is not a particularly strong inducer of breaks but that, as our model suggests, it does contribute to the overall fragility of the CFS. One possible interpretation involves F5 acting as a type of "enabler", where DNA polymerase stalls at the F5 site but rather than causing breakage, it slows down the polymerase so much that when it reaches another difficult-to-replicate region it cannot synthesize past it. Alternatively, the reverse process might take place, where DNA polymerase causes breakage at the F5 site only if it has encountered a significantly difficult region to replicate prior to reaching F5. Both of these scenarios account for detecting no changes in breakage upon the introduction of F5 alone into chromosome II.

P5P5b and its effect on fragility in FRA16D

The YAC breakage assays revealed no significant change in fragility upon deletion of the P5P5b region, with a breakage frequency of 16.36% compared to that of the wild type at 17.75%. According to our palindrome analysis program, P5P5b is expected to form two large palindromes, one that is 90 bp and not AT-rich and the other 70 bp long, which is AT-rich and consists of a 21 bp-stem with a large 28 bp loop. It is possible that these structures do not actually form *in vivo*. Alternatively, they may form but their effect on fragility within FRA16D is marginal, either due to the nature of the secondary structure itself or perhaps because of its location within FRA16D. An

interesting next step would be to assess the fragility of P5P5b as a stand-alone sequence through the additive approach, in an attempt to determine whether the expected palindrome formation affects breakage at all. Furthermore, since the different delete strains were constructed with different markers, it may be worthwhile to investigate whether the deletion of P5P5b along with that of another sequence actually exacerbates the breakage frequency; an especially interesting candidate might be F5 together with P5P5b, since these regions are less than 6 kb apart and might therefore work in a concerted manner to stall the replication machinery to the point of creating breakage.

PFGE to confirm starting YAC801B6

Results obtained for the YAC breakage assays are affected by the limitations of our experimental setup. Indeed YAC801B6 has been observed to undergo many chromosomal deletions not seen in a control YAC lacking the FRA16D region (Zhang & Freudenreich, 2007), such that YAC801B6 is prone to frequent deletions and translocations. There are a number of sites where these rearrangements and deletions would affect the experimental results. For instance, the YAC is known to heal in many different places (a very low percentage of cells actually use the telomere seed sequence G4T4 to heal). It is possible that these other regions that the YAC normally shortens to are deleted, making healing less efficient. Additionally, some of the FRA16D sequence may be deleted while the LEU and URA3 markers are still maintained, thereby making the starting YAC less fragile in the first place. Other deletions could also occur within the URA3 gene, resulting in a strain already resistant to FOA. Performing a physical analysis of the YAC via Pulsed Field Gel Electrophoresis (PFGE) is a useful tool to assess the

structural state of the constructed delete strains and verify that the starting YAC is intact, using probes that anneal to either the *TRP1* or *URA3* sequences present on each end of the YAC. As of now, the Pulse Field Gel is running successfully, despite a poor resolution of the largest bands at the top of the gel. Probing of the chromosome II with an ISW1 probe was successful, while probing of the YAC801B6 with the URA3 probe did not yield any conclusive results (data not shown); since the YAC is much larger than chromosome II, however, it may be the case that the URA3 probe is functional but the upper bands of the gel remain poorly resolved. Continued troubleshooting with the conditions and parameters of the PFGE will therefore be needed to obtain detectable upper bands before performing another Southern blot.

The effect of deleting *MEC1* on fragility

The assessment of fragility upon deletion of *mec1* was a proof of principle experiment, showing that our Direct Duplication Recombination Assay (DDRA) using a subregion of FRA16D in yeast could successfully detect the well-documented increase in FRA16D fragility expected from deleting an ATR kinase in human cells. We chose to use the F1-AT_{34} orientation 2 construct, which showed the greatest rate of FOA^R in the DDRA assay. Results of the experiment did show a statistically significant increase in rate of breakage at $62.91x10^{-5}$ for the *mec1Δsml1Δ* F1-S5'AT_{34}S3'o.2 strain, compared to the wild type F1-S5'AT_{34}S3'o.2 at $41.13x10^{-5}$ (Figure 12). As expected, there was no difference in breakage between the wild type and the *sml1Δ*. Although the increase in breakage was observed, it was mild and lower than the anticipated effect. The reason behind this observation was revealed via a PCR analysis that confirmed the presence of

the *MEC1* gene after its deletion, suggesting its maintained presence in the cell, perhaps as an extrachromosomal element (Figure 11). The fact that an increase in breakage was still detected is likely due to the fact that the *mec1*Δ phenotype must be a partial one; the *MEC1* gene, although preserved extrachromosomally, may not be as efficiently expressed as it is in the wild type. Further studies such as an HU assay might tell us more regarding the phenotypic characteristic of the constructed strains. It would be worthwhile to construct new *mec1*Δ strains both confirmed as true transformants and lacking the knocked out gene, and repeat the DDRAs in an effort to attain a more accurate understanding of the effect of breakage in the total absence of *MEC1*.

The effect of deleting *MUS81* on fragility

The deletion of *MUS81* in the F1-S5'AT_{34}S3'o.2 strain showed a very significant decrease in rate of FOA resistance (13.75×10^{-5}) as compared to the wild type (41.14×10^{-5}). Interestingly, the *mus81* delete in the no repeat strain remained at wild type levels (between 3 and 4×10^{-5}), supporting the fact that Mus81 action must be specifically targeting CFSs. These results are also congruent with the model put forth by the Hickson group, which suggests that Mus81 (in a complex with Eme1) acts as an endonuclease that leads to the formation of DSBs in metaphase chromosomes at CFSs (Ying et. al, 2013). In the absence of Mus81, fewer breaks were therefore created at the F1-S5'AT_{34}S3'o.2 locus, resulting in a lower rate of FOA resistance in our assay. To keep in mind is that lower fragility does not make healthier cells; in fact Ying et. al. 2013, observed an accumulation of DNA damage in cells that were incapable of CFSs expression due to the absence of Mus81. The role of this protein in causing fragility is therefore all the more

interesting as it suggests a mechanism by which CFS expression could in fact be a regulated and a valuable cellular process, not simply an inadvertent chromosomal mishap that occurs during chromosome condensation.

Concluding Remarks

The contribution to fragility of F1 versus F5 is an interesting contrast, not only for the purposes of understanding FRA16D expression, but also towards shedding light on the broader mechanisms underlying breakage at CFSs. In particular, findings from this project lend themselves to conducting further research on the importance of perfect AT repeats like F1, and how its particular hairpin formation might yield greater fragility than a secondary structure that is partially AT-rich like P5P5b or contains interrupted AT repeats like F5. The subtractive approach also supported our hypothesis that multiple regions within FRA16D are in fact responsible for its expression; to continue investigating this hypothesis, it may interesting to carry out concomitant deletions of these AT-rich regions or manipulate their location within FRA16D, as well as perform breakage assays on the P1 and P2P2b delete strains that were constructed in this project. Finally, the findings brought out from the deletion of *MUS81* were not only congruent with Mus81's role as an endonuclease that causes breaks at CFSs, but also showed that our Direct Duplication Recombination Assay is successful in assessing breakage within our F1-AT_{34} orientation 2 construct. The use of the DDRA is therefore an exciting tool that may be used to study the role of any number of proteins in regulating CFS expression.

Resources

Arlt, Martin F., et al. "Molecular characterization of FRAXB and comparative common fragile site instability in cancer cells." *Genes, Chromosomes and Cancer* 33.1 (2002): 82-92.

Arlt, Martin F., et al. "Common fragile sites as targets for chromosome rearrangements." *DNA repair* 5.9 (2006): 1126-1135.

Baran, Nava, Aviva Lapidot, and Haim Manor. "Formation of DNA triplexes accounts for arrests of DNA synthesis at d (TC) n and d (GA) n tracts." Proceedings of the National Academy of Sciences 88.2 (1991): 507-511.

Bignell, Graham R., et al. "Signatures of mutation and selection in the cancer genome." *Nature* 463.7283 (2010): 893-898.

Casper, A. M., Nghiem, P., Arlt, M. F., & Glover, T. W. (2002). ATR regulates fragile site stability. Cell, 111, 779-789

Coquelle, Arnaud, et al. "Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons." *Cell*89.2 (1997): 215-225.

Debatisse, Michelle, et al. "Common fragile sites: mechanisms of instability r revisited." *Trends in Genetics* 28.1 (2012): 22-32.

Finnis, M., Dayan, S., Hobson, L., Chenevix-Trench, G., Friend, K., & Ried, K. et al. (2005). Common chromosomal fragile site FRA16D mutation in cancer cells. Human Molecular Genetics, 14(10), 1341-1349.

Frank-Kamenetskii, Maxim D., and Sergei M. Mirkin. "Triplex DNA structures." Annual review of biochemistry 64.1 (1995): 65-95.

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

Freudenreich, Catherine H. "Chromosome fragility: molecular mechanisms and cellular consequences." *Frontiers in bioscience: a journal and virtual library* 12 (2006): 4911-4924.

Gallo-Fernández, María, et al. "Cell cycle-dependent regulation of the nuclease activity of Mus81–Eme1/Mms4." *Nucleic acids research* (2012): gks599.

Georgakilas, Alexandros G., et al. "Are common fragile sites merely structural domains or highly organized “functional” units susceptible to oncogenic stress?." *Cellular and Molecular Life Sciences* 71.23 (2014): 4519-4544.

Glover, T. W., Arlt, M. F., Casper, A. M., & Durkin, S. G. (2005). Mechanisms of

common fragile site instability. Human Molecular Genetics, 14(2), R196-R205.

Helmrich, A., Stout-Weider, K., Matthaei, A., Hermann, K., Heiden, T., & Schrock, E. (2007). Identification of the human/house syntenic common fragile site FRA7K/Fra12C1—relation of FRA7K and other human common fragile sites on chromosome 7 to evolutionary breakpoints. International Journal of Cancer, 120(1), 48-54.

Helmrich, Anne, Monica Ballarino, and Laszlo Tora. "Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes." *Molecular cell* 44.6 (2011): 966-977.

Huebner, Kay, and Carlo M. Croce. "FRA3B and other common fragile sites: the weakest links." *Nature Reviews Cancer* 1.3 (2001): 214-221.

Huebner, Kay, et al. "The role of the FHIT/FRA3B locus in cancer." *Annual review of genetics* 32.1 (1998): 7-31.

Krasilnikova, Maria M., and Sergei M. Mirkin. "Replication stalling at Friedreich's ataxia (GAA) n repeats in vivo." Molecular and cellular biology 24.6 (2004): 2286-2295.

Le Tallec, Benoît, et al. "Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes." *Cell reports* 4.3 (2013): 420-428.

Le Tallec, Benoît, et al. "Updating the mechanisms of common fragile site instability: how to reconcile the different views?." *Cellular and Molecular Life Sciences* 71.23 (2014): 4489-4494.

Mirkin, Sergei M. "DNA structures, repeat expansions and human hereditary disorders." *Current opinion in structural biology* 16.3 (2006): 351-358.

Naim, Valeria, et al. "ERCC1 and MUS81–EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis." *Nature cell biology* 15.8 (2013): 1008-1015.

Negrini, Simona, Vassilis G. Gorgoulis, and Thanos D. Halazonetis. "Genomic instability—an evolving hallmark of cancer." *Nature reviews Molecular cell biology* 11.3 (2010): 220-228.

O'Keefe, Louise V., and Robert I. Richards. "Common chromosomal fragile sites and cancer: focus on FRA16D." *Cancer letters* 232.1 (2006): 37-47.

Ozeri-Galai, Efrat, Assaf C. Bester, and Batsheva Kerem. "The complex basis underlying common fragile site instability in cancer." *Trends in Genetics* 28.6 (2012): 295-302.

Paciotti V, Clerici M, Lucchini G, Longhese MP. The checkpoint protein Ddc2,

functionally related to S. pombe Rad26, interacts with Mec1 and is regulated by Mec1-dependent phosphorylation in budding yeast. Genes & Dev. 2000;14:2046–2059.

Palakodeti, A., Han, Y., Jian, Y., & Le Beau, M. M. (2004). The role of late/slow replication of the FRA16D in common fragile site induction. Genes Chromosomes Cancer, 39, 71-76.

Palakodeti, Aparna, et al. "Impaired replication dynamics at the FRA3B common fragile site." *Human molecular genetics* 19.1 (2010): 99-110.

Richard, G., Kerrest, A., & Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiology and Molecular Biology Reviews, 72(4), 686-727.

Voineagu, I., Freudenreich, C. H., & Mirkin, S. M. (2009). Checkpoint responses to unusual structures formed by DNA repeats. Molecular Carcinogenesis, 48, 309-318.

Walsh, Erin, et al. "Mechanism of replicative DNA polymerase delta pausing and a potential role for DNA polymerase kappa in common fragile site replication." *Journal of molecular biology* 425.2 (2013): 232-243.

Ying, Songmin, et al. "MUS81 promotes common fragile site expression."*Nature cell biology* 15.8 (2013): 1001-1007.

Zhang, Haihua, and Catherine H. Freudenreich. “An AT-Rich Sequence in Human Common Fragile Site FRA16D Causes Fork Stalling and Chromosome Breakage in *S. Cerevisiae*.” *Molecular cell* 27.3 (2007): 367–379. *PMC*. Web. 26 Apr. 2015.

Zhao, Xiaolan, et al. "The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage." The EMBO Journal 20.13 (2001): 3544-3553.

Zlotorynski, Eitan, et al. "Molecular basis for expression of common and rare fragile sites." *Molecular and cellular biology* 23.20 (2003): 7143-7151.

Appendix

F5 Sequence as sequenced by the Eckert lab:
-28 uninterrupted polyAs
-$(AT)_{24}$ interrupted with AAA and TT
ATTTCTACTAAAAAAAAAAAAAAAAAAAAAAAAAAAAGTATATAAATATAT
ATATATATATATATATATATTATATATATATATATGTAGCCATGCATGGGGGC
CTGCCCTTGTAGTCCCAGCTACTTGGGAGGCTGAGACATGAGA

The F5+pBL007 constructs transformed into yeast strain CFY #2268 were sequenced via EtonBio using primers #1032 and #1033. The transformants were confirmed by sequencing through the 5' flanking sequence, AT24i repeats, and the 3' flanking sequence. The flanking region that was downstream of the repeats for that direction of synthesis always yielded poor quality (the polymerase likely stalled through the repeats). Sequencing results are shown in the table below:

Transformant	Primer #1032				Primer #1033			
	polyAs	AT repeats	5' flanking	3' flanking	polyAs	AT repeats	5' flanking	3' flanking
#520 T1 (orientation 2)	27 uninterrupted polyTs (could be 28; poor quality with A/T peaks at 5'end)	poor quality – T/A peaks together	TAAAGATGA	Quality too poor	26 uninterrupted polyAs	24 interrupted with AAA and TT	AGAGTACAG AGTCGGAGG GTTCATCGAC CCTGATGTTC CCGTCCGGGG GTACGTACCG ATG	Quality too poor
#520 T2 (orientation 2)	27 uninterrupted polyTs (could be 28; poor quality with A/T peaks at 5'end)	poor quality – T/A peaks together	TAAAGATGA	Quality too poor	27 interrupted polyAs with 1T (poor quality at 3'end)	22 interrupted with TT	AGAGTACAG AGTCGGAGG GTTCATCGAC CCTGATGTTC CCGTCCGGGG GTACGTACCG ATG	Quality too poor
#521 T1 (orientation 1)	27 uninterrupted polyAs (poor quality at 3'end)	24 interrupted with AAA and TT	AGAGTACAGA GTCGGAGGGT TCATCGACCC TGATGTTCCC GTCCGGGGGG TACGTACCG ATG	Quality too poor	27 uninterrupted polyTs (poor quality at 3'end)	Quality too poor	TAAAGATGA	Quality too poor
#521 T2 (orientation 1)	27 uninterrupted polyAs (poor quality at 3'end)	24 interrupted with AAA and TT	AGAGTACAGA GTCGGAGGGT TCATCGACCC TGATGTTCCC GTCCGGGGGG TACGTACCG ATG	Quality too poor	27 uninterrupted polyTs (could be 28; poor quality at 3'end)	Quality too poor	TAAAGATGA	Quality too poor

Homology to 5' flanking of F5 Eckert sequence
Homology to 3' flanking of F5 Eckert sequence

Transformants named "orientation 1" signifies that the human genome coordinates increase from the poly-As to the AT-repeats, in keeping with the F1 orientation 1 nomenclature. Conversely, "orientation 2" indicates that the human genome coordinates increase from the TA repeats to the poly-Ts. Orientation is named with respect to replication.

Deletion of regions within FRA16D (on YAC801B6)

Deletion of F1

Marker to replace F1: KANMX6
Template to amplify marker: pFA6a-KANMX6 (plasmid #136)
Primes to amplify marker with flanking 40bp homology to F1: #1236/1237 (homology highlighted blue)
Region of F1 deleted (red letters):

CTATCTAAAG TTGTCGGTGG TTGCTTCCTT CCTGAAAGCA ACAAGAAGCC 78608417
CATGAATGTC TTATGGCCTG AAGATCTGCC TTCAAAGACT ACTACCTGTG 78608467
TTGTCACAAA TGGCAAGATT TCATCCTTTT TATGGCTGAA TAATATTTCA 78608517
CTGTATATGT GAACCACATT TTTAAAAATG CATTTCTGAA CCTGAAGGAC 78608567
ATTATGTTAA GTCAAATAAG CCAGACACAG ACAGACAAAG ACATATGATC 78608617
TGACTAACAA TGTGGAATCT GAAAAAACCA AGCTCATAGA ATAGAAACAG 78608667
AGAGTGAGAA TGGTGGTTGC TGGGGCTAGG GGGTCAGTGA AGTGGGGAGG 78608717
TGTTGATCTA AGGGTTCAAA CTTCTCGTTA ATAATAAACC AGTTCTGGAG 78608767
ATCTAATGTA CAGCATGAGT GGTGATGGAT GTGTTAATTA ATTCGATTGT 78608817
GATAATCATT ACACAATGTA TATAGTAATC AAATCATTAC TTTATAGACC 78608867
CTGAATATAT TCAATATTTA TTTTTCAATT ATATATATAT ATATATATAT 78608917
AAATATATAT ATATATATAT ATATATATAT ATATATATAT ATATATATAT 78608967
ATATATATAT ATATATATAT ATATATTTAA AGCTGTCATG GAAAGCCTTA 78609017
AAGTTAAAAT ACGAAGATTT TTGAGAAAAA CTTTGCATAT TTTAATTGCT 78609067
GTCTGGAATC CTCCTTCAGC TGGGATGAGA AACCATCTCT GGGTTAGTTC 78609117
TGTCCCTGGA GGTGAGGGCC AGACAACACA TCAAAACTGA CTTTTTATTT 78609167
TTTTTCAAAG TTGTTAAGAC TATCTAAGAT TCAGATCTCC CTAATACTGT 78609217
TCCTAGTTAG TATTTAGAAC GTATTTGTTT AGAGGATACT AGCCTGAAAT 78609267
AGCCCTATCT CCAGGTTGTC TGATGCGATG AGGATGTGGT GTTAATCTTA 78609317
CACCTACCCA TTAGTCTATA AAACTCCTGA GGGATGCTCA GGAAACCAAA 78609367
AAGGATGTCT GCATGGAGGA CAAGAAGGCA CAAATCGTCT GATTTCACCC 78609417

*AT repeats highlighted yellow
*Numbers shown are hg19 coordinates from UCSC

Deletion of F5

Marker to replace F1: HIS3MX6

Template to amplify marker: pFA6a-HIS3MX6 (plasmid #138)

Primes to amplify marker with flanking 40bp homology to F5: #1448/1449 (homology highlighted blue)

Region of F5 deleted (red letters):

```
ggttgggagt cctagatcaa ggtgtcagca gggttggttt ctctctgtac  78652297
cattgtccct ggcttggaga tgaccatctt ctctgtcttc acatggtctt  78652347
CCCTCTGTGT GCGTTTGCAT CCTTATGGCC TCTTCTTGTG AGGGCACTGG  78652397
TCATATTGGA TTAGGGTCCA CCATATAATG ACTTTATTTT ACCTGAACTG  78652447
TCTCTTAAAA GGCTCTGTCT CCAAACACAG TCACATTGTG AGGTACTGGG  78652497
GGTTTGGGCT TCCACATACA AAGTCTGGGA CACAGTACGC TCTCATAAGA  78652547
GACCTCTTCT CAACAGGATC CATGTTTGAT CCGGGGGAAT AAAGCTAGAG  78652597
TTGCTTTCTT CCTTGTAATA TGTATAATGA CTCATGAGAG AATTCCCCAC  78652647
TGTCTTCCCC ACCTGAGACC TGAGAGGAGG AGAAGGAAGC TTGGAGTCGC  78652697
TGCCCTGTGG CCTGAATTGG TCAGTGACCA ATCACTCAGA CCCAGCTGAC  78652747
AGGCTATACG CGGACACCTT GGATGTGCCC AGCATGGATC TTCTCCAAGT  78652797
AGGAGGCCTT GGAGTCAGAG AAAGGTGCAA CTCCCACTGC TGCTGCAAGA  78652847
ACAAAGACTT TAAGATACTG AAATTTCTGG GCAAAGTTTC CTCAACTCTT  78652897
CATTTAAGTG TTAGGCCAGG TGGGGTGGCT CATGTCTGTA ATTCCAGCAC  78652947
TTTGGGAGGC TGAGGTGGGA GGATCACTGG AGGTCGGGAA TTTGAAACCA  78652997
GCCTGGCCAA CGTGGCAAAA CCCCATTTCT ACTAAAAAAA AAATATATAT  78653047
ATATATATAT ATATATATAT ATATATATAT ATATATATAT ATATATATAT  78653097
ATATATATAT ATATGTAGCC ATGCATGGGG GCCTGCCCTT GTAGTCCCAG  78653147
CTACTTGGGA GGCTGAGACA TGAGAATCGC TTGAACCTGG GAGTCGGAGG  78653197
TTGCAGTAAG CTGAGATCGC CACTGCACTC CAGCCTGGGC AACAGAGCAA  78653247
GACGCCATCT CAAAAAAAAA AAAAAAAAGT GTAATGCAAG TAGAAAAGGG  78653297
CACACACTCC AGGTGTATAT CTCAACAAAT TTGTACACAC TGCATGCACA  78653347
TGGGTGACCC GCACCCACAT GTTATACACA GAATAATTAC AGGACCCTGC  78653397
CAACCTCCTC CTATCCCCTT TCAGCTCCTG CTGTGATGCC ACTTCTAATC  78653447
TCAATTTTAT TATACAATTT GTTAAATACA GAAGTAAATG GAGGCAATAG  78653497
TGATAGTAAA AAGAACCATG TGAATTGGGC GTGGTGGCTC ATGTCTGTAA  78653547
AGCCAGCACT TCGGGAGTCC AAGGTAGGAG GGTCACTTGA GCCCAGCAAT  78653597
TGGAGACCAG CCTGGGCAAC ATCATGAGAC CCCATCTGTA TCCAAATAAA  78653647
AGCAACAAAA AAATGAGACA TGGGGGTGCA TGTCTGTGGT CCCAGCTACT  78653697
CAGGAGGCTG AGGCAGGAGA ATCACTTGAG CAAAGGAGGT TGAACCATGA  78653747
TTGCACCGTT GTACTCCATC TAGCCTGGGC GACAGAGTGA GAACCTGTCT  78653797
CCAAAATAAA TAAATAAATA AATAAAAATA AAAGTAACAG AACCATAGGC
```

*polyAs and AT repeats highlighted yellow

*Numbers shown are hg19 coordinates from UCSC

Deletion of P3P4

Marker to replace F1: HYG

Template to amplify marker: pAG32 (plasmid #241)

Primes to amplify marker with flanking 40bp homology to P3P4: #1555/1556 (homology highlighted blue)

Region of P3P4 deleted (red letters):

```
cagccccact ctttagcctc agtccctctg tggttcccct gactgagctc  78659190
ctcactaaag gggaggaatc cctggagcct ccccatccaa ggccagcgtg  78659240
TTCCATTTGA GGGTAACTGG ATCCCTGAAA TTCACCCCCA CACAATCCAC  78659290
AACCTACTAC TAGAATGCCA TATGCTTTTC TATATGCTGT TGTTTAATTT  78659340
CAGTTGGTAC ATATTTTATT TTTTGAAACA GACTTCGTTA GAACATCTTT  78659390
AGATTTAAAG AAAAATCAAG ACTGTAGTAC CAAGTTCCCA TGTGTGTAAC  78659440
AGCCAGTTTA TCCTATTATT TTAACGTCTT ATGTTATTTG GTACATTTAT  78659490
TAAGATTAAA TTATTGATAC TGAAATATTA GGTTGTGCAA AAGCAATTGT  78659540
GGTTTTTGCC ATTAAAAGTA ATAGTAATAA TAATACTCGG TTGGTGCAAA  78659590
AGTAAATGTG TTTTTTGCTA TTAAAAGTAA TGCGGTTTTA CTTCTGCAGT  78659640
TTTTTTGCCA TTTGCCAATT ACTTTTGCAG TTTTTGCCAG TACTTAAAAG  78659690
TAATGGCAAA AACTGCAAAA GTAATTGGCA AATGGCAATT GCGGTTTTTG  78659740
CCGTTACTTA AAATTGCATT ACTTAAAAGT TTTTAAGTAA TGCAATTACT  78659790
TAAAACTTAA GTTTTTAAGT AAGTTTAAGT AAAATTTGCA TTACTTAAAA  78659840
GTAATCGCAG TTTTTGCCAT TACTTAAAAG TAATGGCACC AACCTAATAT  78659890
TGTTATTACT TTTTACCAAC CTAATATTGT TATTAACTAA AGTCCATCGT  78659940
TTATCATATT GCCTTAGGTT TTTAAAAAAT TTTTTACCTA ATGACTTTTT  78659990
TGTGTTCCAG GAGCCCATCC AGGATACCAC TTTTATTTAT TTTTTTTTTC  78660040
TTTTTGAGAT GGAGTCTGGC TCTTTTGCCC AGGATGGAGT GCAGTGGCGC  78660090
GATCTTGGCT CACTGCAACC TCCGCCTCGC AGGTTTAAGT GATTCTCCTG  78660140
CCTCAGCCTC CTGAGCAGGT GGGACCACAG GCATGCGCCA CCACGCTCAG  78660190
ctaatttttg caggatacca cattttattt aattgtcagg tctccttagc  78660240
atcttcttgg ctgtgacagt ttctcagaat tgtctttgtg atgaccttga  78660290
```

*numbers shown are hg19 coordinates from UCSC