Determination of the Role of Irc20 in Preventing CAG Repeat Fragility

Oliver Takacsi-Nagy
Tufts University
Biology Senior Honors Thesis, 2018

Principal Investigator: Dr. Catherine Freudenreich
Research Mentor: Dr. Erica Polleys
Thesis Second Reader: Dr. Mitch McVey

ABSTRACT

CAG repeat sequences pose a formidable challenge to DNA replication, transcription and repair due to their propensity to form stable secondary structures. Accordingly, CAG repeat tracts, which are common throughout the human genome, have been deemed prone to chromosomal breakage (fragility) and changes in repeat copy number (instability). Given these characterizations, CAG repeats are causally associated with a host of neurodegenerative disorders. It is thus valuable to identify and investigate the specific mechanisms by which hallmark disease pathologies diseases arise molecularly. This study utilizes the model organism *S. cerevisiae*, which can be altered to harbor CAG repeat tracts in various genomic locations, to characterize the role of Irc20 in preventing CAG repeat fragility, and more generally DNA repair.

Irc20, previously categorized as an E3 ubiquitin ligase with chromatin remodeling capacity, was observed to prevent fragility using a YAC (yeast artificial chromosome) assay system. Subsequent epistatic analysis using this assay system yielded a synergistic relationship with Rad5, and function in the same pathway as Uls1. Moreover, Irc20 was shown to be necessary for efficient repair by sister chromatid recombination. These results point strongly towards a role in the Rad5-mediated error-free branch of post replication repair (PRR).

ACKNOWLEDGMENTS

This project would not have been possible without the help and support of many individuals. I would like to thank Dr. Catherine Freudenreich for giving me the opportunity to work in her lab; it was an honor and a privilege. Without the unconditional support of Dr. Erica Polleys, day in day out, night in night out, this project would not have gotten half as far. Dr. Mitch McVey offered his expertise and further ignited my scientific curiosity through two fascinating courses. I would also like to recognize the rest of the Freudenreich Lab members for guidance and support whenever I was in need. Finally, although they often encouraged me to work less, my family and friends helped keep me sane.

INTRODUCTION

I. Microsatellites and Human Disease

Although the high fidelity of DNA synthesis and repair machinery ensures faithful maintenance and replication of the genome, sequences that inherently pose complications for these processes do exist. Microsatellites compose approximately 2% of the human genome and are defined as regions of repetitive sequences that are highly polymorphic (i.e. varying number of repeat units comprise the tract) across individuals (Neil et. al. 2017). The trinucleotide repeat (TNR) is a common example of a repetitive sequence that exists in some microsatellite regions. Trinucleotide repeats have been identified as the drivers of a host of human neurodegenerative and developmental disorders coined the repeat expansion diseases (Neil et. al. 2017). This class encompasses Huntington disease, myotonic dystrophy, fragile X mental retardation and several spinocerebellar ataxias (Usdin et. al. 2015). These diseases and their individual variants have largely been traced to specific repeat tracts in the genome, and disease pathologies emerge when this tract exists above a threshold length, which can vary between diseases. For example, Huntington disease is caused by expansion (addition of TNR unit(s)) of a CAG-repeat tract within exon 1 of the *Huntingtin* gene above 29-37 repeats (McMurray 2010). Each CAG unit translates to the amino acid glutamine, so the repeat tract manifests as a poly-glutamine (poly-Q) protein region that is neurotoxic and prone to aggregation, resulting in neurodegeneration (Mirkin 2007). Expansions can occur during inheritance from a parent (intergenerational transmission) and additionally during an individual's lifetime (somatic instability) (Neil et. al. 2017). Repeat expansion diseases are marked by a genetic phenomenon known as anticipation, which accounts for earlier disease onset and increasing severity of symptoms with each successive

generation (Mirkin 2007). Biologically, this phenomenon is supported by evidence suggesting that longer repeat tracts expand more readily and are symptomatically more injurious (Usdin et. al. 2015).

II. Mechanisms for Fragility & Instability of TNR Tracts

Repetitive sequences, like CAG repeats, challenge normal replicative and repair processes. DNA generally exists in double-stranded, antiparallel form, meaning the CAG repeat is inter-strand base-paired with a CTG repeat. Both of these sequences are prone to intra-strand base-pairing, which manifests as a non-B form secondary structure known as a hairpin (Usdin et. al. 2015). In addition to the hairpin structure, CAG repeats can also form slightly offset hairpins opposite one another, termed a slip stranded structure (Usdin et. al. 2015).

As repeat instability is inherently dependent upon DNA synthesis, expansions and contractions can occur during replication or repair events, which almost always require synthesis to some extent. The leading model for repeat instability arising during replication is known as polymerase slippage. Essentially, the presence of a hairpin creates a misalignment between the template and nascent strands, and can be visualized as the replicative polymerase "slipping" and erroneously synthesizing (Usdin et. al. 2015). Okazaki fragments on the lagging strand have relatively more opportunity to form secondary structures than their leading strand counterpart, indicating that instability dominantly arises during lagging strand synthesis (Freudenreich et. al. 1997). Moreover, CTG hairpins are more stable, so if the lagging strand is synthesizing CAG (CTG hairpin on template), contractions occur frequently (Freudenreich et. al. 1997). Conversely, if the CTG hairpin forms on the lagging strand itself (CAG template), an

expansion is likely (Freudenreich et. al. 1997). A second possibility is the formation of a hairpin on the 5' end of an Okazaki fragment during lagging strand synthesis. The intra-strand base pairing "hides" previously synthesized DNA from the incoming polymerase and causes re-synthesis of the region encompassed by the hairpin, and an expansion (Usdin et. al. 2015).

The aberrant ligation of Okazaki fragments involves a DNA single-strand break (SSB), which are common in the context of lagging strand replication. Nicks or gaps (SSBs) are, however, also present in non-dividing cells (i.e. neurons) due to stressors such as oxidative damage (Mirkin 2007). By a similar mechanism, aberrant repair can lead to genomic instability and fragility.

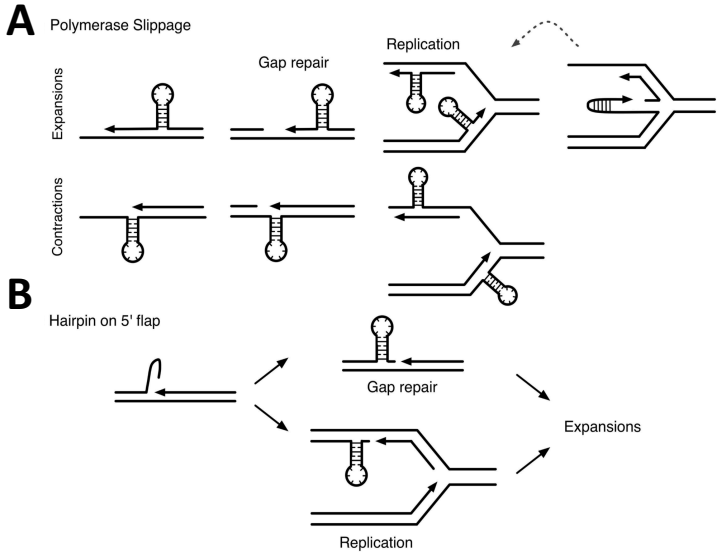


Figure 1A: Examples of mechanisms by which hairpin formation during DNA synthesis can contribute to instability via polymerase slippage. **B:** Nicks/gaps arise during lagging strand replication and SSB repair processes, which can also cause instability
(Usdin et. al. 2015)

CAG/CTG repeat tracts often comprise chromosomal regions prone to appearing as gaps or breaks (DSBs) on metaphase chromosomes (Freudenreich 2007). Thus, DNA double-strand break (DSB) repair pathways are important for preventing CAG repeat fragility and instability.

DNA DSBs in the context of CAG repeats can arise by a plethora of mechanisms. A stalled replication fork caused by a hairpin is prone to DSB formation (fork collapse) due to enhanced mechanical stress and nuclease activity, as well as by the cleavage of the Holliday junction structure formed by fork reversal and/or template switch (Freudenreich 2007). Moreover, a replication fork proceeding through a nick or gap left by incomplete or aberrant repair (i.e. MMR, BER, NER) of a lesion may lead to a DSB. (Freudenreich 2007). The research presented here highlights several specific mechanisms by which cells seek to repair DSBs (prevent fragility), and simultaneously maintain genomic integrity (prevent instability).

Generally, molecular machinery repairs DSBs, repetitive or otherwise, using two distinct mechanisms. The first, termed non-homologous end joining (NHEJ), ligates the two broken ends together and can result in insertions or deletions at the junction. Notable players in the dominant pathway include the stabilizing "Ku" proteins and DNA Ligase 4 (Krogh & Symington 2004) (Figure 2A). The second, termed homologous recombination (HR), utilizes a homologous region of DNA as a repair template. Generally considered to be an error-free form of DNA repair, HR is important in preventing CAG repeat breakage and instability (Usdin et. al. 2015). Although DNA damage signaling cascades may vary depending upon cell cycle stage, the MRX complex (Mre11, Rad50, Xrs2) is recruited to the breakage site to begin resection and allow further processing by helicase and endonuclease activity to form 3' overhang(s) (Krogh & Symington 2004). Next, extensive resection is carried out by helicases like Sgs1/Dna2 and the exonuclease Exo1 (Krogh & Symington 2004). The single stranded DNA (ssDNA) is bound by the RPA complex and subsequently, Rad52 mediates loading of the recombinase protein Rad51 forming a nucleofilament poised for homology search, strand invasion and eventual D-Loop

establishment (Krogh & Symington 2004). Once DNA synthesis using the homologous template has been completed, the extended invading strand can dissociate and anneal to its 3' end from the original DSB.

HR encompasses a diverse array of repair pathways that are directed by the severity of the break and available region of homology. Below is a brief outline of the different HR pathways that will be encountered in this paper:

- **Single-strand annealing** (SSA) (Figure 2B) is not dependent upon Rad51 or its paralogs, but does require the mediator protein Rad52 (Haber). This process involves resection to intra-chromosomal homology, and is finalized by degradation of the resultant ssDNA end and fill-in synthesis (Haber). As this resection can range up to approximately 50 kb, large scale deletions are possible (Haber).
- **Break-induced replication** (BIR) (Figure 2D) is traditionally employed at collapsed replication forks and in the ALT pathway of telomere extension, but more fundamentally in the context of a DSB, in which one side of the chromosome has been lost (Lydeard et. al. 2010). As this pathway acts on substrates marked by large-scale chromosomal loss, extensive DNA synthesis is required and classically requires the polymerase delta factor Pol32 and Pif1 helicase (Lydeard et. al. 2010).
- **Synthesis-Dependent Strand Annealing** (SDSA) (Figure 2C) follows the characteristic steps of HR pathways, but resolution occurs by unwinding of the inter-chromosomal base-pairing (dissolution). Although this prevents crossover events, synthesis from a non-exact homologous template (i.e. homologous chromosome) can result in a mismatch situation and loss of heterozygosity (LOH) by gene conversion.

- **Sister Chromatid Recombination** (SCR) events are characterized by their homologous template, rather than a unique pathway. As repair occurs using the sister chromatid, which is ideally perfectly homologous, SCR is a high-fidelity option and employed by 90% of G2 diploid cells in lieu of repair using a homologous chromosome (Kadyk & Hartwell 1992).

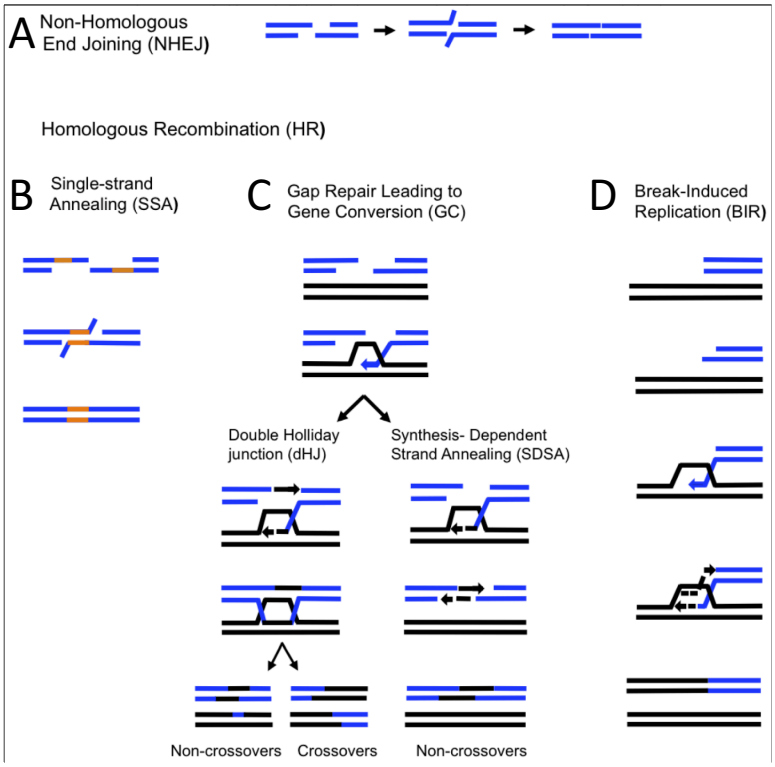


Figure 2A: NHEJ occurs by ligation of broken ends. **B:** SSA involves resection to homology and fill-in synthesis. **C:** SDSA involves formation of a D-loop and dissolution. **D:** BIR occurs when one side of the DSB is lost. (Sakofsky et. al. 2012)

The overarching takeaway from the discussion of HR repair pathways is the involvement of DNA resection and synthesis. Consider that if these processes occur within repeat tracts, the molecular machinery is subject to the same challenges introduced as part of normal replication (misalignment and slippage) and can thereby result in instability and fragility.

III. Saccharomyces Cerevisiae and TNR Systems

Although several decades of work have identified expanded TNR tracts as the root cause of various neurodegenerative diseases and ongoing research is focused on molecular mechanisms of repeat expansion or contraction (instability) and breakage (fragility), significant gaps remain in our knowledge. Due to the relative ease of genomic manipulation and the inherent speed of replication, the baker's yeast, *Saccharomyces cerevisiae,* is a powerful model for studying TNRs.

The Freudenreich Lab has pioneered several TNR-based systems, differing chiefly in the repeat tract length, genomic location and flanking selectable markers. Here, I will outline the constructs used in the presented research.

Yeast Artificial Chromosome system to study TNRs

Yeast artificial chromosomes (YACs) are constructs containing a centromere, origin of replication, telomeres and genes/sequences of interest, which allows propagation of the artificial chromosome alongside endogenous chromosomes from generation to generation. The original CAG_n-URA3 construct (YAC CF1) (Figure 3) was primarily designed to monitor fragility of increasing CAG repeat lengths, and has also been adapted for monitoring CAG repeat instability. CAG repeat tracts of various lengths have been inserted between a URA3 reporter gene and a $(G_4T_4)_{13}$ telomere seed sequence derived from *Oxytrica* (Figure 3). Also, a LEU2 selectable marker provides a direct way to ensure presence of the YAC in colonies of interest. If the YAC experiences chromosomal breakage within the repeat tract, short-range resection to the G_4T_4 sequence offers a template for *de novo* telomere addition. Along with the chromosome region distal to the break (end loss), the URA3 reporter gene is lost, allowing for

genetic scoring of this event by growth in the presence of 5-fluororotic acid (5-FOA) (Polleys & Freudenreich 2018). It is important to note that because YAC CF1 shares little homology to any endogenous chromosomes and the telomere seed sequence is located within short range of the repeat tract, repair by recombination is largely inhibited in favor of the more efficient telomere addition (Polleys & Freudenreich 2018). Moreover, as very few proteins of interest are involved in telomere extension, the fragility assay detailed in "Methods" largely measures increased breakage rather than decreased healing (Polleys & Freudenreich 2018). Lastly, when interpreting the output of the various fragility assays, consider that these fragility assays do not score healing resulting in instability (see "CAG repeat stability assays") or breakage that results in loss of the entire YAC due to improper telomere addition.

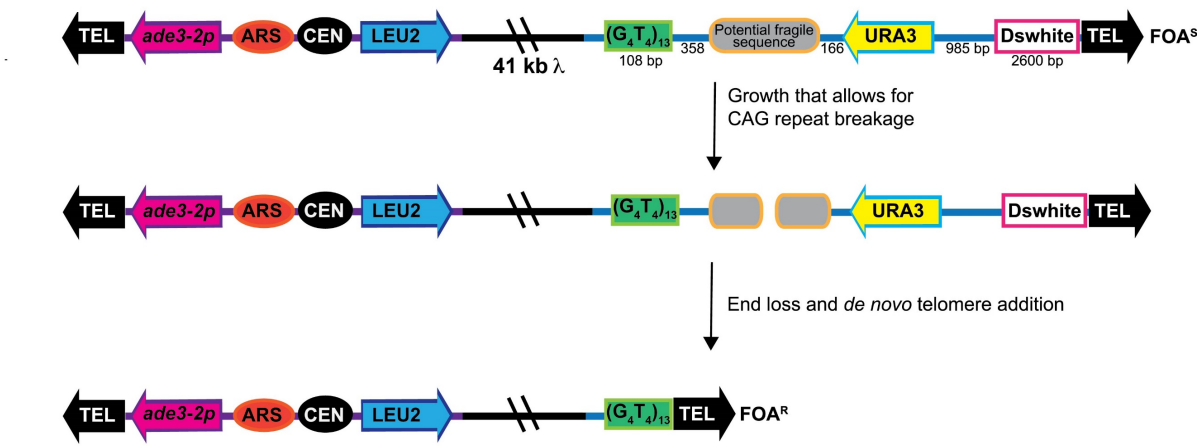


Figure 3: The original YAC CF1 construct. A potential fragile sequence (CAG repeat tract) is integrated between a URA3 reporter gene and telomere seed sequence. Upon breakage, resection to the seed sequence allows de novo telomere addition and FOA^R. Modifications include tract length/orientation change and integration of additional markers or transcriptional elements. (Polleys & Freudenreich 2018)

One of the major strengths of the YAC system is that additional markers or DNA elements (i.e. transcriptional enhancers or terminators) can be easily added. This allows testing in specific cellular conditions and monitoring resulting influence on fragility and instability. Much of the presented research was conducted using several variations of the general YAC CF1 introduced above:

- **ADE2 URA3 CAG-85 YAC** includes a second selectable marker gene (ADE2) to eliminate false fragility positives stemming from point mutations in the URA3 gene, which are undetectable using the original construct. Additionally, this marker is positioned transcriptionally opposite of the URA3 marker and both markers are distant from the repeat tract, effectively lowering repeat transcription, which will be explored in more detail later. An alternative version of this construct with galactose-inducible repeat tract transcription was also used. (Figures 5A and 7A) (Both developed by Elliot Philips)
- **2 Terminator (2T) CAG-70 YAC** includes transcriptional terminators that flank the repeat tract to decrease repeat transcription, which will be explored in more detail later. (Figure 6A) (Developed by Allen Su)

Additionally, the Freudenreich Lab utilizes two chromosomal systems to study the fate of TNRs. The LYS2 direct duplication recombination assay (LYS2 DDRA) measures breakage of a CAG repeat tract and subsequent resection and repair by SSA (Figure 12A). On the other hand, the fill-in synthesis system involves an induced break and subsequent resection through a CAG/CTG repeat tract that must be filled in by synthesis to preserve cell viability (Figure 12C).

IV. Irc20 and Post Replication Repair

S. cerevisiae protein Irc20 was identified by Olivia Familusi in a screen for CAG repeat fragility in Catherine Freudenreich's 2015 Biology 54 course. She found that upon deletion of IRC20, the mutant strain expressed increased levels of YAC end loss. Irc20 is a largely understudied protein, though it has been implicated in DNA repair pathways and is thought to be a member of the Rad5 family of proteins.

Post-translational modification is common across eukaryotes to regulate both the activity and abundance of a protein substrate. Ubiquitylation and SUMOylation (Small Ubiquitin-like MOdifier) are examples of post-translational modification that proceed via a characteristic enzymatic cascade, and most often mark proteins for proteosomal degradation. Given its RING domain, Irc20 is classified as an E3 ubiquitin ligase, the last enzyme of the conserved pathway, which facilitates the transfer of a ubiquitin monomer from an upstream E2 enzyme onto its final target (Richardson et. al. 2013). Moreover, Irc20 contains a Snf2/Swi2 ATPase/helicase domain that is common among chromatin remodelers (Richardson et. al. 2013).

In a genetic screen testing for transcription of a promotorless SUC2 gene, overexpression of Irc20 was able to promote expression of a SUC2 reporter gene lacking an upstream activating sequence (Richardson et. al. 2013). Transcription factors are common targets of ubquitylation and the assay employed has successfully identified several transcriptional regulators, suggesting a possible role for Irc20 in transcription (Richardson et. al. 2013). Creation of a host of domain mutants revealed that both the ATPase and RING domains were required for expression of the phenotype, but the link between the two activities is

unknown. Moreover, a Co-Immunoprecipitation experiment indicated a physical interaction with Cdc48, which is thought to act temporally between E3 ubiquitin ligase activity and proteosomal degradation (Richardson et. al. 2013). Finally, Irc20 contains a SIM (SUMO interacting motif) and showed in vitro interaction with SUMO raising the possibility of classification as a STUbL (SUMO-targeted ubiquitin Ligase) (Richardson et. al. 2013).

IRC20's role in DNA repair has been minimally characterized. Using a tri-parental system (one plasmid and two endogenous chromosomes) designed to score the frequency of SDSA and NHEJ events, Miura et. al. concluded that Irc20 functions in the same pathway as the helicase Srs2 in promoting repair by SDSA (2012). However, an *irc20*Δ mutant reduced Holliday junction-mediated HR, while a *srs2*Δ mutant did not. Srs2 is known to control D-loop migration, and SDSA and Holliday junction-mediated HR events differ only after D-loop formation (Figure 2C), so the group suggested that Irc20 functions upstream of Srs2 (Miura et. al. 2012). As upstream of D-loop formation in HR pathways conceivably indicates the resection process, epistasis analysis of Irc20 and the numerous helicases and nucleases implicated in resection were performed (Miura et. al. 2012). Deletion of IRC20 partially rescued the DNA damage hypersensitivity phenotype of a *mre11*Δ mutant (MMS, HU and CPT), leading the group to suggest that Irc20 may be involved in mediating the extent of resection, and thereby the subsequent repair pathway (Miura et. al. 2012). Considering that a tri-parental system involving a plasmid poorly recapitulates the native environment, and analysis of DNA damage sensitivities by growth on plates cannot discriminate between specific repair processes, these results do not indicate a clear pathway for Irc20 function.

Domain analysis of Irc20 has revealed that it shares homology to Rad5, Uls1 and Rad16 in *S. cerevisiae* and HLTF and SHPRH in humans (Unk et. al. 2010). All of Irc20's orthologs are known proteins involved in post replication repair (PRR) pathways. If a replicating polymerase encounters a lesion such as a hairpin on the template strand, the fork may stall. Subsequently, the Rad6-Rad18 complex binds RPA-coated ssDNA on the unwound template strand and mono-ubiquitylates PCNA, which serves to scaffold the replication machinery around DNA (Kramarz et. al. 2017). This modification acts as a signal for synthesis by trans-lesion synthesis (TLS) polymerases that feature accommodating active sites, yielding a high probability of mutagenic replication (Kramarz et. al. 2017). However, if the ubiquitin residue on PCNA is polymerized by the Mms2/Ubc13 heterodimer and E3 ubiquitin ligase Rad5, an error-free branch of PRR is activated (Kramarz et. al. 2017). Although the exact mechanism of Rad5 has yet to be elucidated, the leading hypothesis posits that the replication fork regresses such that the nascent strands anneal and synthesis can continue using this new template (template switch) until the region corresponding to the lesion is passed, at which point the fork can be restarted (Unk et. al. 2010). Another possibility is a SDSA event in which the 3' end of nascent DNA upstream of the lesion invades its sister chromatid to form a D-loop (Unk et. al. 2010). These processes are referred to as PRR as the lesion is not repaired by the template switch, but must be repaired later (i.e. post replication), and the template switch may occur after replication bypass of the lesion. Human homolog of Rad5, HLTF, has been shown to contain a conserved HIRAN domain capable of binding the 3' end of the nascent leading strand to facilitate fork regression (Kile et. al. 2015). HLTF and SHPRH share extensive homology and both can poly-ubquitylate PCNA, but the two are likely not redundant and regulate replication depending on

the specific nature of DNA damage (Lin et. al. 2011). Moreover, in mouse cells lacking both proteins, poly-ubiquitylation was still observed, indicating the presence of an additional E3 ubiquitin ligase and lending credence to the utility of the Rad5 homolog Irc20 in yeast (Kanao & Masutani 2017).

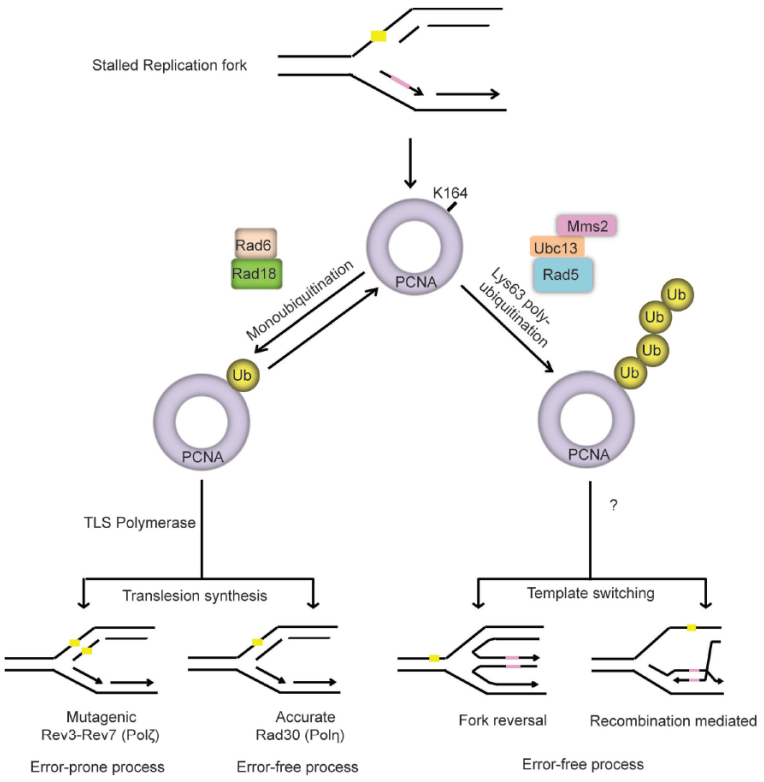


Figure 4: PRR repair pathways operating at a stalled fork. PCNA monoubiquitination leads to mutagenic TLS. Rad-5-mediated PCNA polyubiquitination leads to template switch mechanisms. (Ghosal & Chen 2013)

Clearly, the identification Irc20's role in transcription and DNA repair is far from complete. As *irc20* yielded a dramatic fragility phenotype, it is necessary to further elucidate the specific molecular processes to which Irc20 contributes and uncover how these processes may drive genomic instability and fragility.

METHODS

Yeast strain construction. The yeast strains used in this study are outlined in Table 1. Genes of interest were deleted via PCR-based gene replacement with selectable gene markers. Viable products of lithium acetate transformation were recovered on selective media; correct integration of the selectable marker was verified using PCR at the 3' and 5' integration junctions, primers internal to the target gene to verify ORF absence, and primers flanking repeat tracts to confirm repeat tract stability.

CAG repeat fragility assays. *For URA3 CAG_n & 2T YACs*: (Figure 6B) 7-10 single colonies from YC-Leu-Ura (to select for repeat tract) with verified $(CAG)_n$ tracts were inoculated in 2 ml YC-Leu to OD_{600} of 0.02 - 0.04 to select for the YAC. Cells were grown for 6-7 generations to an OD_{600} of 1.28 - 5.12, at 30°C with rotation. Cells were plated without dilution onto 5-fluororotic acid (5-FOA) plates and total viable cell count was measured by combining 100 μL of each culture (combined TCC) and plating 10^{-5} dilutions on YC-Leu in duplicate. Rate of FOA^R determined by maximum likelihood method (10^{-6}). Statistical significance determined by Student's t-test. ***For ADE2 URA3 CAG_{85} YAC*:** (Figure 5B and 18) 7-10 single colonies from YC-Leu-Ura with verified tracts* were inoculated in 5 ml YC-Leu to OD_{600} 0.1 – 0.3 to select for the YAC. Cells were grown for 24 hrs. to saturation, at 30°C with rotation. 1 ml of each saturated culture was re-suspended in 100 μL sterile H_2O and plated on 5-FOA Low Ade without serial dilution. Combined TCC serially diluted to 10^{-5} on YC-Leu in duplicate. Rate of FOA^R Ade^- (pink/red colonies only) determined by method of the median (10^{-8}). Statistical significance determined by Student's t-test. ***For pGAL CAG_{85} YAC*:** (Figure 7B) 7-10 single colonies from YC-Leu-Ura-Ade (Raffinose) with verified tracts* were inoculated in 2.5 mL YC-Leu-Ura to OD_{600} 0.1 – 0.3 to select for the YAC and repress transcription. After 1 doubling (8-12 hrs.), cultures were split into 1.25 ml aliquots, which were washed twice with sterile H_2O. The split cultures were inoculated in 1.25 ml YC-Leu + Galactose (induced transcription) and 1.25 ml YC-Leu + Glucose (repressed transcription) to saturation. Galactose and glucose cultures reached saturation at 40 and 24 hours from initial inoculation, respectively. Saturated colonies were plated as described for the ADE2 URA3 CAG_{85} YAC. Fragility rate and statistical significance determination also the same. In all cases, FOA plates and YC plates incubated at 30°C for 5 and 3 days, respectively.

CAG repeat stability assays. (Figure 8-10) Single colonies with verified $(CAG)_{85}$ tract lengths* were inoculated in 2 ml YC-Leu to OD_{600} of 0.02 - 0.04 to select for the YAC. Cells were grown for 6-7 generations to an OD_{600} of 1.28 - 5.12, at 30°C with rotation. A 10^{-4} dilution of cells was plated on YC-Leu-Ura to select for both arms of the YAC. The CAG tract lengths* of 33 daughter colonies per mother colony inoculated were tested by colony PCR. PCR products were run on a custom gel mix on a fragment analyzer (Advanced Analytical Technologies, Inc), and median repeat length was determined by overall frequency in the population. Expansions and contraction rates were calculated +/- 1 CAG or +/- 2 CAG repeats from the median tract length. The number of expansions and contractions were evaluated for significant deviation from wild-type cells using Fisher's Exact Test.

Sister chromatid recombination assays. (Figure 11A) Adapted from Mozlin et. al. (2008). Trp+ Ade- cells were resuspended in 5 ml YEPD and grown to saturation at 30°C (16-18 hrs.). Total viable cell count was measured by plating 10^{-5} dilutions on yeast complete (YC) media and recombinants were selected by plating 10^{-3} dilutions on YC-Trp-Ade. Incubated 3-4 days at 30°C. Recombination rates were calculated by the method of the median and rates were tested for statistical deviation using the Student's t-test.

LYS2 Direct duplication recombination assay. (Figure 12A) Verified CAG_{130} tract length* single colony from YC-Lys-Ura plate (if applicable). Plated 10^{-4} dilution of full colony onto YEPD plate and incubated for 3 days at 30°C. 10 single colonies of uniform size were picked from YEPD plates and resuspended in 400 μL sterile H_2O (OD_{600} ~0.1) and plated on 5-FOA-Lys. CAG_{130} strains were serially diluted to 10^{-2}, while CAG_0 strains were not diluted. 100 μL from each re-suspension was combined and diluted to 10^{-5} for total viable cell count on YC. FOA plates and YC plates incubated at 30°C for 5 and 3 days, respectively. Rate of FOA^R calculated by method of the median (10^{-5}). Statistical significance determined by Student's t-test.

Gap fill-in/Single strand annealing assays. (Figure 12C) Modified from Vaze et. al. (2002). Strains were plated for single colonies on YPD. Strains containing CAG_{70} or CTG_{70} repeat tract, were first tract length* confirmed. Next, colonies were resuspended in 3 ml YPLactate and grown for 2-3 divisions (16-18 hrs) at 30 °C with rotation. Cultures were then serially diluted to 10^{-4} and plated in duplicate on YEPD and YEPGalactose. Percent viability was calculated as number of colonies on YEPGalactose/by the number of colonies on YEPD. For strains containing a CAG_{70} or CTG_{70} repeat tract, 22 daughter colonies from both YEPD and YEPGalactose were subjected to tract length* PCR and run out on a fragment analyzer. Incubated 3-4 days at 30°C.

Break-induced recombination assays. (Figure 13A) Assays were modified from Anand et. al. (2014). Briefly, colonies on YEPD+Nourseothricin were picked, serially diluted to 10^{-4} and plated on YEPD and YEPGalactose in duplicate. Colonies were counted and percent viability determined (number of colonies on YEPGalactose/by the number of colonies on YEPD). To determine the frequency of BIR and other types of repair, every colony on YPGalactose was pinned onto YEPD, YEPD+Nourseothricin, and YC-URA. BIR rate was calculated as number of URA+ NAT- cells/number of colonies on YEPD. NHEJ rate was calculated as number of Ura-/NAT+ cells/number of colonies on YEPD. Incubated 3-4 days at 30°C.

Ectopic gene conversion assays. (Figure 14A) Assays were performed as previously described (Keogh et. al. 2006). Briefly, colonies on YPD were resuspended in 3 ml YPLactate and grown for 2-3 divisions at 30°C with rotation. Cultures were then serially diluted to 10^{-4} and plated in duplicate on YEPD and YEPGalactose. Gene conversion rates were calculated as percent viability (number of colonies on YEPGalactose/by the number of colonies on YEPD). Incubated 3-4 days at 30°C.

Nuclear Localization assays. (Figure 15) Done as previously described (Su et. al. 2015). Briefly, colonies were CAG_{130} tract length* confirmed and grown overnight in YC (2 at a time/mutant). In the morning, cells were appropriately diluted into 10 ml of YC and grown for 2 divisions

(about 4 hours) to a density below 5 x 10^6 cells/ml, determined using a hemocytometer. Cells were fixed using 4% paraformaldehyde. 30 Z-stack pictures at 0.15 μm intervals were captured under 100x magnification using a Zeiss AX10 fluorescent microscope and Slidebook software. Exposure time was DIC: 50 ms; GFP: 300 ms. Images were deconvolved, and three-zoning criteria was used to evaluate the location of GFP foci for S phase cells using ImageJ point picker plugin (Meister et. al. 2010). Cell morphology determined as described in (Hediger et. al. 2004). Statistical significance by Fisher's exact test.

Telomeric Southern Blotting. (Figure 16) Done as previously described (Lendvay et. al. 1996) . Briefly, cells were grown overnight to saturation in 10 mL YPD. Genomic DNA was isolated and digested overnight with *Xho*I at 37°C. Samples were applied to a 0.8% agarose gel and electrophoresed at 35 V for 28 h. Following electrophoresis, the gel was imaged and transferred to a nylon membrane using upward denaturing transfer conditions. Following transfer, the membrane was pre-hybed for 2 hrs at 65°C in Church buffer. A telomeric restriction fragment from plasmid AP135 (a gift from the Bertuch Lab) was radioactively labelled with ^{32}P using a random labelling kit (Roche Megaprime Labelling kit). The blot was hybridized at 65°C overnight with the labelled probe in Church's buffer. Finally, the blot was washed twice in a low stringency wash (2X SSC + 0.1%) for 15 minutes each and twice with a high stringency wash (0.1X SSC +0.1%SDS) for 10 minutes each. Blot was exposed to film for 2 days prior to imaging.

DNA Damage plating. (Figure 17) To assess fitness under conditions of DNA damage, ten-fold serial dilutions of saturated liquid cultures were plated on YPD, YPD + 0.5 μM phleomycin, YPD + 1 μM phleomycin, YPD + 200 mM hydroxyurea, YPD + 0.035% MMS, and YPD + 20 μM camptothecin. Plates were grown at 30°C for 3-5 days with the exception of one YPD plate which was incubated at 37°C.

*Tract length confirmation: PCR reagents, FastTaq program and Metaphor high resolution gel/Fragment Analyzer as described in (Polleys & Freudenreich 2018). Primers for all YACs: New CAG For/Rev, except T720B/CTG Rev 2 for 2T YAC. For nuclear localization and LYS2 DDRA strains: T720/CTG Rev 2. For fill-in strain CAG orientation: New CAG For/T720. For fill-in strain CTG orientation: 2117/2118.

Strain Number	Strain Background	Genotype
CFY1162/766	BY4705	YAC: LEU2, URA3, CAG-70
CFY3580	BY4705	YAC: LEU2, Ttef1-CAG-70-Tcyc1-URA3
CFY4466	BY4705	CFY3580 irc20::KanMX6
CFY4064/4065	BY4705	CFY766 irc20::KanMX6
CFY810	BY4705	YAC: LEU2, URA3, CAG-85
CFY4386/4387	BY4705	CFY810 irc20::KanMX6
CFY3237/3238	VPS105	YAC: LEU2, ADE2, URA3 CAG-10
CFY 4007/4008	VPS105	CFY3238 irc20::KanMX6
CFY3273/3274	VPS105	YAC: LEU2, ADE2, URA3 CAG-85
CFY3931/3973	VPS105	CFY3274 irc20::KanMX6
CFY653	VPS105	YAC: LEU2, URA3 CAG-85
CFY4398/4399	VPS105	CFY653 irc20::KanMX6
CFY4536	VPS105	CFY3931 rad5::TRP1
CFY4608/4609	VPS105	CFY3274 uls1::NAT
CFY4610	VPS105	CFY3931 uls1::NAT
CFY4606/4607	VPS105	CFY3931 rad16::TRP1
CFY3340	VPS105	YAC: LEU2, ADE2, URA3 CAG-85 pGAL (CAG tx)
CFY4602	VPS105	CFY3340 irc20::KanMX6
CFY3343	VPS105	YAC: LEU2, ADE2, URA3 CAG-85 pGAL (CUG tx)
CFY4603/4604	VPS105	CFY3343 irc20::KanMX6
CFY3878	S288c	BIR yRA52 (Anand et. al. 2014)
CFY4508/4509	S288c	irc20::KanMX6
CFY4444	YJK17	Ectopic GC (Vaze et. al. 2002)
CFY4464	YJK17	CFY4444 irc20::KanMX6
CFY2867	W303	SCR LSY1519-1D (Mozlin et. al. 2008)
CFY4155/4166	W303	CFY2867 irc20::KanMX6
CFY2744	W303	LacO-ARS607-CAG-130
CFY4461/4462	W303	CFY2744 irc20::KanMX6
CFY3647	YMV80	CFY3596 ilv6Δ::CTG70-HPH
CFY4464	YMV80	CFY3647 irc20::KanMX6
CFY4018	YMV80	CFY3596 ilv6Δ::CAG70- HPH
CFY4465	YMV80	CFY4018 irc20::KanMX6
CFY3682	YMV80	CFY3596 ilv6Δ::HPH
CFY4026	YMV80	CFY3682 irc20::KanMX6
CFY104	YPH500	LYS2 DDRA CAG-0 (Freudenreich et. al. 1998)
CFY4467/4468	YPH500	CFY104 irc20::KanMX6
CFY109	YPH500	LYS2 DDRA CAG-130 (Freudenreich et. al. 1998)
CFY4502	YPH500	CFY109 irc20::KanMX6

Table 1. Strains used in this study.

RESULTS

I. Deletion of IRC20 results in increased CAG repeat fragility

In a genetic screen of non-essential genes containing the CAG85 ADE2 URA3 YAC (Figure 5A), it was discovered that deletion of *IRC20* increased FOA^R, our measure of chromosomal fragility. Since IRC20 has been previously linked to DNA repair pathways, the increase in fragility warranted further confirmation and investigation into how Irc20 prevents CAG repeat breakage. CAG repeats are known to inhibit faithful replication or repair (e.g. fork stalls, nicks, gaps, DSBs), resulting in chromosomal instability or fragility at a relatively high frequency (Usdin et. al. 2015). As TNR fragility and instability inherently arise under similar circumstances, we sought to characterize the influence of IRC20 deletion on CAG repeat instability and fragility.

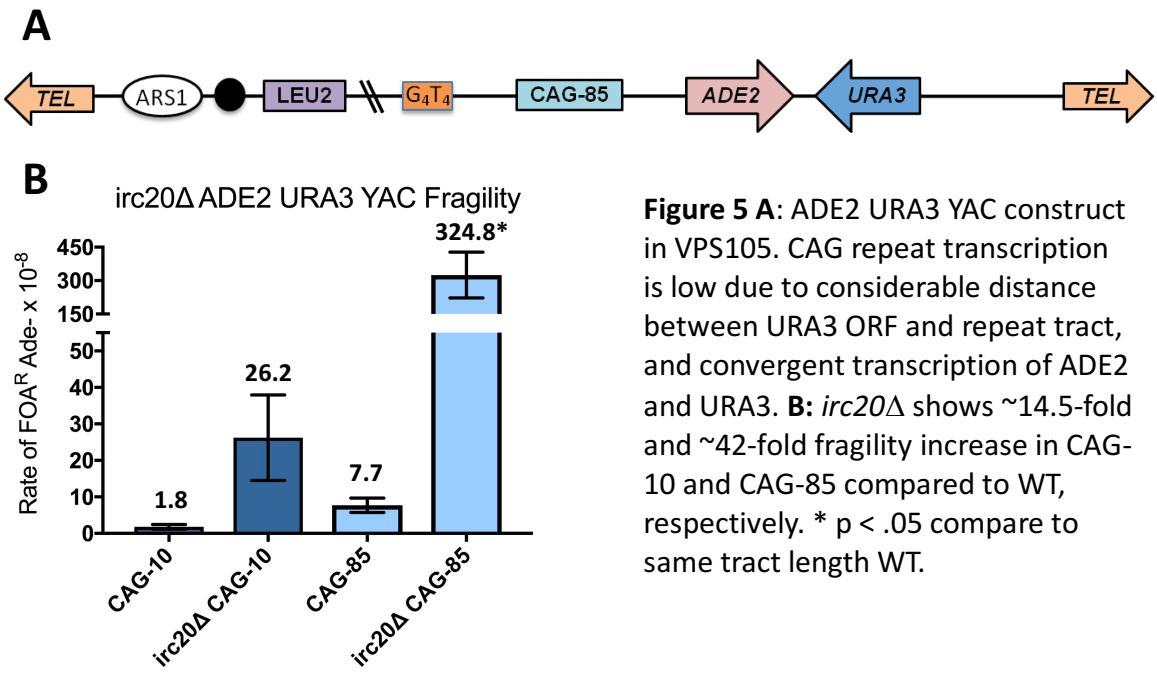


Figure 5 A: ADE2 URA3 YAC construct in VPS105. CAG repeat transcription is low due to considerable distance between URA3 ORF and repeat tract, and convergent transcription of ADE2 and URA3. **B:** *irc20*Δ shows ~14.5-fold and ~42-fold fragility increase in CAG-10 and CAG-85 compared to WT, respectively. * p < .05 compare to same tract length WT.

The dominant fragility phenotype of *irc20*Δ has been documented using the ADE2 URA3 YAC (Figure 5A). A ~14.5-fold increase in end loss compared to wildtype is exhibited in the CAG-

10 construct, and this ratio increases to ~42-fold in the CAG-85 construct, indicating a tract-length dependent increase in fragility (Figure 5B). Interestingly, such a dramatic increase in fragility has not been observed when IRC20 was deleted in other genetic backgrounds, which harbor YACs of various CAG repeat lengths (Figure 6B). We compared the same YAC construct (CAG-85 URA3) in two genetic backgrounds (BY4705 vs. VPS105) and observed no difference in CAG repeat fragility relative to wildtype (Figure 6A and B). Additionally, assays performed using a shorter CAG-70 tract length in the BY4705 genetic background yielded no change in CAG repeat fragility relative to wildtype. Finally, a modest increase in fragility was observed in a YAC construct wherein the CAG-70 repeat tract is integrated in between two transcriptional terminators (CAG-70 2T, Figure 6A bottom). This variability in fragility phenotype is interesting, as this discrepancy does not typically arise among other mutants studied in the Freudenreich Lab. Therefore, we questioned whether the fragility phenotype is an artifact of the ADE2 URA3 YAC itself, rather than the deletion of *IRC20*. More simply, why is fragility only observable in a singular system?

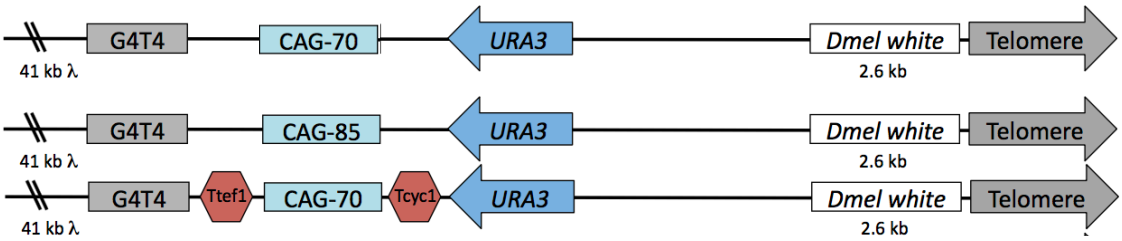


Figure 6A: Overview of the YAC constructs wherein CAG repeat breakage results in FOA^R colonies. Top: CAG-70 URA3. BY4705 genetic background. Middle: CAG-85 URA3. BY4705 and VPS105 genetic backgrounds. Bottom: CAG-70 2T. Low repeat transcription. BY4705 genetic background.

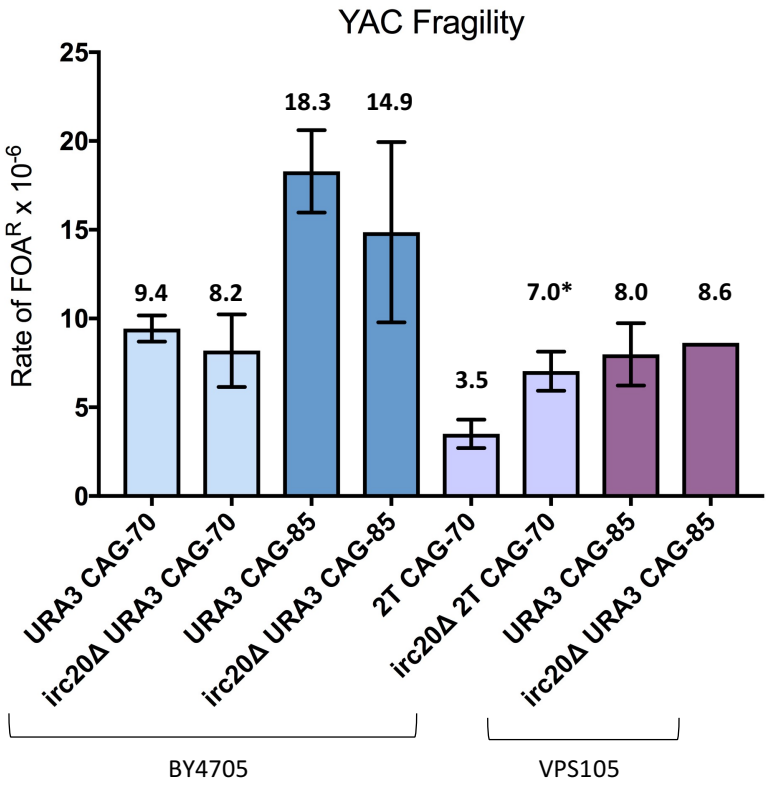


Figure 6B: *irc20*Δ fragility rate matches reported wildtype values across multiple strain backgrounds (bracketed), repeat tract lengths (CAG-70 vs. CAG-85) and constructs (URA3 vs. 2T). * $p < .05$ compare to WT same background, construct, tract length.

The fragility assays featuring various YAC constructs monitor breakage events occurring during logarithmic growth of liquid cultured cells, so DNA replication and repair are most commonly implicated as processes driving chromosomal breakage. However, given that the YAC is maintained by endogenous cellular machinery, its genes are subject to transcription, which is another process known to cause CAG-repeat fragility (Usdin et. al. 2015). In the original YAC CF1 construct introduced earlier, a URA3 reporter gene is situated about 200 base pairs (bp) downstream of the repeat tract, but transcriptionally oriented towards the repeat (Figure 6A top). On the other hand, in the ADE2 URA3 YAC, the ADE2 reporter gene was placed between the repeat tract and URA3 reporter (Figure 5A). The ADE2 reporter is oriented to oppose URA3 transcription, and engineering of the construct resulted in 2.8 kilo base pairs (kb) between the URA3 reporter and repeat tract (Figure 5A). Moreover, in order to inhibit transcription through

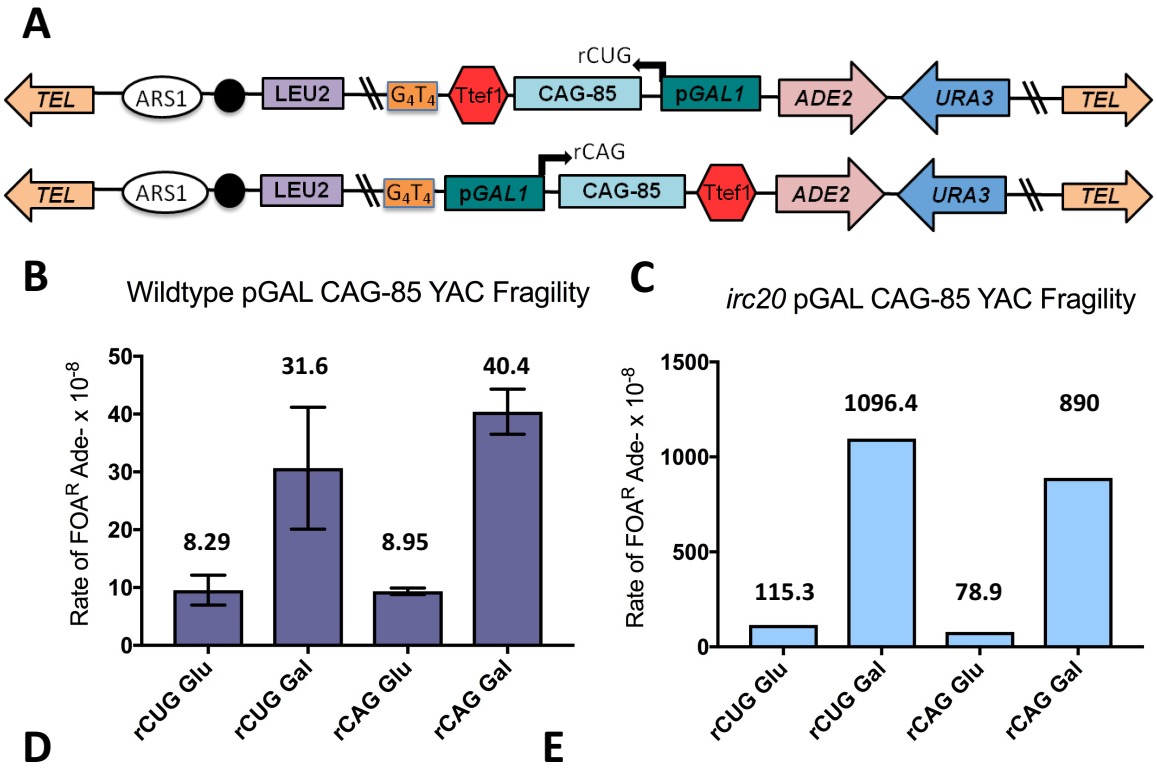
the repeat tract, a construct coined the 2-Terminator (2T) YAC was developed (Figure 6A bottom).

Previous work monitoring transcription through the repeat using qRT-PCR showed a 60-80% and ~70% decrease in CAG repeat transcript in the ADE2 URA3 and 2T YACs, respectively (Koch et. al. 2018). Initially, we hypothesized that the *irc20*Δ fragility phenotype was heightened when repeat transcription was minimal (i.e. ADE2 URA3 YAC). This hypothesis was somewhat supported by 2-fold increase in fragility observed in the 2T YAC, where there is also minimal repeat transcription (Figure 6B).

Given inconclusive evidence for a "low-transcription, high-fragility" hypothesis, we decided to delete IRC20 in a strain harboring a YAC, which allows transcriptional upregulation. In this construct, the ADE2 URA3 YAC construct features galactose-inducible repeat transcription (pGAL CAG-85) (Figure 7A). In the absence of glucose as a carbon source, yeast is capable of metabolizing galactose, for which an enzyme coded by GAL1, is required. Thus, by utilizing a GAL1 promoter adjacent to the repeat tract, in the presence of galactose and absence of glucose, it is possible to induce transcription through the repeat (Koch et. al. 2018). Transcription through the repeat, in either orientation, is up to 20-fold greater under galactose-induced conditions when compared to non-induced glucose conditions (Koch et. al. 2018).

Assuming *irc20*Δ only yielded a fragility phenotype given minimal repeat transcription, we reasoned that fragility rates would match wildtype values when assaying with the pGAL CAG-85 YAC (i.e. high transcription). This was, however, not the case. If the transcription-associated fragility in this system is defined as fold-difference in FOA^R Ade^- under induction versus not (i.e. rCUG Gal/Glu), preliminary data suggests that *irc20*Δ exhibits a fragility

phenotype that is ~2.5-fold greater than wildtype in the rCAG and rCUG orientations (Figure 7B, C and D). This indicates that increasing transcription has a greater effect in an *irc20*Δ mutant than in wildtype.



D

	Wildtype	irc20Δ
rCUG (Gal/Glu)	3.6	9.5
rCAG (Gal/Glu)	4.6	11.3
rCUG Fold Over WT		**2.6**
rCAG Fold Over WT		**2.5**

E

	Wildtype ADE2 URA3	irc20Δ pGAL rCUG (Glu)	irc20Δ pGAL rCAG (Glu)
Mean Rate ($x10^{-8}$)	7.7	115.3	78.9
Fold Over WT	-	**15**	**10.2**

Figure 7A: Schematics of the ADE2 URA3 galactose-inducible (pGAL) YAC in both transcriptional orientations (VPS105 strain background). Ttef1 transcriptional terminator placement ensures efficient detection of repeat transcripts by qRT-PCR. Optimal induction in galactose allows up to 20-fold increase in transcription compared to glucose conditions. **B:** Wildtype fragility of pGAL YACs with respect to orientation and carbon source. **C:** *irc20*Δ fragility of pGAL YACs with respect to orientation and carbon source. **D:** Outline of transcription-dependent fragility. **E:** Comparison of transcription-independent fragility between constructs. Note, Y-axis for **B** and **C** are on different scales.

Although further replicates need to be performed and wildtype rates solidified, this data likely confirms that the fragility phenotype of *irc20*Δ is driven by a molecular scenario more nuanced than high or low transcription. Moreover, this data offered an important confirmation of the drastic fragility phenotype observed in the ADE2 URA3 YAC assay system. If comparing the fragility rate of the *irc20*Δ mutant under non-induced ("Glu") conditions to wildtype rate in the ADE2 URA3 construct, we observe a ~15-fold and ~10-fold increase in fragility in the rCUG and rCAG orientations, respectively (Figure 7E). These increases are not nearly as pronounced as the ~42-fold increase in the ADE2 URA3 YAC, but partial recapitulation of the phenotype in an analogous system (i.e. ADE2 URA3 non-induced transcription), that was independently constructed, lends support to the original result.

II. Deletion of IRC20 does not impact CAG repeat instability

Instability is defined as the propensity of the CAG repeat tract to gain (expand) or lose (contract) repeat units. As mentioned earlier, instability often arises during replication due to hairpin formation on either the template or non-template strand and resultant misalignment of the complementary strands.

Figure 8: 11 PCR products from AATI Inc.'s Fragment Analyzer. Read out includes product length specific to the base-pair. Arrow indicates amplification of a stable CAG_{70} repeat tract. "E" denotes an expansion, and "C" a contraction.

Mutants expressing a fragility phenotype have previously been shown to also exhibit an increase in repeat instability. Thus, we set out to determine whether deletion of IRC20 influenced stability of the repeat tract. To determine tract length changes occurring during replication or repair of spontaneous breaks within the repeat tract, a colony of confirmed tract length (i.e. stable at onset) is grown for a set number of divisions and plated while still in the logarithmic phase of growth. PCR across the repetitive tract (see "Methods") is performed on a population of daughter colonies and products are run on a fragment analyzer (Figure 8). As the system outputs specific base-pair lengths of the PCR products, and the vast majority of tract lengths remain unchanged, it is possible to construct a standard curve of instability and make comparisons between distributions of wildtype and mutant (Figure 9). In the ADE2 URA3 CAG-85 YAC, the wildtype median tract length is 484-486 bp. Based upon the distribution, it was determined that one repeat (3 bp) change in either direction is considered unchanged as it is within the error of the sizing system (481-489), but a length above or below this range is classified as an expansion or contraction, respectively (Figure 9). Mutants exhibiting an instability phenotype can be visualized by either a shift in their median range relative to wildtype (Figure 9) or increased incidence of expansions and/or contractions (Figure 10). As illustrated in Figure 9, *irc20*Δ's instability profile matches wildtype remarkably well. We were curious to see whether an *irc20*Δ instability phenotype arises if the cut-off parameters were changed. Accordingly, frequency of expansions and contractions with +/- 1 CAG repeat and +/- 2 CAG repeats from the median tract length as the cut-offs were considered (Figure 10). In both cases, other than a slight increase in contractions with +/- 1 CAG as the cut-off, *irc20*Δ exhibited

wildtype instability (Figure 10). These data indicate that an *irc20*Δ mutant does not express an instability phenotype.

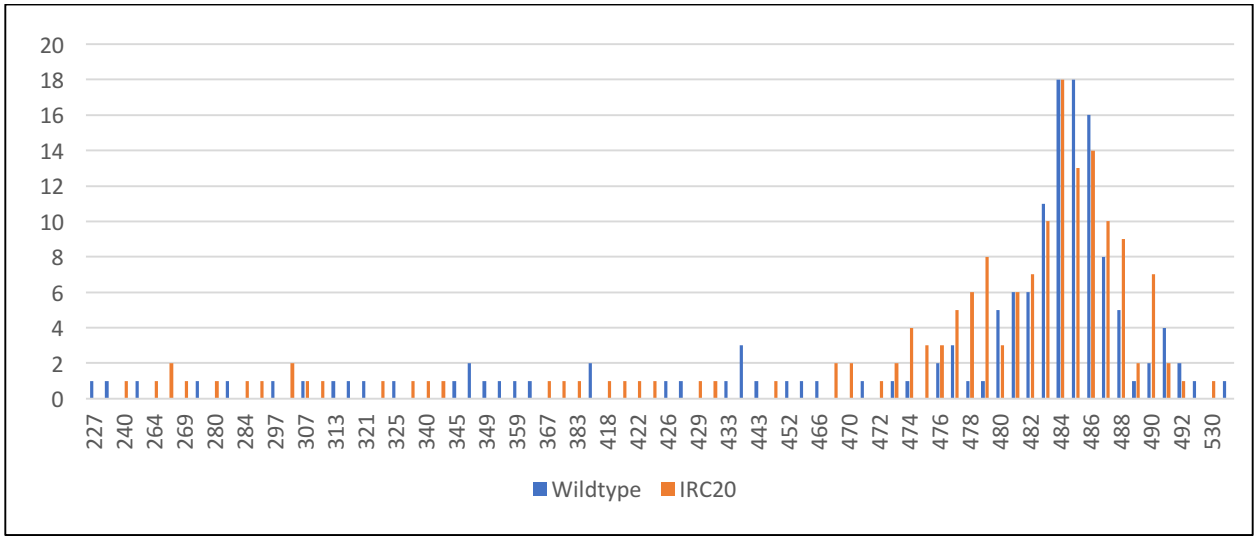


Figure 9: Representation of the frequency of each tract length, as amplified by PCR from daughter colonies of instability assay. Overlaying of data allows visualization of a shift in a mutant's instability profile relative to wildtype. Here, *irc20*Δ shows distribution nearly identical to wildtype. 165 and 144 reactions for *irc20*Δ and WT, respectively.

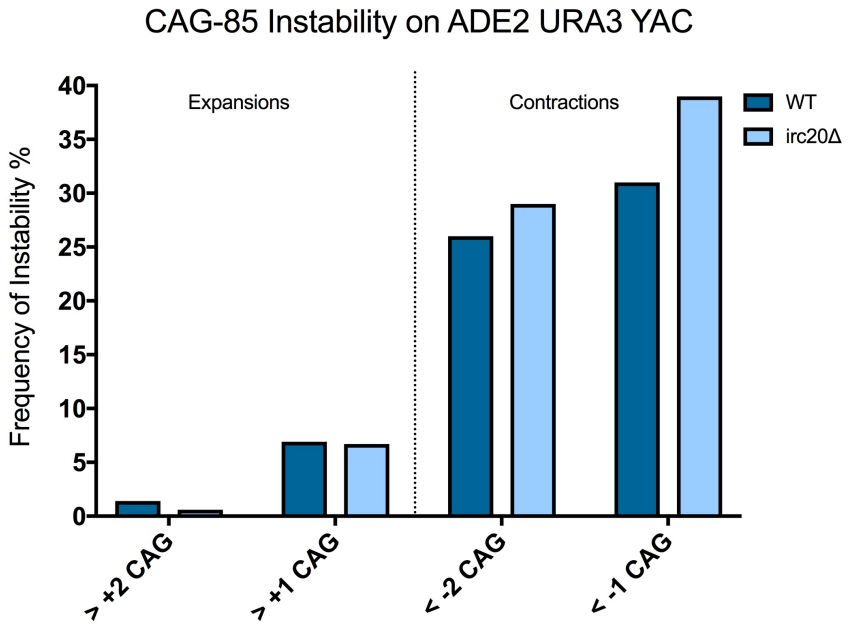


Figure 10: The wildtype median range was determined to be 484-486 bp. Instability was calculated using two different cut-offs: +/- 1 CAG- & +/- 2 CAG-repeat(s), where 1 CAG = 3 bp. Thus, for +/- 1 CAG, any reactions amplifying above 489 bp were considered expansions and below 481 were considered contractions. Same logic applied for +/- 2 CAG; *irc20*Δ did not yield significant instability in either scenario.

III. Deletion of IRC20 decreases Sister Chromatid Recombination

As sister chromatids arise via replication in S phase, cells are granted a genetically identical template for repair processes, which is generally more favorable than a homologous template due to its proximity and higher likelihood of facilitating error-free repair (Kadyk & Hartwell 1992). Although discrimination between a sister chromatid and homologous chromosome cannot be assessed in haploid cells (the assays presented in this paper), G2 haploid cells repairing via sister chromatid show higher viability than G1 cells, where this option does not exist (Haber).

Given that sister chromatids are genetically identical, it is inherently difficult to measure the rate of the sister chromatid repair pathway. Thus, the Symington Lab developed a strain expressing two mutant alleles of a gene encoding a specific selectable marker (ADE2), separated by an intact gene for another selectable marker (TRP1) (Mozlin et. al. 2008) (Figure 11A). If the mutant alleles of the ADE2 genes (ade2-n & ade2-i) spontaneously break, it is possible, via several repair pathways to restore the function of the fragmented, non-functional gene, and using selective media, identify recombinants (Mozlin et. al. 2008) (Figure 11A). SCR recombinants are scored as ADE^+ TRP^+ meaning that resection and annealing (SSA), and accordant loss of the TRP1 marker is ruled out. It is hypothesized that sister chromatid repair occurs primarily via a short-tract non-crossover pathway (SDSA) (Jain et. al. 2009). Note that

spontaneous breakage and strand invasion could occur from either ADE2 locus, but only one scenario is depicted in the schematic (Figure 11A).

Upon deletion of *IRC20*, SCR rate was decreased ~2.5-fold (Figure 11B). A pathway similar to SDSA is thought to contribute to sister chromatid repair, and Miura et. al. reported a ~4-fold decrease in SDSA efficiency in *irc20*Δ mutants, suggesting that Irc20 may be involved in non-crossover recombination pathways (2012). SCR has previously been linked to repeat instability and fragility, so this result potentially identifies a molecular mechanism by which the *irc20*Δ fragility phenotypes arise (House, Polleys, Freudenreich et. al. unpublished, House et. al. 2014). Specifically, a decrease in repair of spontaneous breaks within the repeat tract by SCR could lead to an increase in YAC end loss.

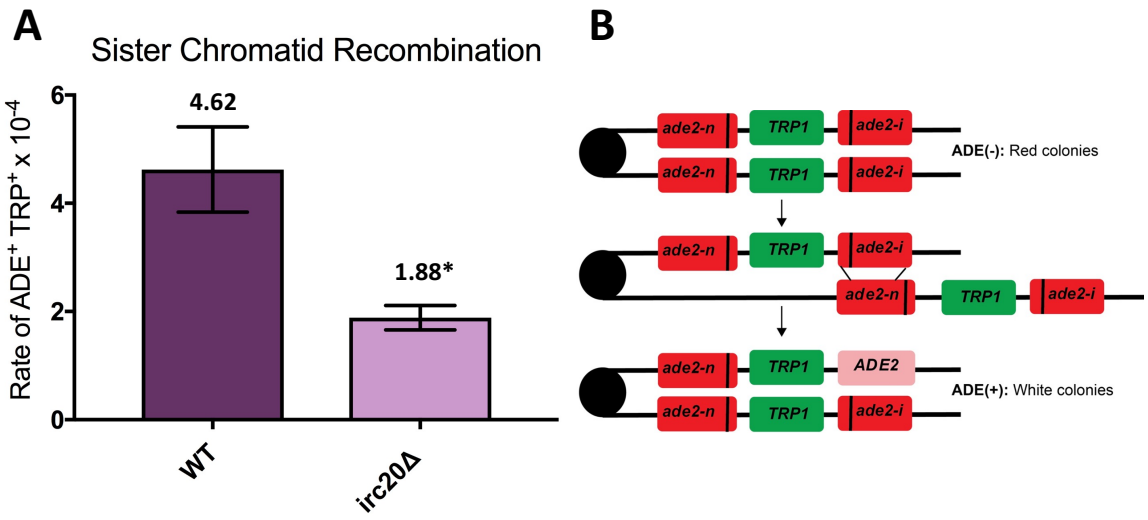


Figure 11A: Schematic of a possible repair mechanism scored in this assay. Misalignment allows invasion from ade2-i, synthesis from ade2-n sister chromatid template, and restoration of ADE2 function (Mozlin et. al. 2008). **B:** *irc20*Δ exhibits a ~2.5-fold reduction in SCR efficiency, which parallels a reported role in SDSA. * $p < .05$ compare to WT.

IV. Irc20 does not play a role in resection of a DSB

Irc20 was tested for a role in fragility of the CAG repeat wherein a break is repaired via the DSB-repair pathway SSA. Two separate assays were employed, as the context of the repeat tract differed in both cases. The LYS2 direct duplication recombination assay (LYS2 DDRA) features a repeat tract integrated into endogenous chromosome II adjacent to a URA3 reporter gene and in between an intact LYS2 marker on the 5' end and a fragmented LYS2 gene on the 3' end (Polleys & Freudenreich 2018) (Freudenreich et. al. 1998) (Figure 12A). Upon breakage within or close to the repeat tract, resection to the homologous LYS2 genes causes FOA^R via loss of the URA3 reporter and preservation of $LYS2^+$ status (Polleys & Freudenreich 2018). Deletion of *IRC20* did not cause an increase in fragility in the no repeat control (CAG-0) or CAG-130 construct, suggesting lack of a role in SSA (Figure 12B).

On the other hand, the fill-in synthesis assay utilizes a galactose-inducible HO endonuclease cut site within the intact LEU2 gene (Figure 12C). Previous work utilizing this system, which originally did not contain a repeat tract, was instrumental in identifying factors essential for break end processing (MRX complex, Sae2, Sgs1, Srs2) and Rad52, all of which are specifically important for this type of DNA repair (Vaze et. al. 2002). In the Freudenreich Lab, Dr. Erica Polleys modified this system to include a CAG repeat tract. This modified system features a fragmented LEU2 gene situated 25 kb 3' of the cut site and a repeat tract in both CAG and CTG orientations integrated 15 kb 5' of the cut site. Upon induction of a DSB, resection proceeds on both sides of the break in search of homology. Once successful repair is completed via SSA between the duplicated regions of LEU2 (U2 in Figure 12C), the 3' overhang must be degraded and the single-stranded template filled in (Vaze et. al. 2002). This fill-in synthesis

proceeds through the repeat tract, which is transiently single-stranded. As outlined earlier, this single-stranded repeat tract can contribute to instability, which may be monitored by PCR. Thus, this assay has two potential outputs: the viability of SSA and fill-in synthesis process (% viability in YPGal/YPD) and resultant instability of the repeat tract. To further confirm the result of the LYS2 DDRA assay, *irc20*Δ did not express a decrease in viability compared to the wildtype no repeat and repeat containing controls. Taken together, this suggests Irc20 is not needed for the short- or long-range resection mandated by the SSA pathway. Moreover, preliminary instability analysis does not indicate that Irc20 is important in repeat tract stability during fill-in synthesis (data not shown).

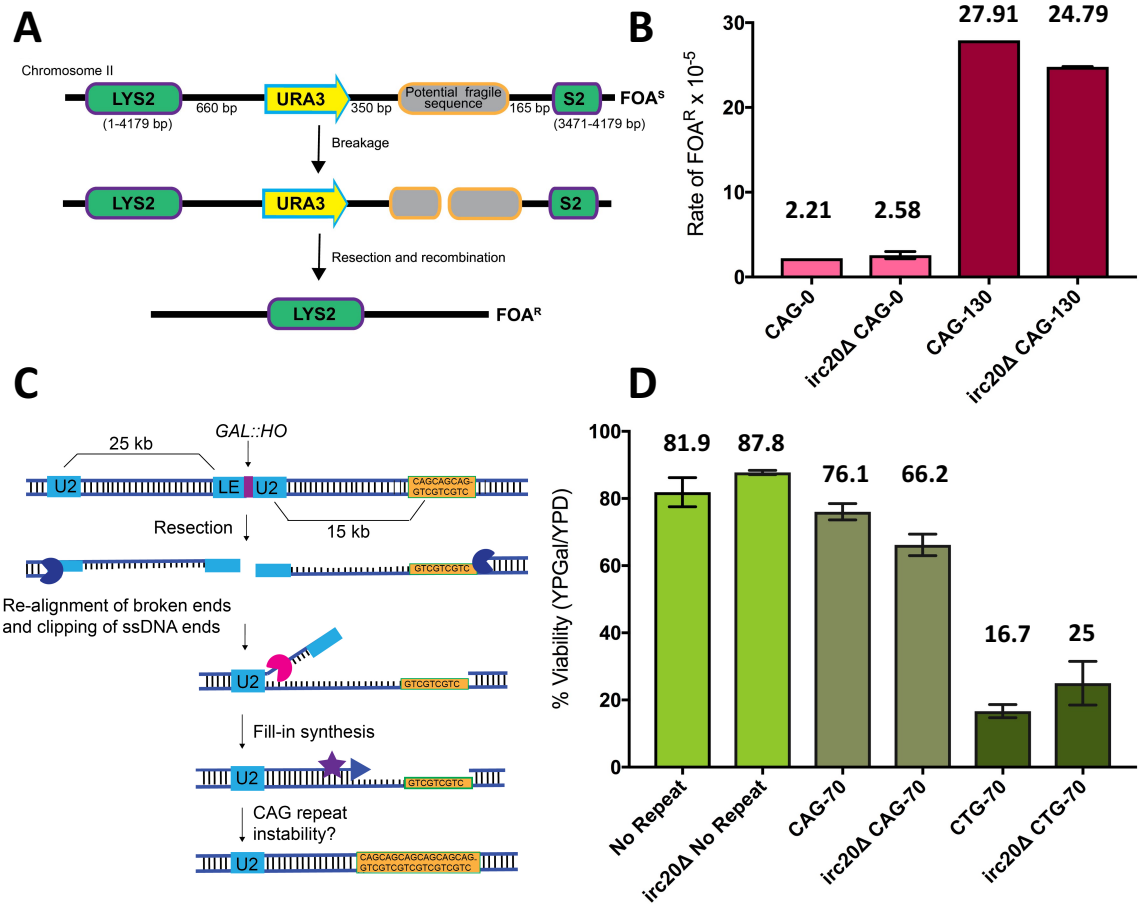


Figure 12A: Schematic of the LYS2 DDRA. Breaks occur near the repeat, healing via resection and repair by SSA between the two duplicated S2 (part of LYS2 gene) portions (Polleys & Freudenreich 2018) **B:** *irc20*Δ did not exhibit increased fragility in either tract length condition. **C:** Schematic of the SSA/Fill-in Synthesis Assay. Repeat tract is integrated in both CAG and CTG (read 5' to 3') orientations. (modified from Vaze et. al. 2002) **D:** *irc20*Δ did not exhibit decreased viability in any tract length condition.

V. Irc20 does not play a role in break-induced replication

Many DSBs are repaired utilizing both sides of the break, which share homology to the repair template. However, in certain instances the DSB occurs such that only one side of the break remains and is able to invade a homologous template. This process, termed BIR, shares extensive molecular machinery with other HR processes, and is of particular importance due to its association with large-scale deletions and chromosomal rearrangements – hallmarks of numerous human diseases (Lydeard et. al. 2010).

The BIR assay employed also operates using a galactose-induced HO cut-site (Figure 13A). The cut site is within a fragmented URA3 reporter, which shares homology to its other half on the proximal side of the same chromosome. Upon induced breakage, the fragmented URA3 reporter finds its homologous other half on the same chromosome. This invasion allows long-range synthesis, which also incorporates an adjacent TRP1 marker, while the NAT marker is lost, emulating a DSB scenario in which one side of the chromosome is lost. Chromosome V, into which this construct is integrated, inherently houses essential genes, so failure to properly repair results in cell death. Accordingly, viability on YPGal (relative to non-induction on YPD) estimates completed repair. *irc20*Δ exhibited wildtype viability (Figure 13B). Moreover, as non-BIR repair can occur, the multiple reporter and marker genes integrated into the construct allow repair typing. Most notably, an NHEJ event would result in ligation of the broken ends

and exhibit a URA$^-$ NAT$^+$ phenotype that can be scored by pinning galactose-viable colonies on selective media. Of the viable *irc20*Δ colonies, 99.5% repaired by BIR and 0.5% by NHEJ, which is very similar to a wildtype repair profile (House, Polleys, Freudenreich et. al. unpublished). This suggests that *irc20*Δ cells do not have a defect in BIR nor NHEJ.

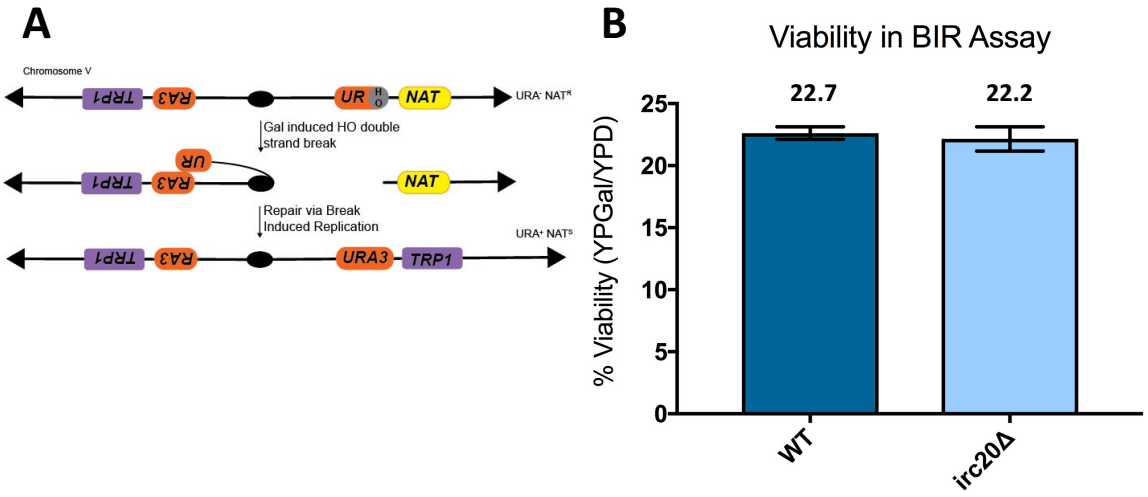


Figure 13A: Schematic of the BIR Assay and pathway of repair by intra-chromosomal BIR. (Anand et. al. 2014) **B:** *irc20*Δ exhibits wildtype viability and repair profile.

VI. Irc20 does not play a role in ectopic gene conversion

Gene conversion refers to DSB repair from a homologous template which results in a mismatch between newly synthesized DNA and the original complementary sequence. Accordingly, the mismatch is corrected and can result in loss of heterozygosity (allelic) or conversion to paralogous gene (ectopic). The Haber Lab has developed a system that utilizes the mating-type switch of *S. cerevisiae* to monitor efficiency of ectopic gene conversion (Kim & Haber 2009). Normally, upon environmental stress, an HO-endonuclease cut site within the MATα locus of chromosome III causes a DSB that is repaired by intra-chromosomal homology to HML and HMR sequences, resulting in mating-type switch (Kim & Haber 2009). The assay

construct is devoid of these intra-chromosomal homologous sequences, driving repair from an ectopic MATa-inc site on chromosome V (Figure 14A). In the experimental system, the original HO cut of MATα on chromosome III is galactose-induced, and the chromosome V MATa-inc ectopic site has a mutated HO cut-site to avoid constitutive cutting in galactose conditions (Kim & Haber 2009). Similarly to other galactose-induced systems discussed, repair efficiency is assessed by viability because failure to repair would result in loss of essential genes and cell death. *irc20*Δ did not express a viability phenotype significantly different from wildtype (Figure 14B).

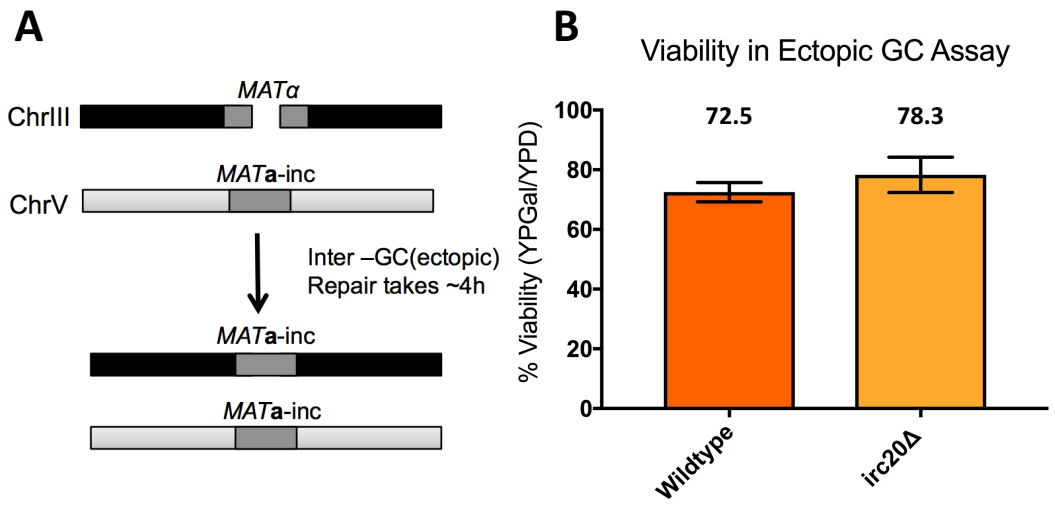


Figure 14A: Schematic of Haber Lab Ectopic GC construct. ChrIII MATα locus features mutated HML & HMR sequences to inhibit intra-chromosomal HR. ChrV Mata-inc locus features mutated HO cut-site to inhibit constitutive galactose-induced cutting. (Kim & Haber 2009) **B:** Irc20 not involved in ectopic GC pathway.

VIII. Irc20 does not play a role in relocation of collapsed forks to the nuclear periphery

The nucleus of a cell is a compact, yet highly organized entity that efficiently facilitates DNA-related processes. HR-mediated repair processes, rely on homology search, which is often

fulfilled in close proximity by a sister chromatid template, but some substrates, such as stalled forks, persist and may require re-localization to the nuclear periphery for restart or repair (Freudenreich & Su 2016). Such a pathway is believed to involve recognition of a SUMOylated intermediate by the Slx5/8 STUbL complex anchored to Nup84 at the nuclear pore complex (Freudenreich & Su 2016). Considering the implication of Irc20 homologs in recovery of stalled forks and its potential STUbL activity, determining whether an *irc20*Δ mutant had a nuclear pore localization defect was of interest. A construct, which features a GFP-marked Nup49 (periphery protein) and GFP-marked CAG_{130} tract allows visualization of S-phase nuclear location by microscopy (Figure 15A) (Su et. al. 2015). By analysis of 3D focal stacks, the repeat tract location can be assigned to one of three nuclear zones. Zone 1 is limited to the nuclear periphery, so a mutant expressing a localization defect would show decreased occupation of Zone 1 (Figure 15B) (Su et. al. 2015). This was not the case for an *irc20*Δ mutant (Figure 15C).

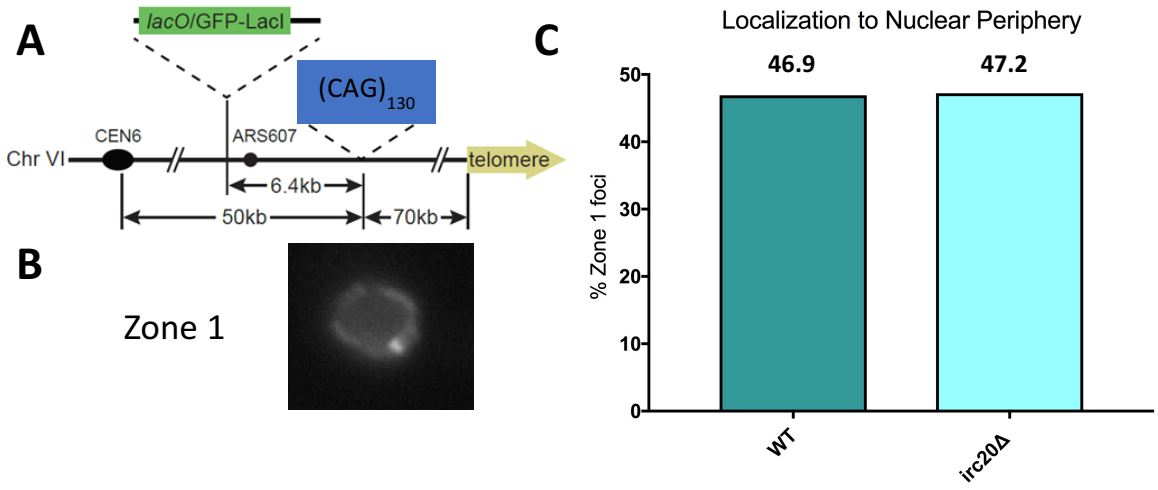


Figure 15A: Schematic of the ChrVI construct that allows tracking of a potentially stalled fork at to the nuclear periphery (Su et. al. 2015). **B:** A foci appearing at the nuclear periphery (Zone 1) (Su et. al. 2015). **C:** *irc20*Δ does not express a decreased frequency of Zone 1 occupation. 556 WT images, 125 *irc20*Δ images were analyzed.

VIII. Irc20 does not play a role in telomere length maintenance

Telomeres are the ends of linear chromosomes and their major function in the genome is to protect the ends of chromosomes from degradation and inhibit fusion with other chromosomes. Failure to maintain telomeres by telomerase or ALT pathways leads to exposure of the "end replication problem" relatively quickly (60-80 generations in yeast), and thus loss of genetic material (Gatbonton et. al. 2006). As the ALT pathways of telomere maintenance rely heavily on HR mechanisms (e.g. BIR), proteins implicated in these pathways often have a telomere maintenance function. Given the uncharacterized role of Irc20 in DNA repair, we wanted to determine whether loss of IRC20 results in changes to telomeric length. Accordingly, *irc20*Δ strains were subjected to telomeric Southern blotting. The mutant strain yielded telomeric restriction fragments of wildtype length (see "Bulk telomeres" in Figure 16), indicating lack of a role for Irc20 in telomere length maintenance. Though this is not a comprehensive study of Irc20's role in telomere maintenance, two additional genetic screens have been previously conducted to identify genes important in telomere maintenance and IRC20 was not a hit in either case, validating the conclusion that IRC20 is dispensable at the telomere (Gatbonton et. al. 2006; Askree et. al. 2004). These results suggest that healing by telomere maintenance in the YAC fragility assays is likely not affected in *irc20*Δ cells, and therefore the increase in end loss is likely due to an increase in breaks within the CAG repeat tract or decreased healing of these breaks.

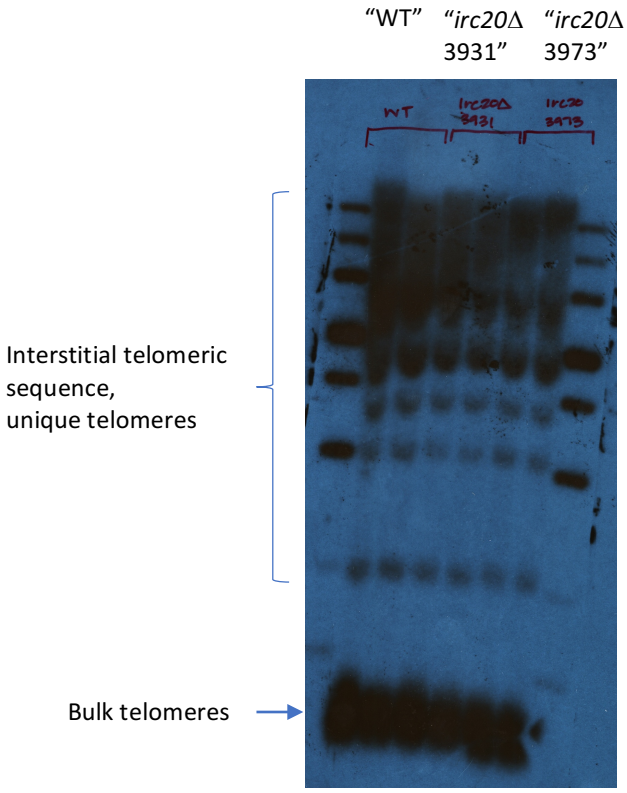


Figure 16: Telomeric Southern blot of wildtype and *irc20*Δ strains indicate that telomere length in an *irc20*Δ mutant is the same as wildtype.

IX. DNA Damage Plating does not help elucidate a role for Irc20 in DNA repair

Given the negative results in characterizing Irc20 DNA repair pathway involvement, a more general approach was taken; *irc20*Δ mutant cells were plated on a variety of DNA damaging agents. Plating cells on hydroxyurea is often an indicator of replication fork stalling, as hydroxyurea depletes dNTPs required for synthesis. Methane methylsulfonate (MMS) is a DNA alkylating agent that modifies guanine and adenine bases causing mispairing and replication barriers. Such damage is thought to repaired primarily by base excision repair (BER), but also by PRR (Lundlin 2005). Phleomycin treatment is believed to induce DSBs, which can be repaired by NHEJ or HR. Finally, camptothecin is a topoisomerase inhibitor and dominantly causes SSBs that are commonly repaired by HR. Treatment with these DNA damaging agents did not significantly impact *irc20*Δ mutant viability, further confounding Irc20's role in CAG

repeat breakage (Figure 17). However, the negative results did match previous reports (Miura et. al. 2012).

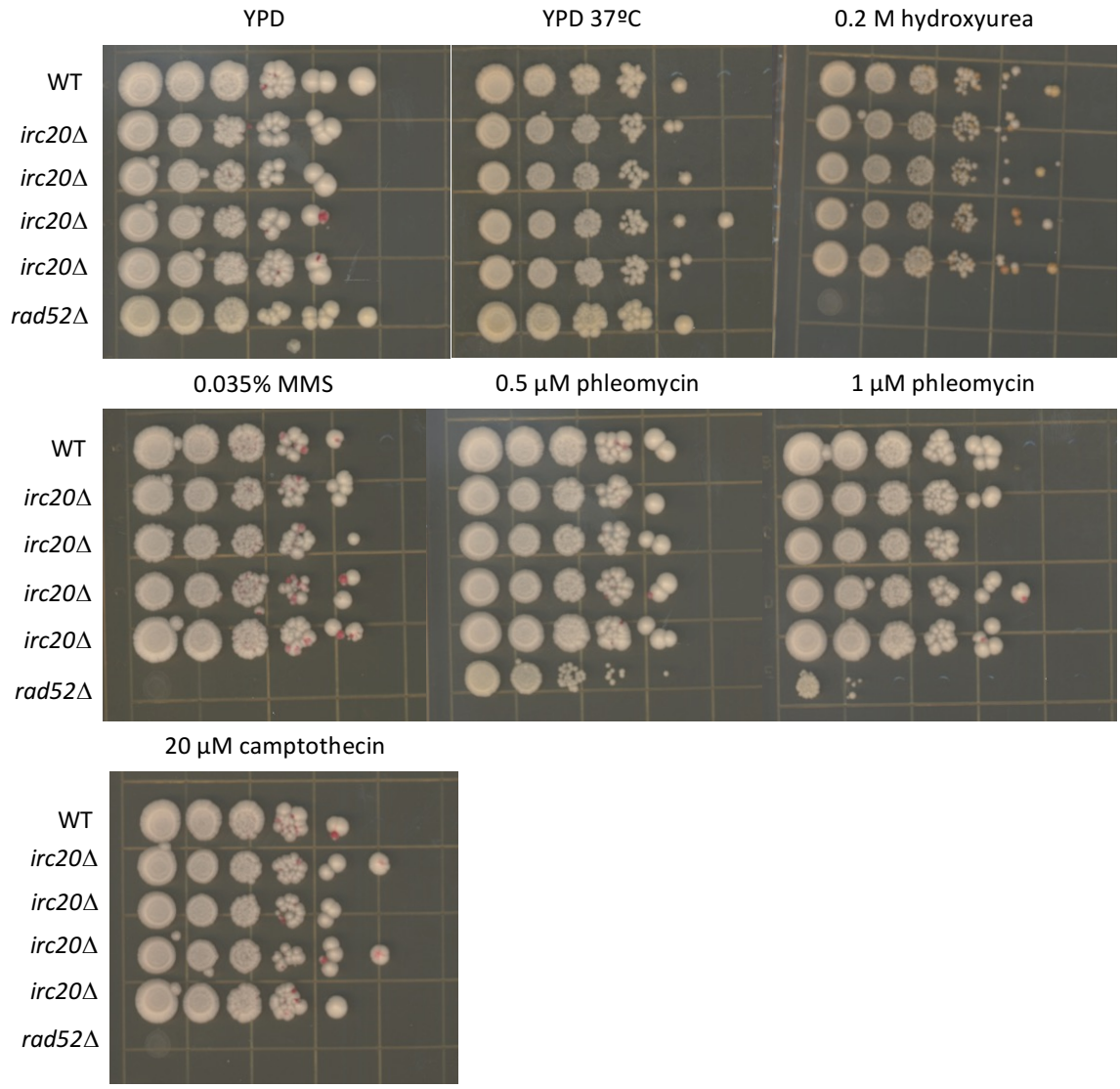


Figure 17: Each spot from left to right indicates a 10-fold serial dilution from a culture grown to saturation, the furthest left being undiluted. DNA damaging agents were added in specified concentrations to YPD plates, and one untreated plate was grown at 37°C to assess for a temperature sensitivity phenotype. WT indicates the ADE2 URA3 CAG-85 YAC. Irc20 mutants in this strain are plated in duplicate (3931, 3931, 3973, 3973) from top. Rad52 mutant also in this strain, served as a positive control for sensitivity to DNA damaging agents. *irc20*Δ proved insensitive to all tested agents/temperature.

X. Epistasis analysis

As experimentation in a multitude of experimental systems did not yield a conclusive role for Irc20 in DNA repair, we turned to an epistasis analysis using the ADE2 URA3 fragility assay. Irc20 orthologs, which fit into the Rad5 family of proteins and are implicated in the PRR pathway, were chosen for this analysis. Although a *rad5*Δ mutant has yet to be tested in the ADE2 URA3 YAC assay system, data using the CAG_{70} URA3 construct suggests that a significant fragility phenotype is not expected (Nguyen et. al. 2017). Thus, the drastic increase in *irc20*Δ *rad5*Δ mutant fragility compared to either single suggests a synergistic relationship (Figure 18). On the other hand, the *irc20*Δ , *uls1*Δ and *irc20*Δ *uls1*Δ mutants express very similar fragility phenotypes, indicating involvement in the same pathway (Figure 18). Finally, the *rad16*Δ mutant has yet to be tested alone, but the *irc20*Δ *rad16*Δ mutant phenotype suggests involvement in the same pathway, or simply lack of a role for Rad16 in preventing FOA^R (Figure 18).

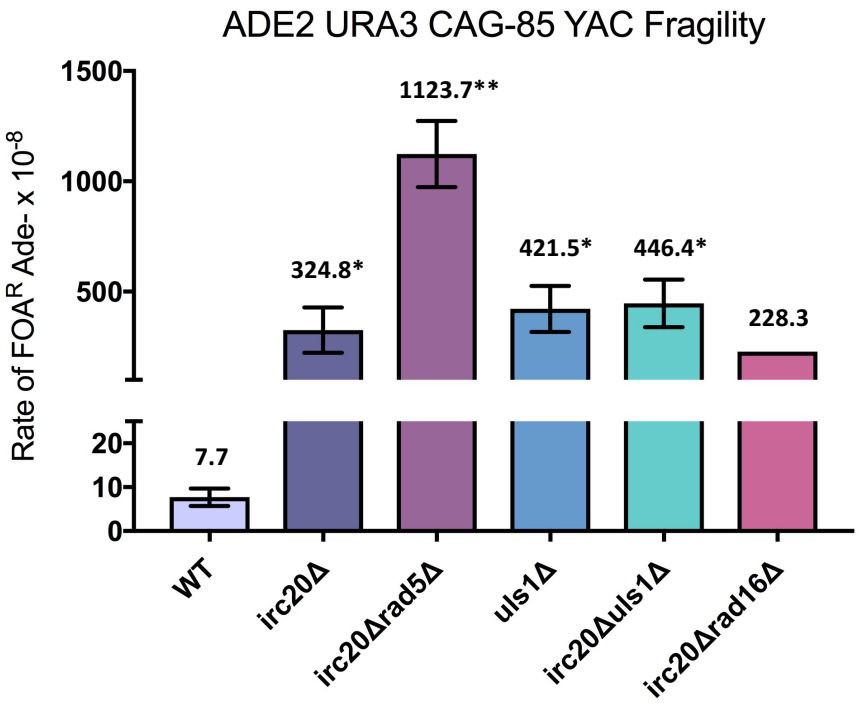


Figure 18: VPS105 strain background. *irc20*Δ *rad5*Δ mutant fragility is ~3.5-fold higher than *irc20*Δ, suggesting a synergistic relationship. *irc20*Δ , *uls1*Δ and *irc20*Δ *uls1*Δ mutants express similar fragility phenotypes, suggesting involvement in the same pathway. *rad16*Δ mutant data is inconclusive thus far. * $p < .05$, ** $p < .01$ compare to WT.

DISCUSSION

The overarching goal of this study was to further investigate the role that Irc20 plays in preventing fragility at CAG repeats, and to identify specific DNA repair pathways that Irc20 is involved in. Although major steps were taken, an abundance of negative phenotypes and a fragility phenotype narrowed to a specific YAC construct has left this characterization rather inconclusive.

The design of the ADE2 URA3 YAC is such that observed fragility rates are approximately two magnitudes lower compared to other more well characterized YAC constructs in the Freudenreich Lab. This low background fragility can be harnessed in genetic screens to detect dramatic fragility phenotypes, such as that expressed by an *irc20*Δ mutant (~42-fold increase over wildtype in CAG-85) (Figure 5B). Ideally, this ease of identification does not preclude a fragility phenotype in other YAC constructs, which can subsequently be employed for deeper exploration and historical comparisons. However, the *irc20*Δ mutant fragility phenotype could not be recapitulated in the other systems, and has thus far not been attributed to low repeat transcription, as this fragility phenotype persists even when repeat transcription is induced (~2.5-fold greater transcription-dependent increase compared to wildtype in rCAG and rCUG orientations) (Figure 6B and 7B/C). This curious phenomenon brings up a question that has continued to elude the Freudenreich Lab – Why is the background fragility of the ADE2 URA3 YAC so low? And more specifically, in the context of this project, why does Irc20 prevent

fragility only in this YAC construct? Perhaps the answer to one question will help to guide the other.

Rad5, a major player in the error-free branch of PRR, has not previously been reported to prevent CAG repeat fragility. A *rad5*Δ *irc20*Δ mutant, however, increases fragility rate drastically, suggesting an interesting synergistic role between the two proteins (Figure 18). Domain analysis categorizes the two as homologs bringing up the possibility of compensation, or perhaps highlighting Irc20 as an indispensable factor in the specific nature of PRR occurring to prevent ADE2 URA3 YAC breakage. It would be informative to test whether a similar phenotype arises for the double mutant in other YAC constructs.

Furthermore, Kramarz et. al. present a model, in which Irc20 homolog Uls1 channels error-free (Rad5-dependent) repair at stalled forks by promoting PCNA-Srs2 interaction via reduction of Srs2 SUMOylation and alleviates replication stress arising in repetitive rDNA (2017, 2014). Due to its hairpin-unwinding capacity, the helicase Srs2 is essential in preventing CAG/CTG repeat fragility (Nguyen et. al. 2017). Additionally, evidence for Srs2 and Irc20 functioning together to promote SDSA exists (Miura et. al. 2012). Finally, our fragility data implicates Irc20 and Uls1 function in the same pathway (Figure 18). Taken together, these results suggest interaction between Irc20, Srs2 and Uls1, and more generally support a role for Irc20 in PRR. We are excited by these results, but conclusive interpretation awaits confirmation using a different assay system. The conclusions of Kramarz's group draw upon DNA damage sensitivity, so testing double and triple mutants of Irc20, Srs2 and Uls1 could be an informative first step.

The lack of an instability phenotype is puzzling, as data suggests that a mutated PRR pathway may drive instability due to aberrant template switch mechanisms (Polleys, House, Freudenreich 2017). However, as instability is suppressed in *rad5*Δ double mutants exhibiting nonfunctional PRR, this result could be explained if PRR is nonfunctional in *irc20*Δ cells.

Richardson et. al. noted the possibility that Irc20 functions in transcriptional regulation (2013). Previous work has shown that R-loops, the transient DNA:RNA hybrids arising during transcription, tend to persist in CAG repeat regions, and can cause hairpin formation on the single-stranded non-template strand (Usdin et. al. 2015). In addition, it has been shown that deletion of R-loop cleaving factors RNH1 and RNH201 results in increased CAG repeat fragility (Su & Freudenreich 2017). This stalling of the transcriptional machinery may cause collision with replication machinery; either convergent or co-directional, depending upon the orientation of the gene being transcribed (Mirkin & Mirkin 2005). Such an occurrence potentially leads to CAG repeat fragility, which may require Irc20 for repair. However, if a stalled fork caused by collision with R-loops or RNA Polymerase II is prevented or repaired by Irc20 in some capacity, no evidence exists to suggest why these instances would occur more frequently in the ADE2 URA3 YAC.

Although the specific nature of repair was not conclusively identified, the SCR defect seen in a *irc20*Δ mutant matches Irc20's reported involvement in SDSA (Miura et. al. 2012). An implication of SDSA as mechanism for template switch by error-free PRR at stalled forks links these results promisingly (Kramarz et. al. 2017). It is important to mention the negative ectopic gene conversion phenotype (Figure 14) as repair via ectopic gene conversion could conceivably occur by SDSA, but the HO nuclease breaks induced by galactose in this system are not confined

to replicating cells. Irc20's potential involvement in PRR would suggest a role in SCR/SDSA specifically during replication, explaining the negative result in the ectopic GC assay. In the future, it would be informative to test Irc20 involvement using a SCR/SDSA system modified with CAG repeats to induce spontaneous repair (currently being developed by Dr. Erica Polleys).

Finally, the negative phenotypes observed in a variety of systems should not be overlooked. Data clearly suggests that Irc20 is not required for the resection and homology search required for SSA, nor the long-range synthesis of BIR (Figures 12 and 13). The former indicates that Irc20 is not involved in two of the major steps of general HR, and sheds further light on a more specific role in template-switch. The latter also excludes a unique role in long-range synthesis, and aligns well with lack of a role in telomere length maintenance (Figure 16). Stalled S-phase forks are sometimes localized to the nuclear periphery for repair or degradation. Although Irc20 does not play a role in this localization process, the possibility remains that Irc20 is needed for a restart/repair process that takes place at the periphery (Figure 15).

Much work remains in elucidating a mechanism by which Irc20 prevents CAG fragility. The first priorities are identifying reasons for the fragility phenotype in a singular system (ADE2 URA3), and continuing to characterize a role in PRR using epistasis analysis.

REFERENCES

Anand, Ranjith P., et al. "Chromosome Rearrangements via Template Switching between Diverged Repeated Sequences." *Genes & Development*, vol. 28, no. 21, Sept. 2014, pp. 2394–2406., doi:10.1101/gad.250258.114.

Askree, S. H., et al. "A Genome-Wide Screen for Saccharomyces Cerevisiae Deletion Mutants That Affect Telomere Length." *Proceedings of the National Academy of Sciences*, vol. 101, no. 23, 8 June 2004, pp. 8658–8663., doi:10.1073/pnas.0401263101.

Freudenreich, Catherine H. "Chromosome Fragility: Molecular Mechanisms and Cellular Consequences." *Frontiers in Bioscience*, vol. 12, no. 12, 2007, pp. 4911–4924., doi:10.2741/2437.

Freudenreich, CH, et al. "Stability of a CTG/CAG Trinucleotide Repeat in Yeast Is Dependent on Its Orientation in the Genome." *Molecular and Cellular Biology*, vol. 17, no. 4, 1997, pp. 2090–2098., doi:10.1128/mcb.17.4.2090.

Freudenreich, C. H. "Expansion and Length-Dependent Fragility of CTG Repeats in Yeast." *Science*, vol. 279, no. 5352, 6 Feb. 1998, pp. 853–856., doi:10.1126/science.279.5352.853.

Freudenreich, Catherine H., and Xiaofeng A. Su. "Relocalization of DNA Lesions to the Nuclear Pore Complex." *FEMS Yeast Research*, vol. 16, no. 8, 1 Dec. 2016, doi:10.1093/femsyr/fow095.

Gatbonton, Tonibelle, et al. "Telomere Length as a Quantitative Trait: Genome-Wide Survey and Genetic Mapping of Telomere Length-Control Genes in Yeast." *PLoS Genetics*, vol. 2, no. 3, Mar. 2006, doi:10.1371/journal.pgen.0020035.

Ghosal, Gargi, and Junjie Chen. “DNA Damage Tolerance: a Double-Edged Sword Guarding the Genome.” *Translational Cancer Research*, vol. 2, no. 3, 20 Sept. 2013, pp. 107–129.

Haber, James E. “Single-Strand Annealing.” *Genome Stability: DNA Repair and Recombination*, Garland Science/Taylor & Francis Group, 2014, pp. 73–87.

Hediger, Florence, et al. “Methods for Visualizing Chromatin Dynamics in Living Yeast.” *Chromatin and Chromatin Remodeling Enzymes, Part A Methods in Enzymology*, 2004, pp. 345–365., doi:10.1016/s0076-6879(03)75022-8.

House, Nealia C.M., et al. “NuA4 Initiates Dynamic Histone H4 Acetylation to Promote High-Fidelity Sister Chromatid Recombination at Postreplication Gaps.” *Molecular Cell*, vol. 55, no. 6, 18 Sept. 2014, pp. 818–828., doi:10.1016/j.molcel.2014.07.007.

Jain, S., et al. “A Recombination Execution Checkpoint Regulates the Choice of Homologous Recombination Pathway during DNA Double-Strand Break Repair.” *Genes & Development*, vol. 23, no. 3, 2009, pp. 291–303., doi:10.1101/gad.1751209.

Kadyk, L C, and L H Hartwell. “Sister Chromatids Are Preferred over Homologs as Substrates for Recombinational Repair in Saccharomyces Cerevisiae.” *Genetics*, vol. 132, no. 2, 1 Oct. 1992, pp. 387–402.

Kanao, Rie, and Chikahide Masutani. “Regulation of DNA Damage Tolerance in Mammalian Cells by Post-Translational Modifications of PCNA.” *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 803-805, 2017, pp. 82–88., doi:10.1016/j.mrfmmm.2017.06.004.

Keogh, Michael-Christopher, et al. “A Phosphatase Complex That Dephosphorylates Î³H2AX Regulates DNA Damage Checkpoint Recovery.” *Nature*, vol. 439, no. 7075, 26 Jan. 2006, pp. 497–501., doi:10.1038/nature04384.

Kile, Andrew C, et al. “HLTF's Ancient HIRAN Domain Binds 3' DNA Ends to Drive Replication Fork Reversal.” *Molecular Cell*, vol. 58, no. 6, 2015, pp. 1090–1100., doi:10.1016/j.molcel.2015.05.013.

Kim, J. A., and J. E. Haber. “Chromatin Assembly Factors Asf1 and CAF-1 Have Overlapping Roles in Deactivating the DNA Damage Checkpoint When DNA Repair Is Complete.” *Proceedings of the National Academy of Sciences*, vol. 106, no. 4, 27 Jan. 2009, pp. 1151–1156., doi:10.1073/pnas.0812578106.

Koch, Melissa R., et al. “The Chromatin Remodeler Isw1 Prevents CAG Repeat Expansions During Transcription InSaccharomyces Cerevisiae.” *Genetics*, vol. 208, no. 3, 5 Mar. 2018, pp. 963–976., doi:10.1534/genetics.117.300529.

Kramarz, Karol, et al. “DNA Damage Tolerance Pathway Choice Through Uls1 Modulation of Srs2 SUMOylation InSaccharomyces Cerevisiae.” *Genetics*, vol. 206, no. 1, 2017, pp. 513–525., doi:10.1534/genetics.116.196568.

Kramarz, Karol, et al. “Swi2/Snf2-Like Protein Uls1 Functions in the Sgs1-Dependent Pathway of Maintenance of RDNA Stability and Alleviation of Replication Stress.” *DNA Repair*, vol. 21, 2014, pp. 24–35., doi:10.1016/j.dnarep.2014.05.008.

Krogh, Berit Olsen, and Lorraine S. Symington. “Recombination Proteins in Yeast.” *Annual Review of Genetics*, vol. 38, no. 1, 2004, pp. 233–271., doi:10.1146/annurev.genet.38.072902.091500.

Lendvay, T S, et al. “Senescence Mutants of Saccharomyces Cerevisiae with a Defect in Telomere Replication Identify Three Additional Est Genes.” *Genetics*, vol. 144, no. 4, Dec. 1996, pp. 1399–1412.

Lin, Jia-Ren, et al. “SHPRH and HLTF Act in a Damage-Specific Manner to Coordinate Different Forms of Postreplication Repair and Prevent Mutagenesis.” *Molecular Cell*, vol. 42, no. 2, 2011, pp. 237–249., doi:10.1016/j.molcel.2011.02.026.

Lundin, C. “Methyl Methanesulfonate (MMS) Produces Heat-Labile DNA Damage but No Detectable in Vivo DNA Double-Strand Breaks.” *Nucleic Acids Research*, vol. 33, no. 12, 11 July 2005, pp. 3799–3811., doi:10.1093/nar/gki681.

Lydeard, J. R., et al. “Break-Induced Replication Requires All Essential DNA Replication Factors except Those Specific for Pre-RC Assembly.” *Genes & Development*, vol. 24, no. 11, Jan. 2010, pp. 1133–1144., doi:10.1101/gad.1922610.

Mcmurray, Cynthia T. “Mechanisms of Trinucleotide Repeat Instability during Human Development.” *Nature Reviews Genetics*, vol. 11, no. 11, 2010, pp. 786–799., doi:10.1038/nrg2828.

Meister, P., et al. “The Spatial Dynamics of Tissue-Specific Promoters during C. Elegans Development.” *Genes & Development*, vol. 24, no. 8, 15 Apr. 2010, pp. 766–782., doi:10.1101/gad.559610.

Mirkin, E. V., and S. M. Mirkin. “Mechanisms of Transcription-Replication Collisions in Bacteria.” *Molecular and Cellular Biology*, vol. 25, no. 3, Feb. 2005, pp. 888–895., doi:10.1128/mcb.25.3.888-895.2005.

Mirkin, Sergei M. “Expandable DNA Repeats and Human Disease.” *Nature*, vol. 447, no. 7147, 2007, pp. 932–940., doi:10.1038/nature05977.

Miura, T., et al. “Homologous Recombination via Synthesis-Dependent Strand Annealing in Yeast Requires the Irc20 and Srs2 DNA Helicases.” *Genetics*, vol. 191, no. 1, 2012, pp. 65–78., doi:10.1534/genetics.112.139105.

Mozlin, Amy M., et al. “Role of the Saccharomyces Cerevisiae Rad51 Paralogs in Sister Chromatid Recombination.” *Genetics*, vol. 178, no. 1, 1 Jan. 2008, pp. 113–126., doi:10.1534/genetics.107.082677.

Neil, Alexander J., et al. “Precarious Maintenance of Simple DNA Repeats in Eukaryotes.” *BioEssays*, vol. 39, no. 9, 2017, doi:10.1002/bies.201700077.

Nguyen, Jennifer HG, et al. “Differential Requirement of Srs2 Helicase and Rad51 Displacement Activities in Replication of Hairpin-Forming CAG/CTG Repeats.” *Nucleic Acids Research*, vol. 45, no. 8, 5 May 2017, pp. 4519–4531., doi:10.1093/nar/gkx088.

Polleys, Erica J., and Catherine H. Freudenreich. “Methods to Study Repeat Fragility and Instability in Saccharomyces Cerevisiae.” *Methods in Molecular Biology Genome Instability*, 2017, pp. 403–419., doi:10.1007/978-1-4939-7306-4_28.

Polleys, Erica J, et al. “Role of Recombination and Replication Fork Restart in Repeat Instability.” *DNA Repair*, vol. 56, Aug. 2017, pp. 156–165., doi:10.1016/j.dnarep.2017.06.018.

Richardson, Aaron, et al. “Physical and Genetic Associations of the Irc20 Ubiquitin Ligase with Cdc48 and SUMO.” *PLoS ONE*, vol. 8, no. 10, 2013, doi:10.1371/journal.pone.0076424.

Sakofsky, Cynthia, et al. “Break-Induced Replication and Genome Stability.” *Biomolecules*, vol. 2, no. 4, 2012, pp. 483–504., doi:10.3390/biom2040483.

Su, Xiaofeng A., and Catherine H. Freudenreich. “Cytosine Deamination and Base Excision Repair Cause R-Loopâ“Induced CAG Repeat Fragility and Instability InSaccharomyces Cerevisiae.” *Proceedings of the National Academy of Sciences*, vol. 114, no. 40, 3 Oct. 2017, doi:10.1073/pnas.1711283114.

Su, Xiaofeng A., et al. “Regulation of Recombination at Yeast Nuclear Pores Controls Repair and Triplet Repeat Stability.” *Genes & Development*, vol. 29, no. 10, 15 May 2015, pp. 1006–1017., doi:10.1101/gad.256404.114.

Unk, Ildiko, et al. “Role of Yeast Rad5 and Its Human Orthologs, HLTF and SHPRH in DNA Damage Tolerance.” *DNA Repair*, vol. 9, no. 3, 2010, pp. 257–267., doi:10.1016/j.dnarep.2009.12.013.

Usdin, Karen, et al. “Repeat Instability during DNA Repair: Insights from Model Systems.” *Critical Reviews in Biochemistry and Molecular Biology*, vol. 50, no. 2, 2015, pp. 142–167., doi:10.3109/10409238.2014.999192.

Vaze, Moreshwar B, et al. “Recovery from Checkpoint-Mediated Arrest after Repair of a Double-Strand Break Requires Srs2 Helicase.” *Molecular Cell*, vol. 10, no. 2, Aug. 2002, pp. 373–385., doi:10.1016/s1097-2765(02)00593-2.

APPENDIX

Table A1 VPS105 ADE2 URA3 fragility assay data

	WT CAG_{10}	*irc20Δ* CAG_{10}	WT CAG_{85}	*irc20Δ* CAG_{85}	*uls1Δ* CAG_{85}	*uls1Δirc20Δ* CAG_{85}	*rad5Δirc20Δ* CAG_{85}	*rad16Δirc20Δ* CAG_{85}
Mean Rate ($x10^{-8}$)	**1.8**	**26.2**	**7.7**	**324.8***	**421.5***	**446.4***	**1123.7****	**228.3**
Individual Assays	2.5	48.3	11.7	465.4	442.0	338.6	1409.4	228.3
	0.5	8.5	5.8	536.1	561.0	554.1	901.7	
	2.4	21.9	5.6	162.0	119.4		1060.0	
				135.5	563.7			
SEM	0.66	11.70	1.99	102.78	104.63	107.75	149.98	-
Fold Over WT		14.6		42.2	54.8	58.0	146.0	29.7
P-Value		0.105		0.048	0.020	0.0121	0.002	-

*p<0.05 to wt, **p<0.01 to wt, Student's t-test

Wildtype data from Elliot Philips. irc20Δ data from Olivia Familusi.

Table A2 Fragility assay data

Strain Background	BY4705						VPS105	
	WT CAG_{70}	*irc20Δ* CAG_{70}	WT CAG_{85}	*irc20Δ* CAG_{85}	WT 2T CAG_{70}	*irc20Δ* 2T CAG_{70}	WT CAG_{85}	*irc20Δ* CAG_{85}
Mean Rate ($x10^{-6}$)	**9.4**	**8.2**	**18.3**	**14.9**	**3.5**	**7.0***	**8.0**	**8.6**
Individual Assays	12.2	7.4	17.4	19.9	5.1	4.0	6.2	8.6
	12.6	5.1	10.2	9.8	4.7	6.8	6.2	
	7.4	12.1	19.8		2.2	8.3	11.5	
	7.0		25.3		2.0	9.0		
	8.5		12.5					
	8.7		26.7					
	8.7		16.2					
	10.4							
SEM	0.74	2.05	2.32	5.08	0.80	1.10	1.76	-
Fold Over WT		0.9		0.8		2.0		1.1
P-Value		0.477		0.519		0.0407		-

*p<0.05 to wt, **p<0.01 to wt, Student's t-test

Wildtype data from Ranjith Anand, Nealia House, Mayurika Lahiri, Jennifer Nguyen, Allen Su & Priya Sundararajan. Mutant data from Olivia Familusi, Erica Polleys & Oliver Takacsi-Nagy.

Table A3 pGAL fragility assay data

Transcript	rCUG				rCAG			
	WT Glucose	WT Galactose	*irc20Δ* Glucose	*irc20Δ* Galactose	WT Glucose	WT Galactose	*irc20Δ* Glucose	*irc20Δ* Galactose
Mean Rate ($x10^{-8}$)	8.29	31.6	115.3	1096.4	8.95	40.4	78.9	890
Individual Assays	12.2	49.8	115.3	1096.4	9.9	36.5	78.9	890
	4.38	13.4			8	44.3		
SEM	3.91	18.2	-	-	0.95	3.9	-	-
Average Galactose/Glucose	**3.6**		**9.5**		**4.6**		**11.3**	
Fold Over WT			2.64				2.45	
P-Value			-				-	

*p<0.05 to wt, **p<0.01 to wt, Student's t-test

Wildtype data from Stacey Fair. Mutant data from Erica Polleys & Oliver Takacsi-Nagy.

Table A4 Instability assay data

		Expansions								Contractions							
		> +2 CAG				> +1 CAG				< -2 CAG				< -1 CAG			
ADE2 URA3 CAG_{85}	Total Reactions	#	%	Fold over wt	p-value	#	%	Fold over wt	p-value	#	%	Fold over wt	p-value	#	%	Fold over wt	p-value
Wildtype	144	38	26.4	-	-	45	31.3	-	-	2	1.4	-	-	10	6.9	-	-
irc20Δ	165	48	29.1	1.1	0.613	65	39.4	1.3	0.154	1	0.6	0.4	0.600	11	6.7	1.0	1.000

p-values by Fisher's Exact Test

Data from Erica Polleys & Oliver Takacsi-Nagy.

Table A5 Sister chromatid recombination assay data

	Wildtype	*irc20Δ*
Mean Rate ($x10^{-4}$)	**4.62**	**1.88***
Individual Assays	4.6	1.5
	7.3	2.5
	4.0	1.7
	4.2	1.8
	1.7	
	6.1	
SEM	0.79	0.23
Fold Over WT		0.4
P-Value		0.026

*p<0.05 to wt, **p<0.01 to wt, Student's t-test

Data from Erica Polleys & Oliver Takacsi-Nagy.

Table A6 SSA/Fill-in synthesis assay data

	Parent	WT No Repeat	*irc20Δ* No Repeat	WT CAG_{70}	*irc20Δ*CAG_{70}	WT CTG_{70}	*irc20Δ*CTG_{70}
Mean Viability (%)	**77.7**	**81.9**	**87.8**	**76.1**	**66.2**	**16.7**	**25**
Individual Assays	69.6	78.2	87.1	76.6	69.4	12	31.5
	92.6	81.5	88.4	66.7	63	23.1	18.5
	76.4	73.8		77.3		16.4	
	72.2	94		79.6		13.6	
				80.1		18.3	
SEM	5.16	4.34	0.65	2.43	3.2	1.94	6.5
Fold Over WT			1.1		0.9		1.5
P-Value	-	-	0.418	-	0.076	-	0.139

*p<0.05 to wt, **p<0.01 to wt, Student's t-test

Data from Erica Polleys & Oliver Takacsi-Nagy.

Table A7 Direct duplication recombination assay data

	WT CAG_0	*irc20Δ* CAG_0	WT CAG_{130}	*irc20Δ* CAG_{130}
Mean Rate ($x10^{-5}$)	**2.21**	**2.58**	**27.91**	**24.79**
Individual Assays	2.21	3.01	27.91	24.75
		2.15		24.83
SEM	-	0.43	-	0.04
Fold Over WT	-	1.2	-	0.9
P-Value	-	-	-	-

*p<0.05 to wt, **p<0.01 to wt, Student's t-test

Data from Oliver Takacsi-Nagy.

Table A8 Break-induced replication assay data

	Wildtype	*irc20Δ*
Mean Viability (%)	**22.7**	**22.2**
Individual Assays	21	23.3
	23	23
	22.9	19.2
	22.5	23.1
	24.6	
	22.3	
SEM	0.48	0.99
Fold Over WT	-	0.98
P-Value	-	0.579
BIR (%)	97.2	99.5
NHEJ (%)	0.9	0.5
GC (%)	0.3	0
End loss/large deletion (%)	1.3	0

*p<0.05 to wt, **p<0.01 to wt, Student's t-test

Data from Erica Polleys & Oliver Takacsi-Nagy.

Table A9 Ectopic gene conversion assay data

	Wildtype	*irc20Δ*
Mean Viability (%)	**72.5**	**78.3**
Individual Assays	66.8	83.5
	74.6	84.8
	71.5	66.5
	64.8	
	65	
	75.7	
	88.9	
SEM	3.21	5.90
Fold Over WT	-	1.1
P-Value	-	0.376

*p<0.05 to wt, **p<0.01 to wt, Student's t-test

Data from Erica Polleys & Oliver Takacsi-Nagy.

Table A10 Nuclear localization assay data

	Wildtype	*irc20*Δ
No. images analyzed	556	125
No. S-phase Zone 1 foci	261	59
%	**46.9**	**47.2**
P-Value	-	1.000

p-values by Fisher's Exact Test

Data from Oliver Takacsi-Nagy & Jenna Whalen.