MECHANISMS OF CHROMOSOMAL INSTABILITY INDUCED BY UNSTABLE DNA REPEATS IN YEAST *S.CEREVISIAE*

A Dissertation
Presented to
The Academic Faculty

By

Yu Zhang

In partial fulfillment
Of the requirements for the degree
Doctor of Philosophy in Biology

Georgia Institute of Technology

August 2013

MECHANISMS OF GENOME INSTABILITY INDUCED BY UNSTABLE DNA REPEATS IN YEAST *S.CEREVISIAE*

Approved by:

Dr. Kirill Lobachev, Advisor
School of Biology
Georgia Institute of Technology

Dr. Brian Hammer
School of Biology
Georgia Institute of Technology

Dr. Yuri Chernoff
School of Biology
Georgia Institute of Technology

Dr. Katie Rudd
School of Medicine
Emory University

Dr. Francesca Storici
School of Biology
Georgia Institute of Technology

Date approved: June 03, 2013

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Dr. Kirill Lobachev, for his patient and inspiring guidance throughout my Ph.D research. He introduced me to the exciting research field of genome instability and DNA repair, and is always helping me to learn the correct way of scientific research and critical thinking. His enthusiasm and motivation in science will continue to influence me for my whole academic career in the future.

I am also thankful to my committee members, Dr. Yuri Chernoff, Dr. Francesca Storici, Dr. Brian Hammer and Dr. Katie Rudd for their valuable suggestions during my PhD study.

Special thanks to my fellow labmates, Dr. Yuri Nishida, Natalie Saini and Ziwei Sheng, for their friendship and great help in my research, which made the years of Ph.D study as an important memory in my life. I am also grateful to the undergraduate student Kathleen Heller, for her valuable assistance in my research.

I want to thank my husband Jianrong Wang, for his remarkable patience and unfailing love and support, which are essential for me to overcome the difficulties over the course of my Ph.D study.

Last but not the least, I would like to thank my parents, my mother Ling Li, and my father Renjian Zhang, for their deep love and unconditional support during my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

Alu-IRs	Inverted *Alu* Repeats
CHEF	Contour-clamped Homogeneous Electric Field
ChIP	Chromatin Immunoprecipitation
CST	Cdc13-Stn1-Ten1
DSB	Double Strand Break
FRDA	Friedreich's ataxia
GCR	Gross Chromosomal Rearrangement
MMR	Mismatch Repair
PATRRs	Palindromic AT-rich Repeats
ssDNA	Single-stranded DNA
2D gel analysis	Two-dimensional Gel Electrophoresis Analysis
5-FOA	5-Fluoroorotic Acid
5-FAA	5-Fluoroanthranilic Acid

SUMMARY

DNA repetitive sequences capable of adopting non-B DNA structures are a potent source of instability in eukaryotic genomes. They are strong inducers of chromosomal fragility and genome rearrangements that cause various hereditary diseases and cancers. In addition, a subset of repeats also has an ability to expand, which leads to more than 20 human genetic diseases that are collectively known as repeat expansion diseases. However, the mechanisms underlying the potential of these structure-prone motifs to break and expand are largely unknown.

In this study, a systematic genome-wide screen was employed in yeast *Saccharomyces cerevisiae* to investigate the contributing factors of the instability of two representative non-B DNA-forming repeats: the triplex-adopting GAA/TTC tracts and the inverted repeats that can form hairpin and cruciform structures. A complete set of 4786 deletion mutations and 800 essential genes for which expression is either downregulated by doxycycline or compromised due to mRNA perturbation were used in the screens.

The GAA/TTC screen revealed that DNA replication and transcription initiation are the two major pathways governing the GAA/TTC stability in yeast, as corresponding mutants strongly induce both fragility and large-scale expansions of the repeats. Strikingly, we found that GAA/TTC tracts could recruit transcription initiation factors and promote gene expression. Furthermore, fragility increased in non-dividing cells in a time-dependent manner and was amplified in transcription initiation mutants. These data

led us to propose that triplex structure frequently formed in replication deficient mutants is the driving force for repeat-mediated instability in actively-dividing cells. At the same time, accumulation of R-loops or/and stalled RNA-polymerases inside tracts can create another barrier for replication fork during S-phase and also trigger breakage in non-diving cells.

The inverted repeats screen and follow-up experiments revealed that both replication-dependent and –independent pathways are involved in maintaining the stability of palindromic sequences. DSB intermediates analyses indicate cruciform resolution as the mechanism of break generation at the repeats. Moreover, the homologous recombination protein Rad51 was identified as a key mediator of cruciform formation and the repeat instability specifically in replication mutants.

We propose that similar mechanisms could operate in the human cells to mediate the deleterious metabolism of GAA and inverted repeats.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 DNA repeats can adopt non-B DNA structures

Eukaryotic genomes are abundant with repetitive sequence, for example, more than 50% of the DNA consists of repeats in the human genome (Lander et al, 2001). Although these repeats mostly exist in the form of B-DNA as right-handed helical duplex, a subset of them can adopt a variety of non-canonical structures. So far, more than 10 such non-B DNA conformations have been identified (Choi & Majima, 2011). Classical examples include G-quaquplex, triplex, cruciform, hairpin and Z-DNA (Sinden, 1994.) (Figure 1.1). Repetitive G-rich sequences can form square planar arrays of four guanines (G-quadruplex) where adjacent guanines interacting with each other through hydrogen bonds. On the other hand, homopurine-homopyrimidine DNA with mirror symmetry has the propensity to form a triplex, where the third strand pairs with the duplex DNA via Hoogsteen hydrogen bond. Additionally, palindromic sequences enable transition from inter-strand to intra-strand base paring that promotes formation of cruciform structures when both strands are engaged or hairpins when only one strand is involved. Z-DNA, named after the zig-zag contour of the phosphate backbone, is usually formed by alternating purine-pyrimidine sequences, especially (GC)n tracts. Formation of these structures depends on both the structural parameters and sequence composition of the repeats, and the physiological conditions generated by dynamic DNA metabolic processes. As a rule in general, the structures are promoted by negative superhelicity and

single-stranded DNA that are natural intermediates during DNA replication, transcription and repair.

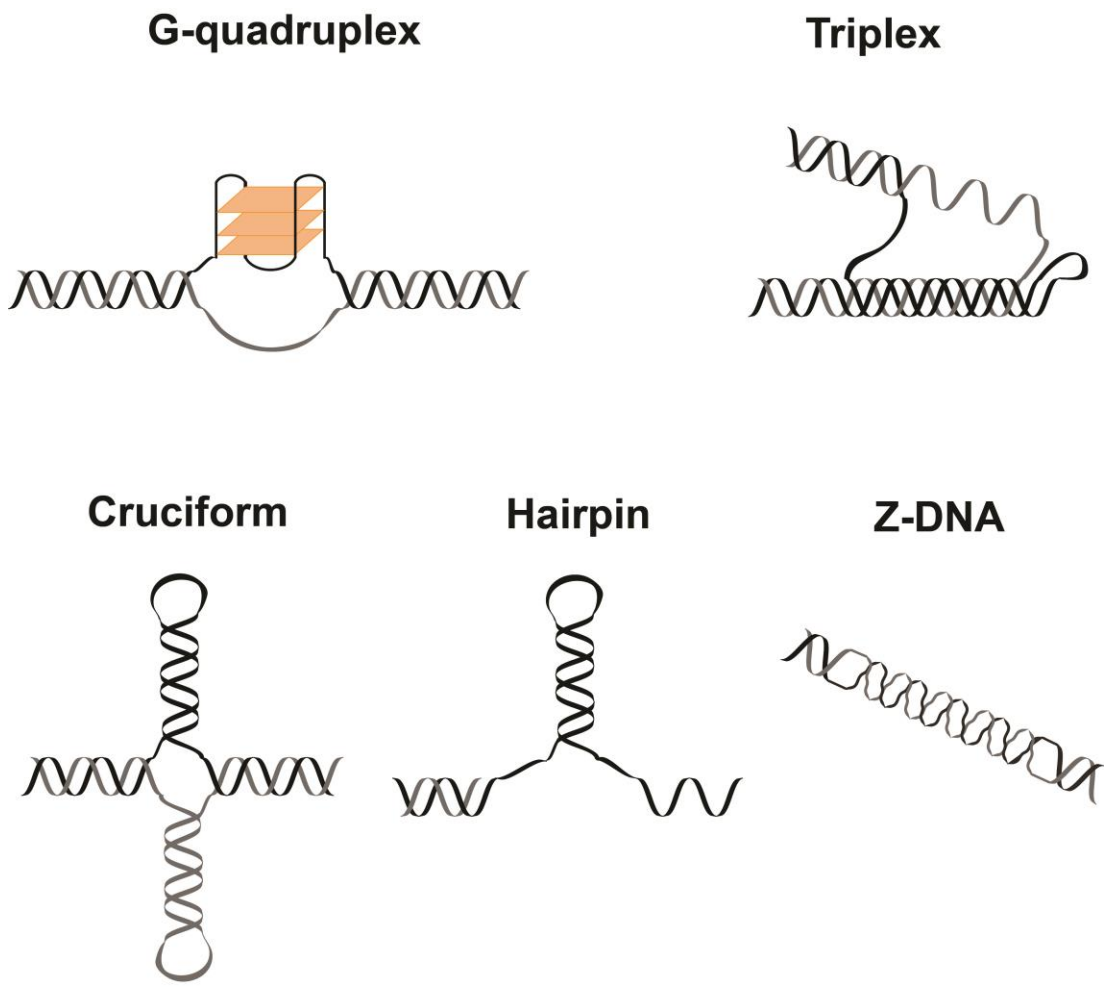

Figure 1.1 Examples of non-B DNA secondary structures

Illustrations of G-quadruplex, triplex, cruciform, hairpin and Z-DNA.

1.2 Non-B structure-adopting sequences and genome instability

1.2.1 Chromosomal fragility

Extensive evidence has shown the potential of non-B DNA-forming sequences to threaten genome integrity. These repeats are prone to induce DNA double strand breaks (DSBs), which are the triggering events of various genome rearrangements, including deletions, insertions, translocations and gene amplifications in human and model organisms. These chromosomal aberrations are hallmarks of a number of cancers and hereditary diseases (Bacolla et al, 2006; Leach, 1994; Raghavan & Lieber, 2006).

Analyses of the gross rearrangement breakpoint database and the human gene mutation database revealed that non-B DNA motifs are overrepresented at the vicinity of breakpoints in the human genome (Abeysinghe et al, 2003; Bacolla et al, 2006; Chuzhanova et al, 2003; Stenson et al, 2003), supporting a role of structure-prone sequences in the etiology of human diseases. Careful genetic characterization of translocation junctions has provided direct evidence for the involvement of non-B DNA structures in initiating chromosomal breaks. In the case of follicular lymphoma, the triplex-forming repeats located at the major breakpoints region of Bcl-2 gene was proven to generate breaks on chromosome 8 that participates in the translocations responsible for the disease (Raghavan et al, 2005; Raghavan et al, 2004). Another famous example is Emanuel syndrome, where two palindromic AT-rich sequences mediate the translocation between chromosome 11 and 22. The offspring of these balanced-translocation carriers could develop the Emanuel syndrome manifested as both physical and mental impairment (Edelmann et al, 2001; Zackai & Emanuel, 1980). Another indication that non-B DNA could cause DNA fragility comes from the studies of fragile sites, the gaps or breaks observed on metaphase chromosome that occur either spontaneously in a cell or after

exposure to certain replication stress-inducing agents (Arlt et al, 2006; Durkin & Glover, 2007; Schwartz et al, 2006; Zlotorynski et al, 2003). Long CGG and AT-rich repeats that can adopt hairpin structures have been found to form a variety of rare fragile sites. Breakage at these fragile sites has been implicated in several syndromes, including the Jacobson syndrome caused by breakage at FRA11B (Jones et al, 2000). Similar to their effects in the human genome, secondary structure-forming repeats are hotspots for DSBs and frequently trigger genome rearrangements in *E. coli* and yeast (Darmon et al, 2010; Freudenreich et al, 1998; Kim et al, 2008; Lemoine et al, 2005; Lobachev et al, 2002; Narayanan et al, 2006).

1.2.2 Size variations

In addition to inducing chromosomal fragility, the unstable DNA repeats are also prone to size variations. Expansion of microsatellites (repeating units of 2-8 bp of nucleotides) has been found to cause more than 20 human hereditary diseases that are collectively called repeat expansion diseases (La Spada & Taylor, 2010; Pearson et al, 2005). Of importance, only repeats capable of forming non-B DNA structures are linked to these diseases, suggesting a crucial role of abnormal secondary structures in driving the repeat expansion.

The expansion of the repeats can occur in both the coding and non-coding regions of various genes, resulting in either the inactivation of genes, generation of toxic RNA transcripts, or production of proteins exhibiting gain or loss of function, which lead to diseases. Examples for each class are listed below.

The fragile X mental retardation syndrome is caused by expansion of CGG repeats located in the 5' UTR of the FMR1 gene, followed by hypermethylation in the promoter that inactivates the gene (Bell et al, 1991; Heitz et al, 1991). In the case of myotonic dystrophy type 1 and 2, RNA toxicity is believed to be the major driver of the two diseases. The expanded hairpin-forming CUG (myotonic dystrophy 1) or CCUG (myotonic dystrophy 2) repeats in the RNA transcripts of DMPK (myotonic dystrophy 1) or ZNF9 (myotonic dystrophy 2)) renders cellular toxicity by sequestering the MBNL1 protein that functions in RNA-splicing of many genes (Kanadia et al, 2003). Huntington disease, the representative of the polyglutamine diseases, is caused by protein toxic gain of function. Expansion of CAG (codes for glutamine) repeats in the coding region of huntingtin gene leads to neuronal aggregation of the encoded protein and stimulates autophagy of the cells. (Kegel et al, 2000; Petersen et al, 2001).

1.3 Models for non-B DNA-induced genome instability

The abundance of unstable DNA repeats and their association with various diseases makes it an important task to identify the mechanisms underlying their instability. Although the exact molecular processes leading to the repeat fragility and expansions are still largely unknown, a substantial number of studies from model organisms to human have provided clues about the possible mechanisms of repeat instability. Among those, the influence of DNA replication, transcription and repair are under extensive study. Importantly, usually more than one pathway is identified as contributing factors, likely reflecting the complexity of the repeat metabolism.

1.3.1 DNA replication

During DNA replication, the duplex DNA is unwound into single-stranded DNA. Also, movement of the replicating fork generates positive supercoiling in front of the fork and negative supercoiling behind the fork, providing opportunity for the formation of abnormal secondary structures, which is otherwise an energy-unfavorable process. These structures could further impede DNA replication, leading to fork collapse and repeat fragility. On the other hand, the attempt to bypass the barriers can result into expansions.

Consistent with this model, the unstable DNA repeats have been shown to block DNA polymerase progression in vitro. Moreover, using two-dimensional gel analysis, a technique to monitor replication fork progression *in vivo*, various repeat sequences including CGG, GAA, palindromic sequences and repeats from the common fragile site FRA16D, have been found to hamper fork movement (Kim et al, 2008; Krasilnikova & Mirkin, 2004; Pelletier et al, 2003; Samadashwily et al, 1997; Voineagu et al, 2008; Zhang & Freudenreich, 2007). The degree of replication stalling correlates with the severity of the repeat instability and depends on the length, the sequence composition, and the orientation of repeats. The replication arrest is usually only obvious for repeats that exceed certain size threshold and have the structure-prone sequences located on the lagging-strand synthesis template. Also, interruptions in the repeat sequences that destabilize the structure-formation parameters strongly decrease the replication block and stabilize the repeats (Rolfsmeier & Lahue, 2000; Samadashwily et al, 1997; Voineagu et al, 2008).

In support of the involvement of the replication pathway are also studies of the effect of replication proteins on repeats instability. Fen1, the DNA nuclease required for Okazaki

fragment processing, has been demonstrated to maintain the stability of several repeats, including CAG, CGG and palindromic sequences (Callahan et al, 2003; Lobachev et al, 2000; Spiro et al, 1999). The replication fork stabilizer complex Mrc1-Csm3-Tof1 was shown to stabilize CAG and GAA repeats located on the chromosome and have an important role in facilitating replication progression through CGG and palindromic sequences placed on plasmids (Razidlo & Lahue, 2008; Shishkin et al, 2009; Voineagu et al, 2008; Voineagu et al, 2009). Also, deficiency in DNA polymerases strongly destabilizes repeat sequences such as CAG and palindromes (Gordenin et al, 1993; Petruska et al, 1998; Yang et al, 2003). Moreover, post-replication repair proteins have been shown to promote repeat instability while helping the fork to bypass the repeats (Cherng et al, 2011; Shishkin et al, 2009).

Overall, these studies demonstrate that intact replication machinery is important for preventing repeat instability.

1.3.2 DNA transcription

DNA transcription is suggested as an alternative pathway for repeat instability from various studies. It provides another DNA metabolic process to transiently expose single-stranded DNA as well as generate negative supercoiling behind the RNA polymerase (Witz & Stasiak, 2010), which are the conditions required for the formation of non-B DNA structures.

Consistently, trinucleotide repeats undergoing expansions and contractions are unstable in both proliferating and terminally differentiated cells wherein the tissue type and age of cells are determinants of instability. For example, in Friedreich's ataxia, GAA

contractions are the dominant form of repeat instability in all examined tissues except for neuronal tissues, which have an expansion bias (De Biase et al, 2007a). This phenomenon of somatic mosiacism has also been observed for Huntington's disease and myotic dystrophy type 1 (Fortune et al, 2000; Gonitel et al, 2008; Kennedy & Shelbourne, 2000; Shelbourne et al, 2007). Also, large expansions occur in these diseases in an age-dependent manner, which is consistent with the progressive severity of the diseases (De Biase et al, 2007b; Kennedy et al, 2003). These observations imply the existence of replication-independent pathway operating in the repeat instability.

Direct evidence showing the contribution of transcription in repeat instability first came from the studies of CAG repeats. It was found that the instability of CAG tracts positively correlates with the transcription activity across the repeat region (Bowater et al, 1997; Jung & Bonini, 2007; Mangiarini et al, 1997). Moreover, convergent transcription consisting of both sense and antisense transcription through the repeats largely destabilize CAG tracts with an effect greater than the sum of transcription from either direction (Lin et al, 2010b; Nakamori et al, 2011). Recently, it has also been shown that the potential of GAA repeats to contract and expand depends on the level of the transcription, with moderate transcription drives expansion while high-level transcription majorly promotes contraction (Ditch et al, 2009). An emerging role of abnormal RNA·DNA hybrids (R-loop) in destabilizing repeats has also been revealed. Expanded GAA, CAG and CGG tracts were shown to form transcription-dependent R-loops (Grabczyk et al, 2007; Lin et al, 2010a; Reddy et al, 2011) that were demonstrated to be recombinogenic (Aguilera & Garcia-Muse, 2012). For the CAG/CTG and CCG/CGG repeats, R-loop formation is greater when CTG or CGG tracts are located on

the non-template strand. In the case of GAA/TTC repeats, R-loops could only be generated when GAA is transcribed. Given that CTG, CGG and GAA sequence tracts are more prone to structure-formation than their corresponding complementary strands, these data indicate an important role of non-canonical structures in promoting R-loop formation. The generation of R-loops further impedes transcription elongation and stimulates repeat instability.

1.3.3 DNA repair

Besides replication and transcription, studies in different model organisms, ranging from bacteria to human cells, have suggested the participation of various DNA repair pathways in controlling the stability of non-B DNA motifs.

The complex effect of different repair pathways is well reflected by their participation in CAG instability. The base excision repair enzyme Ogg1 is a DNA glycosylase responsible for the excision of 7,8-dihydro-8oxoguanine, a common base lesion resulting from oxidative damage . Deficiency in Ogg1 was shown to reduce somatic instability for CAG repeats (Kovtun et al, 2011; Kovtun et al, 2007). Further in vitro study implicated that other components of base excision repair, including DNA polymerase β, high mobility group box 1 and FEN1, coordinate to direct the DNA lesion into expansion (Liu et al, 2009; Xu et al, 2013). Transcription-coupled nucleotide excision repair is also a player in CAG stability. Knock-down of the key components of this pathway such as CSB, ERC1, XPG and XPA1, decreases CAG contractions (Hubert et al, 2011; Lin & Wilson, 2007). The activity of mismatch repair (MMR) proteins is another important triggering event of CAG size variations, as deficiency in MSH2, MSH3 and PMS2 suppresses CAG expansions (Foiry et al, 2006; Gomes-Pereira et al, 2004; Kovtun &

McMurray, 2001; Lin & Wilson, 2009; Manley et al, 1999; Savouret et al, 2004; Tome et al, 2009). Current models propose that hairpin-loops formed by CAG or CTG repeats on newly-synthesized DNA during repair synthesis give rise to expansions while hairpin-loops on template strand lead to contractions.

Mismatch repair (MMR) factors are also key players in GAA instability. In yeast, MMR proteins stimulate large contractions and chromosomal fragility and the nuclease activity of PMS1 is crucial for these effects (Kim et al, 2008). Similarly, deficiency in MMR decreases the expansion rate of GAA tracts in cultured iPS cells and suppresses somatic instability of the repeats (Bourn et al, 2012; Du et al, 2012). On the other hand, loss of MMR factors promotes size variations of the tracts during intergenerational transmission in mice and Pms2 is required to prevent large expansion of GAA in neuronal tissues (Ezzatizadeh et al, 2012; protein et al, 2012).

In addition to affecting genome stability, MMR proteins have also been shown to bind to G4-DNA in the switch regions of immunoglobulin genes to stimulate class switch recombination for B cell maturation (Larson et al, 2005).

In summary, these data clearly indicate the importance of repair pathways in the repeat metabolism. In-depth investigation on the specific roles of each pathway and the crosstalk between these pathways in the future will shed more light on the mechanisms of repeat instability.

1.4 Overview of the dissertation

The association of the non-B DNA-forming sequences with a number of human diseases brings forth the importance of understanding the mechanisms underlying repeat

instability. Elucidation of the key players that affect the ability of the repeats to break, expand, affect gene expression and threaten genome integrity, is crucial for understanding the etiology and pathogenesis of the repeats disorders and identifying the molecular targets for therapy.

This dissertation focuses on two representative non-B DNA-adopting motifs, the trinucleotide GAA repeats that could form triplex and R-loops, and the palindromic sequences capable of adopting hairpin and cruciform structures. The overall goal of the research is to characterize the genetic factors that drive the instability of the two unstable DNA repeats in eukaryotic genomes. Both sequence motifs are a common polymorphism in the human genome. They gain their infamy because their association with diseases. Expansion of GAA repeats inactivates the FXN gene, causing the most common hereditary ataxia Friedreich's ataxia (Campuzano et al, 1996). Palindromic sequences-mediated chromosomal aberrations have also been implicated in a number of hereditary diseases (Kato et al, 2012) and are associated with DNA amplification in various cancers (Guenthoer et al, 2012).

By carrying out systematic genome-wide screens and follow-up experiments in yeast *S. cerevisiae*, we gained important mechanistic insights into GAA repeats and palindromic sequences instability. This is the first genome-wide analysis of potential players in unstable DNA repeat instability. The unbiased screen allowed us to identify key proteins influencing the potential of GAA repeats to break and expand and the ability of palindromic sequences to cause chromosomal fragility.

Chapter 2 presents a global genetic analysis of factors driving the expansion and fragility of GAA repeats. Replication and transcription initiation were identified as the two major pathways in governing GAA stability. Significantly, most mutants that stimulate fragility also promote the repeat expansion, implying a common initiation event, likely the formation of abnormal structures, in triggering both two processes. Importantly, GAA repeats were found to recruit transcription initiation factors and promote gene expression. This function of the repeats might be important for stimulating their instability outside of S-phase. Consistently, fragility at the tracts increases in a time-dependent manner in non-dividing cells and is augmented in transcription-initiation defective mutants.

Chapter 3 presents a genome-wide examination of genetic players in inverted repeat fragility. This study identified proteins whose deficiency strongly augments fragility at both 100% homologous and 94% homeologous palindromic sequences, the latter represents the imperfect palindromes that are less characterized in previous studies but prevail over perfect palindromes in the human genome. Analysis of DSB intermediates in wild-type and mutant strains provides strong evidence for cruciform resolution as the mechanism of DSB generation at the repeats. Surprisingly, we found that homologous recombination proteins are the key mediators for cruciform formation and instability under conditions of compromised replication, as the removal of Rad51 significantly decreases the DSB formation and abrogates the replication arrest in replication mutants. These observations revealed an unexpected detrimental role of homologous recombination proteins in promoting genome instability while bypassing abnormal structures during replication.

1.5 References

Abeysinghe SS, Chuzhanova N, Krawczak M, Ball EV, Cooper DN (2003) Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. *Human mutation* **22:** 229-244

Aguilera A, Garcia-Muse T (2012) R loops: from transcription byproducts to threats to genome stability. *Molecular cell* **46:** 115-124

Arlt MF, Durkin SG, Ragland RL, Glover TW (2006) Common fragile sites as targets for chromosome rearrangements. *DNA repair* **5:** 1126-1135

Bacolla A, Wojciechowska M, Kosmider B, Larson JE, Wells RD (2006) The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA repair* **5:** 1161-1170

Bell MV, Hirst MC, Nakahori Y, MacKinnon RN, Roche A, Flint TJ, Jacobs PA, Tommerup N, Tranebjaerg L, Froster-Iskenius U, et al. (1991) Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. *Cell* **64:** 861-866

Bourn RL, De Biase I, Pinto RM, Sandi C, Al-Mahdawi S, Pook MA, Bidichandani SI (2012) Pms2 suppresses large expansions of the (GAA.TTC)n sequence in neuronal tissues. *PloS one* **7:** e47085

Bowater RP, Jaworski A, Larson JE, Parniewski P, Wells RD (1997) Transcription increases the deletion frequency of long CTG.CAG triplet repeats from plasmids in Escherichia coli. *Nucleic acids research* **25:** 2861-2868

Callahan JL, Andrews KJ, Zakian VA, Freudenreich CH (2003) Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. *Molecular and cellular biology* **23:** 7849-7860

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Canizares J, Koutnikova H, Bidichandani SI, Gellera C, Brice A, Trouillas P, De Michele G, Filla A, De Frutos R, Palau F, Patel PI, Di Donato S, Mandel JL, Cocozza S, Koenig M, Pandolfo M (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271:** 1423-1427

Cherng N, Shishkin AA, Schlager LI, Tuck RH, Sloan L, Matera R, Sarkar PS, Ashizawa T, Freudenreich CH, Mirkin SM (2011) Expansions, contractions, and fragility of the spinocerebellar ataxia type 10 pentanucleotide repeat in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 2843-2848

Choi J, Majima T (2011) Conformational changes of non-B DNA. *Chemical Society reviews* **40:** 5893-5909

Chuzhanova N, Abeysinghe SS, Krawczak M, Cooper DN (2003) Translocation and gross deletion breakpoints in human inherited disease and cancer II: Potential involvement of repetitive sequence elements in secondary structure formation between DNA ends. *Human mutation* **22:** 245-251

Darmon E, Eykelenboom JK, Lincker F, Jones LH, White M, Okely E, Blackwood JK, Leach DR (2010) E. coli SbcCD and RecA control chromosomal rearrangement induced by an interrupted palindrome. *Molecular cell* **39:** 59-70

De Biase I, Rasmussen A, Endres D, Al-Mahdawi S, Monticelli A, Cocozza S, Pook M, Bidichandani SI (2007a) Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients. *Annals of neurology* **61:** 55-60

De Biase I, Rasmussen A, Monticelli A, Al-Mahdawi S, Pook M, Cocozza S, Bidichandani SI (2007b) Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. *Genomics* **90:** 1-5

Ditch S, Sammarco MC, Banerjee A, Grabczyk E (2009) Progressive GAA.TTC repeat expansion in human cell lines. *PLoS genetics* **5:** e1000704

Du J, Campau E, Soragni E, Ku S, Puckett JW, Dervan PB, Gottesfeld JM (2012) Role of mismatch repair enzymes in GAA.TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells. *The Journal of biological chemistry* **287:** 29861-29872

Durkin SG, Glover TW (2007) Chromosome fragile sites. *Annual review of genetics* **41:** 169-192

Edelmann L, Spiteri E, Koren K, Pulijaal V, Bialer MG, Shanske A, Goldberg R, Morrow BE (2001) AT-rich palindromes mediate the constitutional t(11;22) translocation. *American journal of human genetics* **68:** 1-13

Ezzatizadeh V, Pinto RM, Sandi C, Sandi M, Al-Mahdawi S, Te Riele H, Pook MA (2012) The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model. *Neurobiology of disease* **46:** 165-171

Foiry L, Dong L, Savouret C, Hubert L, te Riele H, Junien C, Gourdon G (2006) Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Human genetics* **119:** 520-526

Fortune MT, Vassilopoulos C, Coolbaugh MI, Siciliano MJ, Monckton DG (2000) Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Human molecular genetics* **9:** 439-445

Freudenreich CH, Kantrow SM, Zakian VA (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science* **279:** 853-856

Gomes-Pereira M, Fortune MT, Ingram L, McAbney JP, Monckton DG (2004) Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Human molecular genetics* **13:** 1815-1825

Gonitel R, Moffitt H, Sathasivam K, Woodman B, Detloff PJ, Faull RL, Bates GP (2008) DNA instability in postmitotic neurons. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 3467-3472

Gordenin DA, Lobachev KS, Degtyareva NP, Malkova AL, Perkins E, Resnick MA (1993) Inverted DNA repeats: a source of eukaryotic genomic instability. *Molecular and cellular biology* **13:** 5315-5322

Grabczyk E, Mancuso M, Sammarco MC (2007) A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic acids research* **35:** 5351-5359

Guenthoer J, Diede SJ, Tanaka H, Chai X, Hsu L, Tapscott SJ, Porter PL (2012) Assessment of palindromes as platforms for DNA amplification in breast cancer. *Genome research* **22:** 232-245

Heitz D, Rousseau F, Devys D, Saccone S, Abderrahim H, Le Paslier D, Cohen D, Vincent A, Toniolo D, Della Valle G, et al. (1991) Isolation of sequences that span the fragile X and identification of a fragile X-related CpG island. *Science* **251:** 1236-1239

Hubert L, Jr., Lin Y, Dion V, Wilson JH (2011) Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. *Human molecular genetics* **20:** 4822-4830

Jones C, Mullenbach R, Grossfeld P, Auer R, Favier R, Chien K, James M, Tunnacliffe A, Cotter F (2000) Co-localisation of CCG repeats and chromosome deletion breakpoints in Jacobsen syndrome: evidence for a common mechanism of chromosome breakage. *Human molecular genetics* **9:** 1201-1208

Jung J, Bonini N (2007) CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. *Science* **315:** 1857-1859

Kanadia RN, Johnstone KA, Mankodi A, Lungu C, Thornton CA, Esson D, Timmers AM, Hauswirth WW, Swanson MS (2003) A muscleblind knockout model for myotonic dystrophy. *Science* **302:** 1978-1980

Kato T, Kurahashi H, Emanuel BS (2012) Chromosomal translocations and palindromic AT-rich repeats. *Current opinion in genetics & development* **22:** 221-228

Kegel KB, Kim M, Sapp E, McIntyre C, Castano JG, Aronin N, DiFiglia M (2000) Huntingtin expression stimulates endosomal-lysosomal activity, endosome tubulation, and autophagy. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **20:** 7268-7278

Kennedy L, Evans E, Chen CM, Craven L, Detloff PJ, Ennis M, Shelbourne PF (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Human molecular genetics* **12:** 3359-3367

Kennedy L, Shelbourne PF (2000) Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Human molecular genetics* **9:** 2539-2544

Kim HM, Narayanan V, Mieczkowski PA, Petes TD, Krasilnikova MM, Mirkin SM, Lobachev KS (2008) Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *The EMBO journal* **27:** 2896-2906

Kovtun IV, Johnson KO, McMurray CT (2011) Cockayne syndrome B protein antagonizes OGG1 in modulating CAG repeat length in vivo. *Aging* **3:** 509-514

Kovtun IV, Liu Y, Bjoras M, Klungland A, Wilson SH, McMurray CT (2007) OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature* **447:** 447-452

Kovtun IV, McMurray CT (2001) Trinucleotide expansion in haploid germ cells by gap repair. *Nature genetics* **27:** 407-411

Krasilnikova MM, Mirkin SM (2004) Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. *Molecular and cellular biology* **24:** 2286-2295

La Spada AR, Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nature reviews Genetics* **11:** 247-258

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML,

Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921

Larson ED, Duquette ML, Cummings WJ, Streiff RJ, Maizels N (2005) MutSalpha binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions. *Current biology : CB* **15:** 470-474

Leach DR (1994) Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *BioEssays : news and reviews in molecular, cellular and developmental biology* **16:** 893-900

Lemoine FJ, Degtyareva NP, Lobachev K, Petes TD (2005) Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell* **120:** 587-598

Lin Y, Dent SY, Wilson JH, Wells RD, Napierala M (2010a) R loops stimulate genetic instability of CTG.CAG repeats. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 692-697

Lin Y, Leng M, Wan M, Wilson JH (2010b) Convergent transcription through a long CAG tract destabilizes repeats and induces apoptosis. *Molecular and cellular biology* **30:** 4435-4451

Lin Y, Wilson JH (2007) Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. *Molecular and cellular biology* **27:** 6209-6217

Lin Y, Wilson JH (2009) Diverse effects of individual mismatch repair components on transcription-induced CAG repeat instability in human cells. *DNA repair* **8:** 878-885

Liu Y, Prasad R, Beard WA, Hou EW, Horton JK, McMurray CT, Wilson SH (2009) Coordination between polymerase beta and FEN1 can modulate CAG repeat expansion. *The Journal of biological chemistry* **284:** 28352-28366

Lobachev KS, Gordenin DA, Resnick MA (2002) The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* **108:** 183-193

Lobachev KS, Stenger JE, Kozyreva OG, Jurka J, Gordenin DA, Resnick MA (2000) Inverted Alu repeats unstable in yeast are excluded from the human genome. *The EMBO journal* **19:** 3822-3830

Mangiarini L, Sathasivam K, Mahal A, Mott R, Seller M, Bates GP (1997) Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nature genetics* **15:** 197-200

Manley K, Shirley TL, Flaherty L, Messer A (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature genetics* **23:** 471-473

Nakamori M, Pearson CE, Thornton CA (2011) Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. *Human molecular genetics* **20:** 580-588

Narayanan V, Mieczkowski PA, Kim HM, Petes TD, Lobachev KS (2006) The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell* **125:** 1283-1296

Pearson CE, Nichol Edamura K, Cleary JD (2005) Repeat instability: mechanisms of dynamic mutations. *Nature reviews Genetics* **6:** 729-742

Pelletier R, Krasilnikova MM, Samadashwily GM, Lahue R, Mirkin SM (2003) Replication and expansion of trinucleotide repeats in yeast. *Molecular and cellular biology* **23:** 1349-1357

Petersen A, Larsen KE, Behr GG, Romero N, Przedborski S, Brundin P, Sulzer D (2001) Expanded CAG repeats in exon 1 of the Huntington's disease gene stimulate dopamine-mediated striatal neuron autophagy and degeneration. *Human molecular genetics* **10:** 1243-1254

Petruska J, Hartenstine MJ, Goodman MF (1998) Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *The Journal of biological chemistry* **273:** 5204-5210

protein CsB, antagonizes OGG1 in modulating CAG repeat length in vivoBourn RL, De Biase I, Pinto RM, Sandi C, Al-Mahdawi S, Pook MA, Bidichandani SI (2012) Pms2 suppresses large expansions of the (GAA.TTC)n sequence in neuronal tissues. *PloS one* **7:** e47085

Raghavan SC, Chastain P, Lee JS, Hegde BG, Houston S, Langen R, Hsieh CL, Haworth IS, Lieber MR (2005) Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation. *The Journal of biological chemistry* **280:** 22749-22760

Raghavan SC, Lieber MR (2006) DNA structures at chromosomal translocation sites. *BioEssays : news and reviews in molecular, cellular and developmental biology* **28:** 480-494

Raghavan SC, Swanson PC, Wu X, Hsieh CL, Lieber MR (2004) A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature* **428:** 88-93

Razidlo DF, Lahue RS (2008) Mrc1, Tof1 and Csm3 inhibit CAG.CTG repeat instability by at least two mechanisms. *DNA repair* **7:** 633-640

Reddy K, Tam M, Bowater RP, Barber M, Tomlinson M, Nichol Edamura K, Wang YH, Pearson CE (2011) Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic acids research* **39:** 1749-1762

Rolfsmeier ML, Lahue RS (2000) Stabilizing effects of interruptions on trinucleotide repeat expansions in Saccharomyces cerevisiae. *Molecular and cellular biology* **20:** 173-180

Samadashwily GM, Raca G, Mirkin SM (1997) Trinucleotide repeats affect DNA replication in vivo. *Nature genetics* **17:** 298-304

Savouret C, Garcia-Cordier C, Megret J, te Riele H, Junien C, Gourdon G (2004) MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Molecular and cellular biology* **24:** 629-637

Schwartz M, Zlotorynski E, Kerem B (2006) The molecular basis of common and rare fragile sites. *Cancer letters* **232:** 13-26

Shelbourne PF, Keller-McGandy C, Bi WL, Yoon SR, Dubeau L, Veitch NJ, Vonsattel JP, Wexler NS, Arnheim N, Augood SJ (2007) Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Human molecular genetics* **16:** 1133-1142

Shishkin AA, Voineagu I, Matera R, Cherng N, Chernet BT, Krasilnikova MM, Narayanan V, Lobachev KS, Mirkin SM (2009) Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Molecular cell* **35:** 82-92

Sinden R.R., DNA Structure and Function, Academic Press, San Diego, 1994.

Spiro C, Pelletier R, Rolfsmeier ML, Dixon MJ, Lahue RS, Gupta G, Park MS, Chen X, Mariappan SV, McMurray CT (1999) Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Molecular cell* **4:** 1079-1085

Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeysinghe S, Krawczak M, Cooper DN (2003) Human Gene Mutation Database (HGMD): 2003 update. *Human mutation* **21:** 577-581

Tome S, Holt I, Edelmann W, Morris GE, Munnich A, Pearson CE, Gourdon G (2009) MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice. *PLoS genetics* **5:** e1000482

Voineagu I, Narayanan V, Lobachev KS, Mirkin SM (2008) Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 9936-9941

Voineagu I, Surka CF, Shishkin AA, Krasilnikova MM, Mirkin SM (2009) Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nature structural & molecular biology* **16:** 226-228

Witz G, Stasiak A (2010) DNA supercoiling and its role in DNA decatenation and unknotting. *Nucleic acids research* **38:** 2119-2133

Xu M, Gabison J, Liu Y (2013) Trinucleotide repeat deletion via a unique hairpin bypass by DNA polymerase beta and alternate flap cleavage by flap endonuclease 1. *Nucleic acids research* **41:** 1684-1697

Yang Z, Lau R, Marcadier JL, Chitayat D, Pearson CE (2003) Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells. *American journal of human genetics* **73:** 1092-1105

Zackai EH, Emanuel BS (1980) Site-specific reciprocal translocation, t(11;22) (q23;q11), in several unrelated families with 3:1 meiotic disjunction. *American journal of medical genetics* **7:** 507-521

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

Zlotorynski E, Rahat A, Skaug J, Ben-Porat N, Ozeri E, Hershberg R, Levi A, Scherer SW, Margalit H, Kerem B (2003) Molecular basis for expression of common and rare fragile sites. *Molecular and cellular biology* **23:** 7143-7151

CHAPTER 2

GENOME-WIDE SCREEN IDENTIFIES PATHWAYS THAT GOVERN GAA/TTC REPEAT FRAGILITY AND EXPANSIONS IN DIVIDING AND NON-DIVIDING YEAST CELLS

2.1 Summary

Triplex structure-forming GAA/TTC repeats pose a dual threat to the eukaryotic genome integrity. Their potential to expand can lead to gene inactivation, the cause of Friedreich's ataxia disease in humans. In model systems, long GAA/TTC tracts also act as chromosomal fragile sites that can trigger gross chromosomal rearrangements. The mechanisms that regulate the metabolism of GAA/TTC repeats are poorly understood. We have developed an experimental system in the yeast *Saccharomyces cerevisiae* that allows us to systematically identify genes crucial for maintaining the repeat stability. Two major groups of mutants defective in DNA replication or transcription initiation are found to be prone to fragility and large-scale expansions. We demonstrate that problems imposed by the repeats during DNA replication in actively dividing cells and during transcription initiation in non-dividing cells can culminate in genome instability. We propose that similar mechanisms can mediate detrimental metabolism of GAA/TTC tracts in human cells.

2.2 Introduction

Triplex-forming GAA/TTC trinucleotide repeat tracts are a common sequence motif present in many prokaryotic and eukaryotic genomes (Clark et al., 2006; Kassai-Jager et al., 2008). They are particularly abundant in mammalian genomes including humans where GAA/TTC loci were found to be highly polymorphic (Clark et al., 2004). Repeat tracts can both contract and expand, the latter feature is responsible for the inactivation of the *FXN* gene causing Friedreich's ataxia (FRDA) (Campuzano et al., 1996). FRDA patients are carriers of expanded repeats ranging from 66 to 1,700 copies, while normal individuals bear less than 65 repeats (Campuzano et al., 1996; Clark et al., 2004 and references therein). It has been demonstrated that premutation (34-65 repeats) and mutant (>66 repeats) alleles are unstable in both proliferating and terminally differentiated cells wherein the tissue type is one of the determinants of instability (Clark et al., 2007; De Biase et al., 2007a; De Biase et al., 2007b). While in most somatic cells contractions are a predominant class of tract length variations, neuronal cells exhibit an expansion bias. The ability of GAA/TTC repeats to undergo tract length variations is not only confined to human cells. Long tracts are also highly unstable in bacteria, yeast and mice with contractions being the most predominant class in microbial systems (Bourn et al., 2009; Clark et al., 2007; Kim et al., 2008; Krasilnikova and Mirkin, 2004; Pollard et al., 2004).

Besides being prone for size variations, expanded GAA/TTC tracts were also found to induce recombination and gross chromosomal rearrangements (GCRs) in bacteria and yeast (Kim et al., 2008; Tang et al., 2011; Vetcher and Wells, 2004). In yeast, this ability of repeats to compromise genome integrity is attributed to double-

strand break (DSB) formation. We demonstrated that replication is a factor in GAA/TTC-mediated fragility and rearrangements (Kim et al., 2008). In contrast, Tang *et al*., 2011 found that mitotic crossover stimulated by GAA/TTC tracts results from DSB formation on unreplicated chromosomes, where the breakage was deduced to occur at the G1 stage of the cell cycle.

Overall, size variations in GAA/TTC tracts and repeat-associated fragility happen in both actively dividing and non-dividing cells. This likely reflects the complexity of the processes that affect repeat metabolism where DNA replication, transcription, repair, and chromatin organization are all contributing factors to the genetic instability. The homopurine/homopyrimidine nature of GAA/TTC tracts along with their intrinsic mirror symmetry provide the repeats the ability to adopt triplex (or H-DNA) secondary structure (reviewed in Frank-Kamenetskii and Mirkin, 1995). The formation of triplex structure is considered to be one of the main triggering events for repeat-associated instability. Triplex formation is facilitated in single-stranded regions which are natural intermediates during replication, transcription and repair. Consequently, stable secondary structure, which is an abnormal template for DNA and RNA synthesis, can hinder DNA and RNA polymerase progression (Kim et al., 2008; Krasilnikova et al., 2007 and references therein). The attempt to repair or bypass these impediments eventually results in destabilization of the repeats (Bourn et al., 2009; Kim et al., 2008; Shishkin et al., 2009).

In yeast, using two-dimensional (2D) gel electrophoresis, we found that long GAA/TTC tracts can stall the progression of the replication fork in a plasmid and in the chromosome when the GAA strand is a template for lagging strand synthesis (Kim et al., 2008; Krasilnikova and Mirkin, 2004). The propensity of the repeats to block the

replication fork also correlates with their ability to stimulate chromosomal breakage and rearrangements (Kim et al., 2008). At the same time, GAA/TTC expansion potential is not dependent on the orientation of the repeats relative to replication origin (Shishkin et al., 2009). However, expansions are stimulated in the replication-checkpoint surveillance mutants *tof1* and *csm3,* and compromised in *sgs1*, *rad5* and *rad6* mutants defective in repair and post-replicative template switch. These data point strongly toward the existence of replication problems imposed by the repetitive tracts on leading as well as on lagging strands in actively dividing cells.

GAA/TTC tracts can efficiently block transcription elongation both in *vitro* and in *vivo*, presumably due to triplex formation (Krasilnikova et al, 2007 and references therein). Also, it has been shown that expanded alleles in the first intron of the *FXN* gene can promote spreading of repressive chromatin marks that can reach the promoter and affect transcription initiation (reviewed in Kumari and Usdin, 2012). Both deficiencies in elongation and initiation of transcription are implicated in the FRDA pathology. At the same time, a change in the level of transcription across repeats has a strong effect on repeat instability. In human cell lines, induction of transcription promotes deletions (Soragni et al., 2008). Increasing or decreasing the basal level of transcription antagonizes expansions (Ditch et al., 2009). The emerging mechanism of the repeat instability associated with transcription is the accumulation of persistent DNA-RNA hybrids (R-loops) in the tract regions resulting from RNA polymerase stalling (Belotserkovskii et al., 2010; Grabczyk et al., 2007; McIvor et al., 2010; Reddy et al., 2011). R-loops are known recombinagenic substrates (reviewed in Aguilera and Gomez-Gonzalez, 2008), and the processing of these structures has been implicated in

destabilization of GAA/TTC repeats (McIvor et al., 2010). Repeat instability documented in non-replicating cells can be explained by transcription-associated mechanisms. Collision of DNA polymerase with trapped RNA polymerase at R-loops can also be a factor in generation of breaks and size variations in actively-dividing cells (Rindler and Bidichandani, 2011).

In this study, we carried out a systematic and unbiased genome-wide screen in yeast *S. cerevisiae* to identify genes that govern GAA/TTC repeat instability. A complete set of 4786 deletion mutations and 800 essential genes, for which expression is either regulated by doxycycline or compromised due to mRNA perturbation, was tested. The screen identified 37 mutants that destabilize the repeats. We determined that most of the mutations that augment GAA/TTC fragility also increase the rate of repeat expansions, indicating that both types of instability can be driven by a common initiating event. We found that defects in proteins that constitute the core of the replisome and replication-pausing checkpoint surveillance strongly increase GAA/TTC fragility and expansions. A dysfunction of origin and telomere also stimulates breakage and expansions. Surprisingly, another major class of mutants exhibiting hyper-instability belongs to transcription initiation. Consistently, we have found that GAA/TTC repeats serve as promoters and recruit transcription initiation factors. Our findings suggest that problems imposed by the triplex structure during DNA replication are a major triggering event in chromosomal fragility and size variations operating in actively dividing cells. At the same time, proper transcription initiation at GAA/TTC tracts is a guardian of repeat stability in both dividing and non-dividing cells. We propose that similar mechanisms can mediate detrimental metabolism of GAA/TTC tracts in human cells.

2.3 Results

2.3.1 Genome-wide screen of mutants that augment GAA/TTC fragility

To identify mutants exhibiting an increased level of fragility mediated by GAA/TTC repeats, we carried out genome-wide screening of three collections of strains purchased from Open Biosystems. 4786 deletion mutations (YKO collection) and 860 essential genes, whose expression is either regulated by doxycycline (yTHC collection) or compromised due to mRNA perturbation (DAmP collection), were tested. The scheme for combining mutant alleles marked with *kanMX* cassette and GAA/TTC tracts is based on the approach developed by C. Boone and colleagues (Tong et al., 2001) with modifications (Figure 2.1A, B and Materials and Methods). Hyper-GCR mutants were identified based on the increased number of papillae in a visual comparison to wild-type strain (Figure 2.1C). Since the GCR rate induced by $(GAA)_{230}$ repeats (~$2x10^{-6}$) is ~10 times higher than the mutation rate (~$3x10^{-7}$) at the *CAN1* locus in wild-type strains, the screen preferentially selects for GCR inducers or strong mutators.

To verify the effect of hyper-GCR mutants and to distinguish them from strong mutators at the *CAN1* locus, we introduced the alleles into strains carrying the GCR assay described in Kim *et al.*, 2008. The *ADE2* marker placed between *CAN1* and *LYS2* helps to differentiate between DSB-mediated arm loss and mutations in *CAN1*. In this assay, GCR isolates are manifested as canavanine-resistant red colonies while mutations in *CAN1* give rise to canavanine-resistant white colonies (Figure 2.1D). The effect of the mutations was tested in the strains containing 120 and 230 copies of GAA/TTC repeats and in the strains lacking repeats.

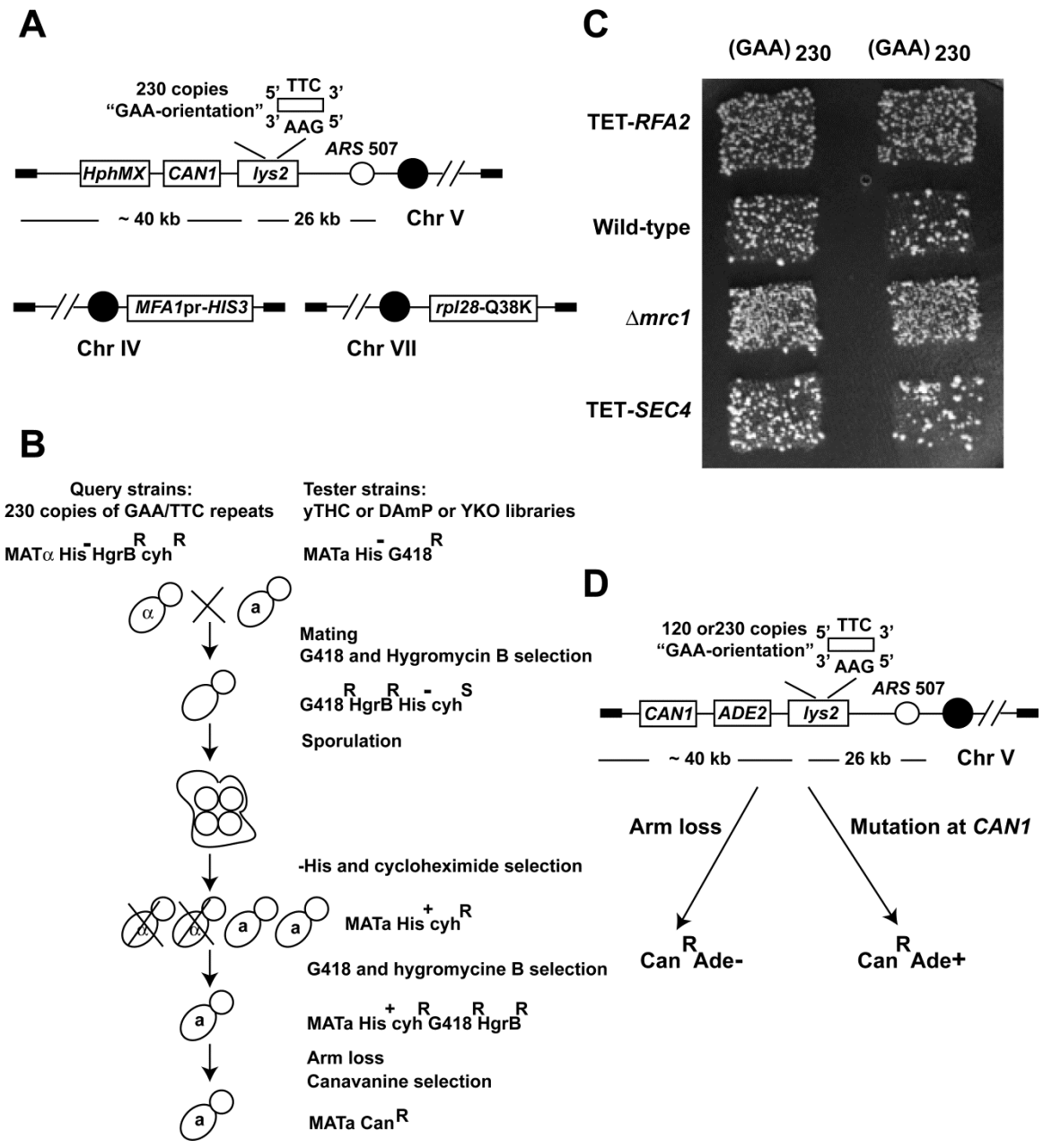

Figure 2.1. Genome-wide screen methodology.

(A) Query strain to uncover genotypes prone to GAA/TTC fragility. The 230 GAA/TTC repeats were positioned on the left arm of chromosome V with GAA as a template for lagging strand synthesis. The arrangement of *LYS2*, *hphMX* cassette and *CAN1* on chromosome V are shown. The location of *HIS3* ORF under control of MATa-specific promoter and *rpl28-Q38K* allele on chromosomes IV and VII are depicted. (B) Schematics of the screen. MATα query strains carrying GAA/TTC tract were crossed with MATa tester strains from YKO, yTHC and DAmP libraries. Diploids were selected by replica plating on media containing both G418 and Hygromycin B. After sporulation of diploids, MATa haploids were selected using histidine drop-out media containing

cycloheximide. Haploids harboring both GAA/TTC tract and mutations of interests were further selected by G418 and Hygromycin B containing media. Cells were transferred to canavanine-containing media to score for arm-loss events. (C) Example plate for the screen. Columns are duplicates of query strains. Each row is one tester strain. The level of arm loss in TET-*RFA2* (hyper-fragile), Δ*mrc1* (hyper-fragile), TET-*SEC4* (no change in fragility) and wild-type are shown. (D) Experimental assay to verify results from the screen. *ADE2* is placed between *CAN1* and *LYS2* cassette. Mutations in *CAN1* will manifest as $Can^{R}Ade^{+}$ white colonies, while arm loss events will give rise to $Can^{R}Ade^{-}$ red papillae.

2.3.2 Hyper-fragility mutants identified in the screen

The screen identified 11 hyper-fragility mutants from the YKO collection (Table A.1). After recreating these deletions in the color-based GCR assay, we excluded *Δtsa1* from further analysis because this mutant exhibited a strong mutator phenotype but not increased arm loss (data not shown). The remaining 10 mutants fall into two groups: DSB repair and replication. The first group consists of mutants deficient in Mre11-Rad50-Xrs2 (MRX complex) and Sae2 activities. The MRX complex as well as Sae2 are nucleases involved in DSB end resection. In addition, MRX provides a bridge between the two broken molecules during DSB repair (reviewed in Mimitou and Symington, 2009). Previously we have demonstrated that breaks at GAA/TTC tracts are predominantly repaired via break-induced replication involving GAA-rich stretches on non-homologous chromosomes (Kim et al., 2008). It is therefore conceivable that the lack of extensive resection in *mrx* or *sae2* mutants will allow for improved recovery of broken chromosome V due to increased probability of recombination between GAA regions. Moreover, in the absence of the MRX bridge, the acentric broken fragment of chromosome V can be lost more efficiently than in wild-type strains. Consistent with this, the GCR rates were ~ 6 fold higher in *mrx* mutants than in *sae2* (Table 2.1). Hence,

it is likely that a deficiency in MRX complex and Sae2 is not related to stability of the repeats, but rather reflects a change in the repair dynamics of GAA/TTC-mediated breaks.

The second group includes mutations in Rad27, the flap-endonuclease, Rtt101-Mms1-Mms22, the proteins involved in the repair of stalled replication forks, and Tof1-Csm3-Mrc1, the replication-pausing checkpoint surveillance complex. Disruption of genes that encode for these proteins caused a 5 to 19-fold increase in chromosome V arm loss in strains carrying 120 and 230 GAA/TTC repeats. These data indicate that unperturbed replication progression and checkpoint sensing of the replication problems imposed by triplex structure are critical for the maintenance of GAA/TTC tracts. The list of alleles of the essential genes that altered GAA/TTC fragility (see below) also corroborates this conclusion.

The effect of the decrease in expression of the essential genes on GAA/TTC fragility was assessed using the yTHC and DAmP collections. Screening using the yTHC collection was more successful in identifying hypomorphic alleles that destabilize the repeats. Although during the initial screen doxycycline was used to downregulate expression of the genes, we noticed that the identified strains with TET-*ORFs* exhibited hyper-fragility even without the addition of the drug to the medium (Table A.3). It has been shown that *tetO7*–driven expression leads to an increased level of mRNA production (Belli et al., 1998). However, we found that increase in transcription does not result in elevated protein levels and the TET-*ORFs* behave as recessive alleles (N. Saini and K. Lobachev, in preparation and Table A.2). Therefore, we consider TET-*ORFs* to

be mutant alleles (henceforth referred to as TET alleles). Overall, these mutants can be grouped into three main categories.

First, in agreement with results from the screening of YKO library, 16 hyper-fragility alleles from the essential gene collections belonged to the DNA replication pathway (Table A.1). They fall into 8 distinct complexes or proteins including polymerase α-primase complex, lagging strand polymerase δ, nuclease/helicase Dna2, single-strand DNA binding protein RPA, clamp loader RFC, processivity clamp PCNA, replicative helicase MCM and origin recognition complex ORC. Although Pol2, the leading strand polymerase ε, was not identified in the initial screen using the library, we found that replacing *POL2* promoter with the doxycycline-regulatable promoter leads to hyper-fragility. The effect of the mutant alleles on the GAA/TTC-induced GCRs varied from 2 to 91-fold (Table 2.1). These results indicate that a defect in replication machinery can tremendously affect the fragility potential of the repeats and intact synthesis of both leading and lagging strands is required for preventing the breakage. It is interesting to note the effect of the TET-*ORC* alleles. The *LYS2* region is replicated from the *ARS507* origin located ~26 kb away. It suggests that the dysfunction of origin firing or improper assembly of the replisome in these mutants affects the efficiency of breakage at a distant site.

Second, *ten1-DAmP* hypomorphic allele was found to increase GAA/TTC-mediated GCRs. Ten1 forms a complex with Cdc13 and Stn1 that protects telomeric ends in yeast (reviewed in Giraud-Panis et al., 2010). To determine if the hyper-fragility of *ten1-DAmP* comes from a defect in expression of solely the *TEN1* gene or results from the malfunction of the complex, we placed *TEN1* and *CDC13* under control of the *tetO7–*

promoter. Both TET alleles caused a 4 to 11-fold increase in the fragility of 120 and 230 GAA/TTC repeats (Table 2.1). This suggests that proper maintenance of telomeres, the closest telomere being ~40 kb away from the repeat locus, is important to prevent DSB formation.

Third, besides TET alleles of the essential genes associated with DNA replication, another major group of hyper-fragile alleles comprised of genes encoding for transcription initiation proteins such as components of TFIID, TFIIA, TFIIB and TFIIF complexes (reviewed in Hahn and Young, 2011) (Table 2.1 and A.1). The effect of transcription initiation mutants on GAA/TTC fragility ranged from an 8 to 44-fold increase over the wild-type. In addition, the TET-*SPN1* hyper-fragile allele that codes for a factor that functions in both the recruitment of RNA polymerase II to the promoter and the elongation steps of transcription (Lindstrom et al., 2003), was identified. Although we acknowledge that the screen could have missed some mutants, especially those with mild effect on GCR rate, these data point towards two major pathways that underlie GAA/TTC fragility in yeast: replication and transcription initiation.

Table 2.1. Effect of mutants identified in the genome-wide screen on GAA-induced GCRs

Genetic background	Rate of GCR ($X10^{-7}$)					
	$(GAA)_0$		$(GAA)_{120}$		$(GAA)_{230}$	
wild-type	0.03 (0.02-0.04)[a]		3 (3-4)		20 (10-30)	
Double-strand break repair genes						
Δmre11	36 (27-42)	1200 [b]	126 (110-140)	42	350 (240-400)	18
Δsae2	0.9 (0.2-1)	30	17 (13-24)	6	60 (50-70)	3
Replication genes						
TET-*RFA2*	4 (3-6)	133	272 (93-380)	91	740 (490-910)	37
TET-*POL12*	4 (1-8)	133	65 (46-72)	22	280 (70-580)	14
TET-*PRI2*	1 (1-2)	33	49 (39-51)	16	630 (550-910)	32
TET-*POL3*	0.5 (0.2-0.7)	17	30 (23-40)	10	150 (140-200)	8
TET-*POL2*	0.2 (0.2-0.5)	7	7 (5-8)	2	80 (70-100)	4
TET-*POL30*	0.6 (0.4-1)	20	32 (21-40)	11	370 (290-540)	19
TET-*RFC2*	0.2 (0.1-0.4)	7	30 (9-41)	10	280 (250-860)	14
TET-*DNA2*	0.5 (0.3-0.7)	17	16 (12-21)	5	170 (150-200)	9
TET-*MCM4*	2 (1-3)	67	28 (20-31)	9	150 (90-210)	8
TET-*ORC4*	3 (2-4)	100	26 (19-33)	9	160 (130-220)	8
Δrad27	12 (7-20)	400	32 (26-46)	11	ND [c]	ND
Δrtt101	3 (2-3)	100	15 (12-16)	5	140 (120-360)	7
Δmms1	0.14 (0.1-0.4)	5	16 (12-21)	5	200 (120-830)	10
Δmrc1	4 (3-8)	133	58 (45-66)	19	290 (210-350)	15
Δtof1	2 (0.9-3)	67	35 (29-53)	12	210 (170-230)	11
Telomere maintenance genes						
TET-*TEN1*	0.2 (0.2-0.5)	7	16 (5-28)	5	160 (90-300)	8
TET-*CDC13*	2 (0.8-2)	67	13 (10-18)	4	220 (180-260)	11
Transcription initiation genes						
TET-*TAF4*	0.2 (0.07-0.2)	7	132 (105-170)	44	460 (390-660)	23
TET-*TAF9*	0.1 (0.05-0.2)	3	23 (9-31)	8	190 (130-370)	10
TET-*TOA1*	0.6 (0.4-1)	20	36 (17-44)	12	210 (200-270)	11
TET-*SUA7*	0.2 (0.1-0.4)	7	86 (60-120)	29	390 (370-460)	20
TET-*TFG1*	0.5 (0.2-1)	17	23 (17-27)	8	370 (310-450)	19
TET-*SPN1*	0.3 (0.2-0.4)	10	10 (7-11)	3	160 (100-240)	8

[a] Numbers in parentheses correspond to the 95% confidence intervals

[b] Fold increase in GCR rates in mutants compared to wild-type strains

[c] Not determined

2.3.3 TET-TAF4 and TET-RFA2 strains exhibit increased DSB formation

In order to determine if the increase in GCR levels observed in mutants results from elevated levels of chromosomal breakage, we analyzed the accumulation of DSB intermediates in repeat-containing strains. No breaks were detected in the wild-type strain containing 20 or 230 copies of GAA/TTC repeats when detection of the breakage was carried out in yeast cultures grown overnight. Two mutants with the highest levels of GCR events from replication and transcription initiation categories, TET-*TAF4* and TET-*RFA2*, respectively, accumulated DSBs (Figure 2.2). Densitometry analysis revealed that 12% and 21% of chromosomes V were broken in TET-*TAF4* and in TET-*RFA2* strains, correspondingly.

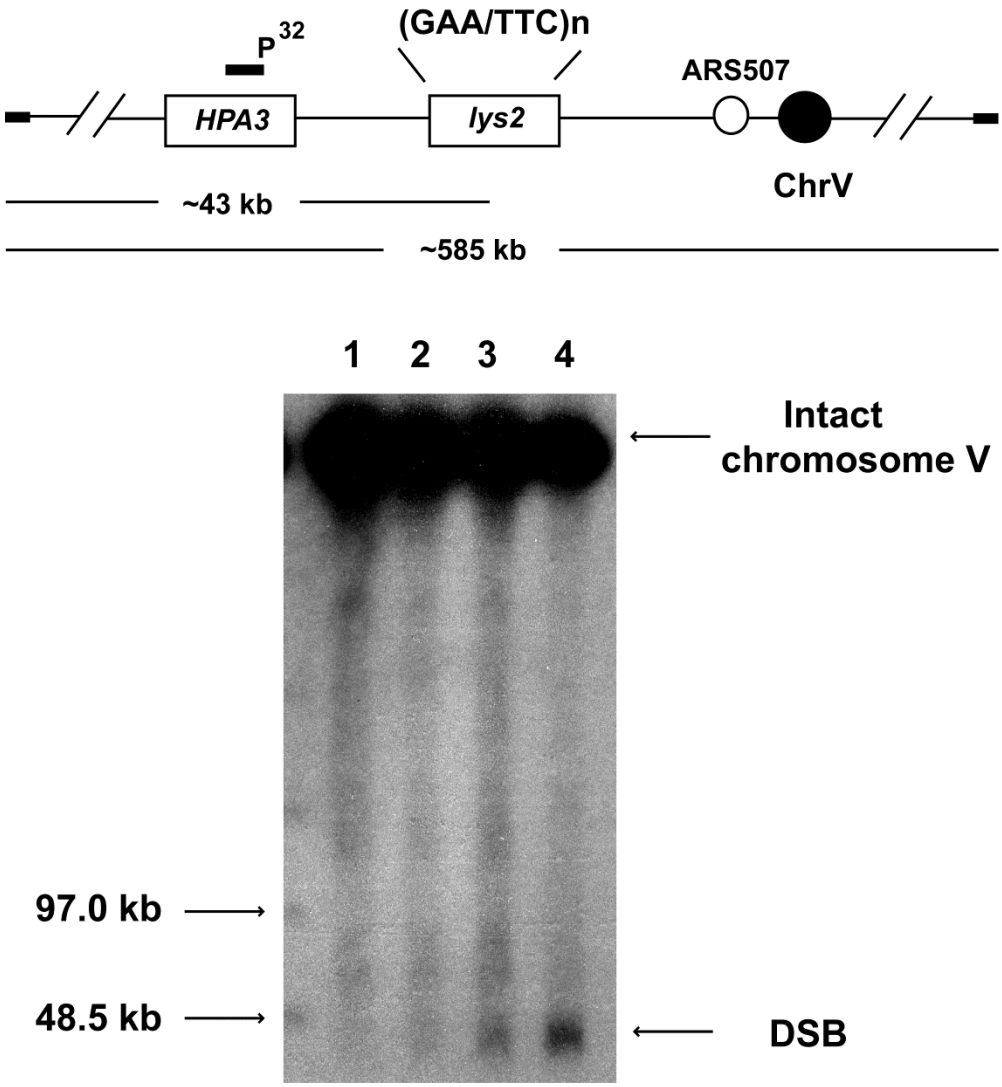

Figure 2.2. Physical detection of GAA/TTC-induced DSBs.

Location of GAA/TTC tracts on chromosome V is depicted. Intact chromosome V is ~585 kb, DSB at the tract results in a 43 kb broken fragment. Chromosomal DNA was separated by CHEF. Both intact and broken chromosomal V were detected by Southern blot hybridization using a *HPA3*-specific probe. The lanes are: 1. wild-type strain with $(GAA)_{20}$; 2. wild-type strain with $(GAA)_{230}$; 3. TET-*TAF4* strain with $(GAA)_{230}$; 4. TET-*RFA2* strain with $(GAA)_{230}$.

2.3.4 GAA/TTC tracts can promote transcription

In the screen for hyper-fragile mutants, we have identified alleles for proteins specifically involved in initiation but not in elongation or termination steps of transcription. It has also been shown that GAA/TTC repeats exclude nucleosomes *in vitro*, perhaps due to the AT-richness of the tracts (Ruan and Wang, 2008). In eukaryotic genomes, AT-rich regions are often loci of open chromatin where transcription can start (reviewed in Hahn and Young, 2011). Based on these observations, we proposed that GAA/TTC tracts can recruit transcription initiation complexes and serve as non-canonical promoter elements in yeast. To test this hypothesis, we replaced the native promoter of the chromosomal *TRP1* gene with GAA/TTC tracts of different lengths. The repeats were placed 56 bps upstream from the ATG codon of *TRP1* in the orientation where during transcription GAA strand will be the sense strand. As expected, strains containing promoter-less *TRP1* ORF did not grow on media lacking tryptophan (Figure A.1). 5 GAA/TTC repeats led to a weak growth of yeast on selective media. Strains with $(GAA)_{20}$ and $(GAA)_{120}$ exhibited normal growth comparable to the strain bearing *TRP1* expressed from its native promoter. Interestingly, growth was inhibited in strains with $(GAA)_{230}$ and $(GAA)_{400}$. Plating serial dilutions on *TRP1* counter–selective media containing 5-fluoroanthranilic acid resulted in reverse growth phenotypes: strains with 5, 120, 230 and 400 repeats grew better than strains containing 20 repeats. Moreover, 20 and 120 repeats placed in another orientation relative to the ATG codon (where TTC strand is the sense strand) also drove expression of *TRP1* (data not shown). Consistent with these data, chromatin immunoprecipitation of TAP-tagged Sua7 protein demonstrated that this TFIIB

transcription initiation factor is associated with $(GAA)_{120}$ tracts with an efficiency comparable to binding to the native *TRP1* promoter.

2.3.5 Fragility is increased in non-dividing cells in a time-dependent manner and is amplified in TET-TAF4 strains

Increased GAA/TTC fragility in replisome-defective mutants suggests that DSBs occur during DNA replication. At the same time, the involvement of transcription-initiation factors in GAA/TTC fragility predicts the existence of a second pathway for breakage where DSBs can be formed outside the S-phase of the cell cycle. To check this premise experimentally, we estimated GCR frequencies in wild-type and TET-*TAF4* strains carrying 120 copies of GAA/TTC repeats in cultures that actively divide during the log phase of growth and in those that are arrested at stationary phase. Yeast were cultured from less than 20 cells to saturation ($\sim 1 \times 10^8$/ml) in liquid YPD and instead of rates, the frequencies of GCRs were calculated (Figure 2.3A and B). Wild-type and TET-*TAF4* strains stop dividing approximately 43 and 49 hours after inoculation, respectively. Cells were held at this stage up to ~ 70 hours. The level of fragility in wild-type and mutant strains increased significantly after they reached stationary phase in comparison to actively dividing cultures (Figure 2.3B and Table A.4). Holding yeast cultures at the arrested stage gradually augmented GAA/TTC fragility. GCR frequencies measured in cultures held arrested for 70 hours were ~ 70-fold higher than those in log-phase cultures for both wild-type and TET-*TAF4* mutants. Overall, fragility frequencies in TET-*TAF4* mutants were 5 to 30-fold higher than in wild-type in both logarithmic and stationary phases.

To obtain direct evidence for time-dependent chromosomal breakage in non-dividing cells, accumulation of DSB intermediates in the log phase and the arrested cultures was monitored (Figure 2.3C). No broken molecules were detected in wild-type strains carrying 20 or 230 GAA/TTC repeats at any analyzed time points. Consistent with the measurements of the GCR rates, in the TET-*TAF4* $(GAA)_{230}$ strain, there was an increase in the amount of broken molecules when cells spent more time in the stationary phase. The fraction of DSBs at the beginning of the stationary phase was 3%, which increased to 13% and 18% at 20 hours and 70 hours after arrest, respectively.

These data demonstrate that fragility occurs and is amplified in non-dividing cells. Elevated fragility in TET-*TAF4* compared to wild-type strains in stationary phase suggests that intact transcription initiation is important in controlling GAA/TTC stability in non-dividing cells.

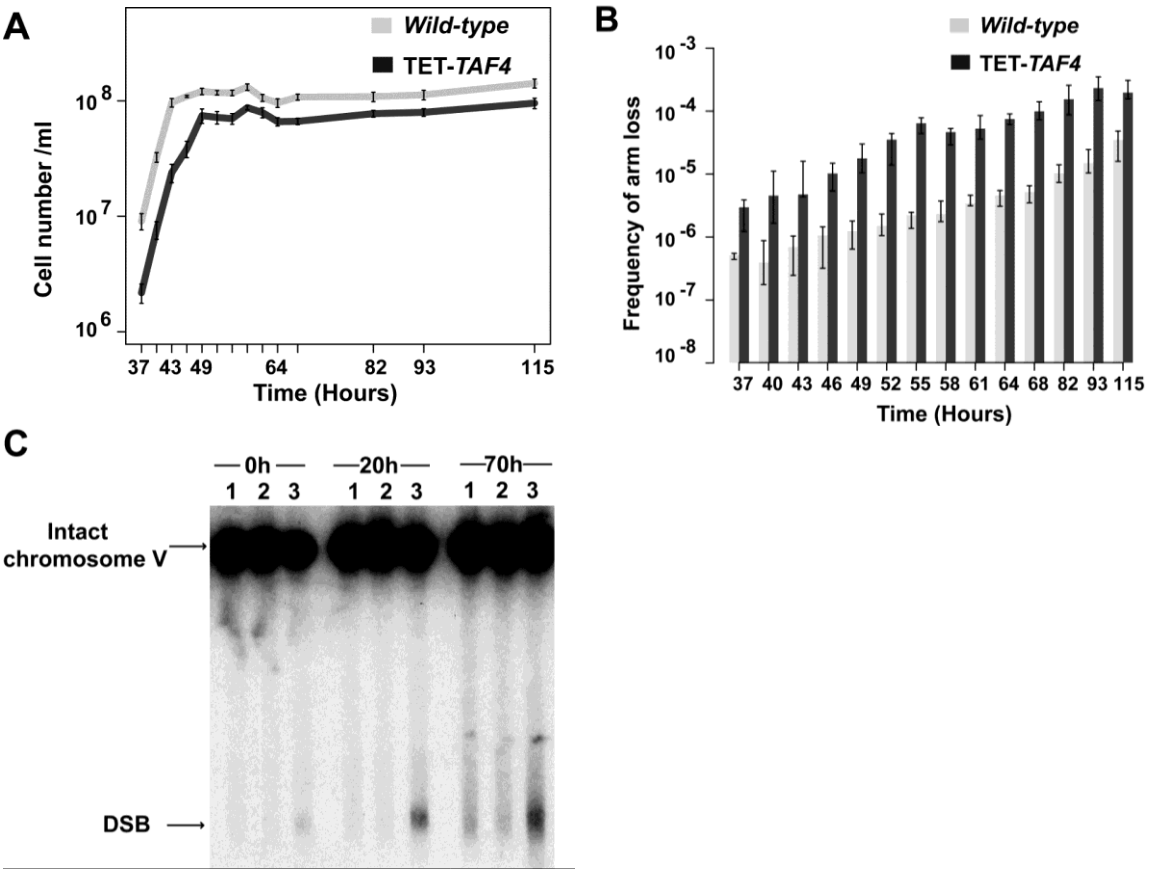

Figure 2.3. GAA/TTC fragility increases in non-dividing cells.

(A) Growth curves of wild-type and TET-*TAF4* strains. Wild-type and TET-*TAF4* strains enter stationary phase at ~ 43 and 49 hours after inoculation, correspondingly. Error bars indicate standard error of the mean. (B) GAA/TTC fragility frequencies of wild-type and TET-*TAF4* strain in dividing and non-dividing cells. Values are the median frequencies obtained from fluctuation tests of at least 8 cultures carried out at the indicated time-points. Error bars denote 95% confidence intervals. (C) Detection of DSBs in wild-type and TET-*TAF4* strains at different time-points during stationary phase. 0 hour corresponds to the time when strains stopped growing. 20 and 70 hours indicate the time each strain spent in stationary phase. The lanes are: 1. wild-type strain with $(GAA)_{20}$; 2. wild-type strain with $(GAA)_{230}$; 3. TET-*TAF4* strain with $(GAA)_{230}$. Positions of DSBs and unbroken chromosome V are indicated by arrows.

2.3.6 Mutations conferring hyper-fragility also induce large-scale repeat expansions

Formation of triplex structure by GAA/TTC repeats was proposed to be a triggering event for both fragility and tract length variations (Kim et al., 2008; Shishkin et al., 2009). Hence, we reasoned that the identified mutants predisposed for GAA/TTC breakage should also be prone to repeat instability including large-scale expansions. To address this experimentally, we introduced a selectable cassette containing 100 GAA/TTC repeats (Shishkin et al., 2009 and Materials and Methods) into the strains with hyper-fragile alleles. The repeat tract is inserted into an artificial intron in the *URA3* gene located 1 kb from the *ARS306* on chromosome III. Expansion of repeats beyond 130 copies leads to the inactivation of RNA splicing blocking *URA3* expression. These events give rise to colonies resistant to 5-fluoroorotic acid (5-FOA^R).

We measured the rates of expansions in colonies that grew for 3 days on YPD plates. A defect in Mre11-Rad50-Xrs2 and Sae2 that caused increase in GCRs events did not change the rate of expansions (Table 2.2 and data not shown). This is in agreement with our interpretation of the effect of these mutations on the GCRs efficiency: *mrx* and *sae2* mutants do not affect the stability of the repeats, but influence the repair step during GCR generation. Remarkably, all the other identified hyper-fragile alleles profoundly affected the ability of GAA/TTC tracts to expand. The degree of impact of the mutant alleles on fragility and expansion also generally correlated. Malfunction of telomere maintenance caused a modest 4 to 11-fold increase in GCRs (Table 2.1) and 12 to 14-fold elevation of the expansion rates (Table 2.2). Along with this, the rate of expansions was elevated 171-fold in the most fragile mutant from DNA replication group, TET-*RFA2* and 58-fold in the most fragile mutant from transcription initiation group, TET-*TAF4*.

To verify that the 5-FOA^R colonies arose due to GAA/TTC expansions, but not due to changes in *URA3* expression levels in transcription initiation mutants, expansions at the repeat locus were confirmed by PCR in both wild-type and TET-*TAF4* strains. In addition, analysis of *URA3* expression using reverse transcriptase PCR showed that wild-type and TET-*TAF4* strains are comparable (data not shown).

Table 2.2. Rates of expansions of $(GAA)_{100}$ in wild-type and hyper-GCR mutants

Genetic background	Expansion rate($X10^{-5}$)	Fold increase
wild-type	1 (0.6-1) [a]	1 [b]
Double-strand break repair genes		
Δsae2	1 (0.6-2)	1
Replication genes		
TET-*RFA2*	171 (84-330)	171
TET-*POL12*	55 (49-82)	55
TET-*PRI2*	110 (77-190)	110
TET-*POL3*	11 (7-25)	11
TET-*POL2*	5 (2-12)	5
TET-*POL30*	26 (16-30)	26
TET-*RFC2*	26 (13-66)	26
TET-*DNA2*	8 (4-12)	8
TET-*MCM4*	23 (17-36)	23
TET-*ORC4*	18 (4-90)	18
Δrad27	20 (9-35)	20
Δrtt101	4 (2-31)	4
Δmms1	3 (2-4)	3
Δtof1	10 (13-24)	10
Telomere protection genes		
TET-*TEN1*	12 (9-17)	12
TET-*CDC13*	14 (10-29)	14
Transcription initiation genes		
TET-*TAF4*	58 (40-68)	58
TET-*TAF9*	37 (32-46)	37
TET-*TOA1*	14 (8-23)	14
TET-*SUA7*	19 (12-23)	19
TET-*TFG1*	22 (17-57)	22
TET-*SPN1*	2 (1-3)	2

[a] Numbers in parentheses correspond to the 95% confidence intervals

[b] Fold increase in rates of expansions in mutants compared to wild-type strains

As demonstrated above, GAA/TTC induced-DSBs accumulate during stationary phase and intact transcription initiation counteracts the fragility. Transcription initiation mutants are also prone to tract expansions. Based on these data, we asked whether large-

scale repeat expansions occur in non-dividing cells during stationary phase. We estimated the expansion frequencies in actively dividing and starvation-arrested cells in wild-type and TET-*TAF4* strains (Figure 2.4 and Table A.5). Interestingly, the expansion frequency in TET-*TAF4* strains increased at the point of transition from actively dividing to the stationary stage. However, unlike fragility dynamics, expansions were not elevated upon holding cultures in the stationary stage, indicating that replication is required for this process. In fact, there was a slight decrease in the expansion frequencies at later time-points in the stationary phase, which is likely due to accumulation of deletions.

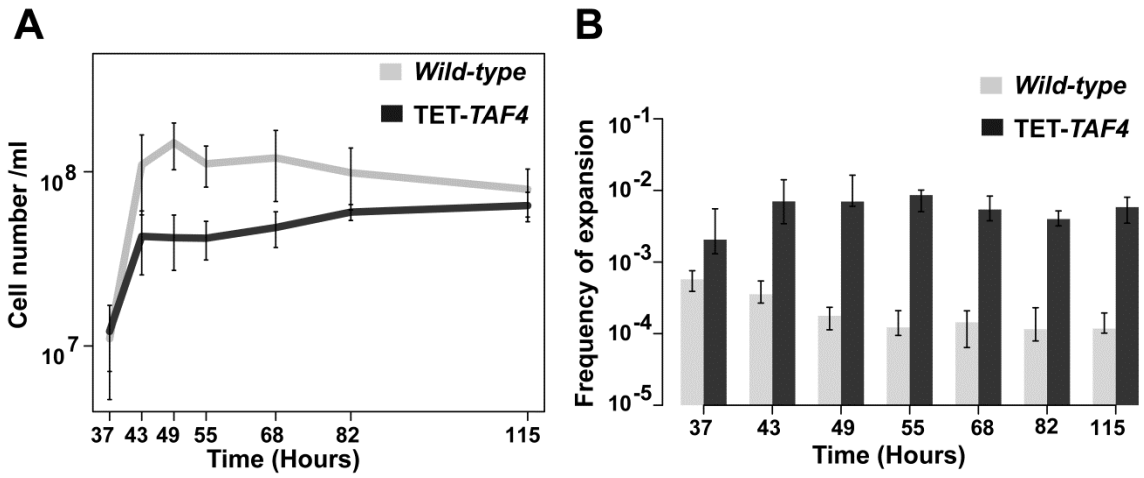

Figure 2.4. Expansion dynamics of $(GAA)_{100}$ repeats in actively dividing and non-dividing cells in wild-type and TET-*TAF4* strains.

(A) Growth curves for wild-type and TET-*TAF4* strains. Error bars represent standard error of the mean. (B) Frequencies of expansion of $(GAA)_{100}$ in dividing and arrested cells. Values are the median frequencies obtained from fluctuation tests of at least 8 cultures. Error bars indicate 95% confidence intervals.

Together, these studies reveal replication and transcription initiation as the two major determinants of GAA/TTC stability in yeast. Both fragility and expansion are

increased in replication and transcription initiation mutants, indicating existence of common intermediates that can drive both types of instability. While fragility can happen in both dividing and non-dividing cells, large-scale expansions are dependent on DNA replication.

2.4 Discussion

In this work, we have isolated 33 mutants prone for GAA/TTC fragility and large-scale expansions in yeast. Moreover, we uncovered the role for GAA/TTC tracts as initiators of transcription. The identity of the mutants, the ability of repeats to promote transcription and the analysis of breakage and expansion dynamics point towards mechanisms that govern GAA/TTC instability during and outside of S-phase. Below we will discuss the results of the genome-wide screen according to the category of the revealed mutants.

2.4.1 Defect in DNA replication

Intact replication machinery and replication-pausing checkpoint surveillance are required to maintain GAA/TTC repeat stability (Table 2.1, 2.2 and A.1). A model of how a defect in DNA replication can influence the propensity of GAA/TTC tracts to break and expand is presented on Figure 2.5I. It is likely that a deficiency in the structural components of the replisome such as Polα–Primase, Polε, Polδ, RFC, PCNA, MCM and Tof1-Csm3-Mrc1, results in the generation of long single-stranded regions on the template for lagging or leading strand synthesis. This creates optimal conditions for triplex formation. Conceivably, depletion of Dna2, Fen1 and PCNA impairs Okazaki fragment maturation and leaves an unprocessed flap that could be folded into a triplex. The secondary structure may be destroyed by components of the replication machinery such as the single-strand DNA-binding protein RPA (Figure 2.5Ia). Otherwise, the triplex could create a strong block for DNA synthesis. Consistent with this, using 2D gel analysis we found that in TET-*RFA2* strains that exhibit the highest levels of fragility and expansions, replication progression through $(GAA)_{120}$ tract is impaired (Figure A.2). Triplex or

triplex-arrested fork might be sensed and removed by the Rtt101-Mms1-Mms22 complex. Since Tof1-Csm3-Mrc1 complex also carries out checkpoint function (reviewed in Tourriere and Pasero, 2007), we can not exclude that triplex-mediated arrest can also be under the surveillance of this replication-pausing machinery. If undetected, the triplex can cause template switch resulting in expansions (Figure 2.5Ib). Alternatively, the secondary structure can be attacked by nucleases resulting in DSBs (Figure 2.5Ic).

From the list of mutants associated with DNA replication that affect GAA/TTC stability it is important to pinpoint the deficiency in the ORC complex. The ORC complex is required for the initiation of DNA replication but it does not travel along with the replisome (reviewed in Prasanth et al., 2004). The increase in breakage and expansions observed in these mutants can be explained in several ways. First, the ORC complex might be loaded at the GAA/TTC tract. This interaction could be promoted either by exposure of ssDNA due to triplex structure formation or the AT rich nature of the repeats. Interestingly, it has been demonstrated in Cos-1 cells that GAA/TTC tract can stimulate initiation of alternative replication on plasmid (Chandok et al., 2012). Based on this observation, it is conceivable that binding of the OCR complex might destroy the triplex structure thereby stabilize the repeats. Second, ORC deficiency might lead to the assembly of a non-canonical and faulty replisome. Improper replication progression can result in the accumulation of persistent single-stranded regions where triplex structures can be formed. Third, in *orc* mutants, the GAA/TTC locus might not be replicated from the closest active origin, *ARS507*. It is possible that the replication machinery travelling from distant origins might be compromised and less processive.

This can mimic what is happening when structural components of the replisome are hampered. The third explanation echoes data obtained in human cells where it was shown that fragility of FRA3B and FRA16D sites is suppressed when replication is initiated in these regions and is increased when the fragile sites are replicated by forks travelling from distant origins (Letessier et al., 2011).

2.4.2 Defect in telomere maintenance

Somewhat surprisingly, in strains carrying *ten1-DAmP,* TET-*TEN1* and TET-*CDC13* alleles, rates of GAA/TTC fragility and expansions increased. Cdc13, Stn1 and Ten1 form a complex that plays a role in chromosome end protection, telomere replication and regulation of telomerase (reviewed in Giraud-Panis et al., 2010). In *cdc13* mutants, telomere uncapping leads to C-rich strand resection, accumulation of single-stranded DNA and activation of the Rad53 and Chk1-dependent checkpoint response (reviewed in Longhese, 2008). One possible explanation among others for the effect of telomere dysfunction on repeat instability could be the sequestration of the replication-pausing checkpoint complex to the uncapped telomeres to counteract Exo1-mediated resection. It has been shown that telomere length is decreased in *Δmrc1* mutants; and *Δmrc1cdc13-1*, *Δtof1cdc13-1* or *Δcsm3cdc13-1* double mutants show a synthetic growth defect (Grandin and Charbonneau, 2007; Tsolou and Lydall, 2007). Moreover, the growth defect and the amount of degradation are reduced in *Δmrc1cdc13-1* strains when Exo1 nuclease is removed (Tsolou and Lydall, 2007). These data indicate that the Mrc1-Tof1-Csm3 complex protects uncapped telomeres by inhibiting DNA resection. Hence, it is possible that telomere uncapping can sequester the replication-pausing complex from the replication fork to single-stranded regions in *ten1-DAmP,* TET-*TEN1* and TET-*CDC13*

strains thereby creating a paucity of Mrc1-Tof1-Csm3 in the replisome. As discussed above, this deficiency can cause accumulation of single-stranded DNA regions at the replication fork or compromise checkpoint response when the replisome encounters the triplex barrier. Also, we can not exclude the possibility of Ten1-Stn1-Cdc13 binding the repeats or the triplex structure and influencing the tract stability.

2.4.3 Defect in transcription initiation

Transcription initiation mutants are the second largest group of mutants that predispose repeats for both breakage and large-scale expansions. In yeast, transcription initiation and start site selection by RNA polymerase II require the activity of several general transcription factors, including TFIIA, TFIIB, TFIID, and TFIIF complexes. TFIID is specifically important for transcription from TATA-less promoters that contain AT-rich regulatory sequences (reviewed in Hahn and Young, 2011). GAA/TTC tracts, similar to poly-AT sequences, bind nucleosomes poorly (Ruan and Wang, 2008; Russell et al., 1983; Struhl, 1985) and thus likely generate nucleosome-depleted regions, a feature of many sites in yeast chromosomes where transcription is initiated (Nagalakshmi et al., 2008; Xu et al., 2009). In this study, we found that Sua7 binds to GAA/TTC tracts (Figure A.1D). Moreover, GAA/TTC repeats serve as a promoter element and drive transcription of the *TRP1* ORF. These data demonstrate that in yeast, GAA/TTC sequences are target sites for binding of the transcription factors that trigger mRNA synthesis. Interestingly, although the underlying mechanism is not known, in *Mycoplasma gallisepticum*, GAA/TTC tracts located upstream of the putative promoter for the M9 gene also promote transcription in a repeat length-dependent manner (Glew et al., 2000 and references therein). In human fibroblast cells, *FXN* antisense transcript was

detected, which is elevated in FRDA patients with expanded alleles, suggesting that these repeats might play a role in transcription initiation (De Biase et al., 2009).

We found that fragility, in contrast to expansions, occurs in non-dividing cells that are preferentially arrested at the G1 stage of the cell cycle. This observation is consistent with data obtained by Tang *et al.,* 2011, who demonstrated that mitotic crossover induced by GAA tracts in diploid cells is likely due to DSB at the G1 stage. Moreover, we showed that fragility and expansions are elevated in transcription initiation mutants during active division and fragility is further amplified in stationary phase. How can a deficiency in transcription initiation machineries destabilize repeat tracts? One possible explanation is that triplex formation outside of the S-phase, likely during the G1 stage of the cell cycle, can be counteracted by transcription initiation factors binding to the GAA/TTC region. In mutants defective in transcription initiation, the triplex can persist and be attacked by a nuclease leading to DSB. If the structure survives until S-phase, template switch (as described above) could result in large-scale expansions. This model leaves the question of how a triplex can be formed in non-dividing cells and does not explain why breaks accumulate while cells are arrested in the stationary phase. Another explanation that accommodates results from other studies includes accelerated accumulation of R-loops in the GAA/TTC region when transcription initiation is abnormal (Figure 2.5II). It has been shown that GAA/TTC tracts are poor substrates for RNA polymerase II progression (Grabczyk and Usdin, 2000; Krasilnikova et al., 2007) and readily accumulate recombinagenic R-loops (McIvor et al., 2010; Reddy et al., 2011). A lack of components of TFIIA, TFIIB, TFIID, and TFIIF complexes can lead to the loading of the hampered transcription elongation machinery. In mutants, abortive

transcription can lead to futile cycles of RNA synthesis inside GAA/TTC tracts, RNA-polymerase trapping, and more efficient R-loop production culminating in increased DNA breakage in non-dividing cells (Figure 2.5IIa). In dividing cells, trapped RNA-polymerase has been shown to be an obstacle for replication progression (Mirkin et al., 2006). Collision of the replication machinery with RNA polymerase at the non-canonical promoter can cause removal of this barrier and unwinding of R-loops (Figure 2.5IIb). Alternatively, replication bypass via template switch can generate large-scale expansions (Figure 2.5IIc) or the block and persistent R-loops can culminate in DSBs (Figure 2.5IId). There are several observations that are in agreement with this model. First, breaks are higher in cells arrested in stationary phase than in cells that actively divide (Figure 2.3). This indicates that intact replication acts as a preventing force for fragility. Second, there is a time-dependent accumulation of GAA/TTC-mediated DSBs during stationary phase. This points towards a dynamic process promoting fragility where transcription is the most likely mechanism. Third, unlike fragility, large-scale expansions do not occur in non-dividing cells and require DNA replication (Figure 2.4). Fourth, increase in the size of repeats from 20 to 400 leads to a decrease in *TRP1* expression level, indicating that longer tracts are more problematic templates for RNA polymerase progression, or longer R-loops prevent efficient translation.

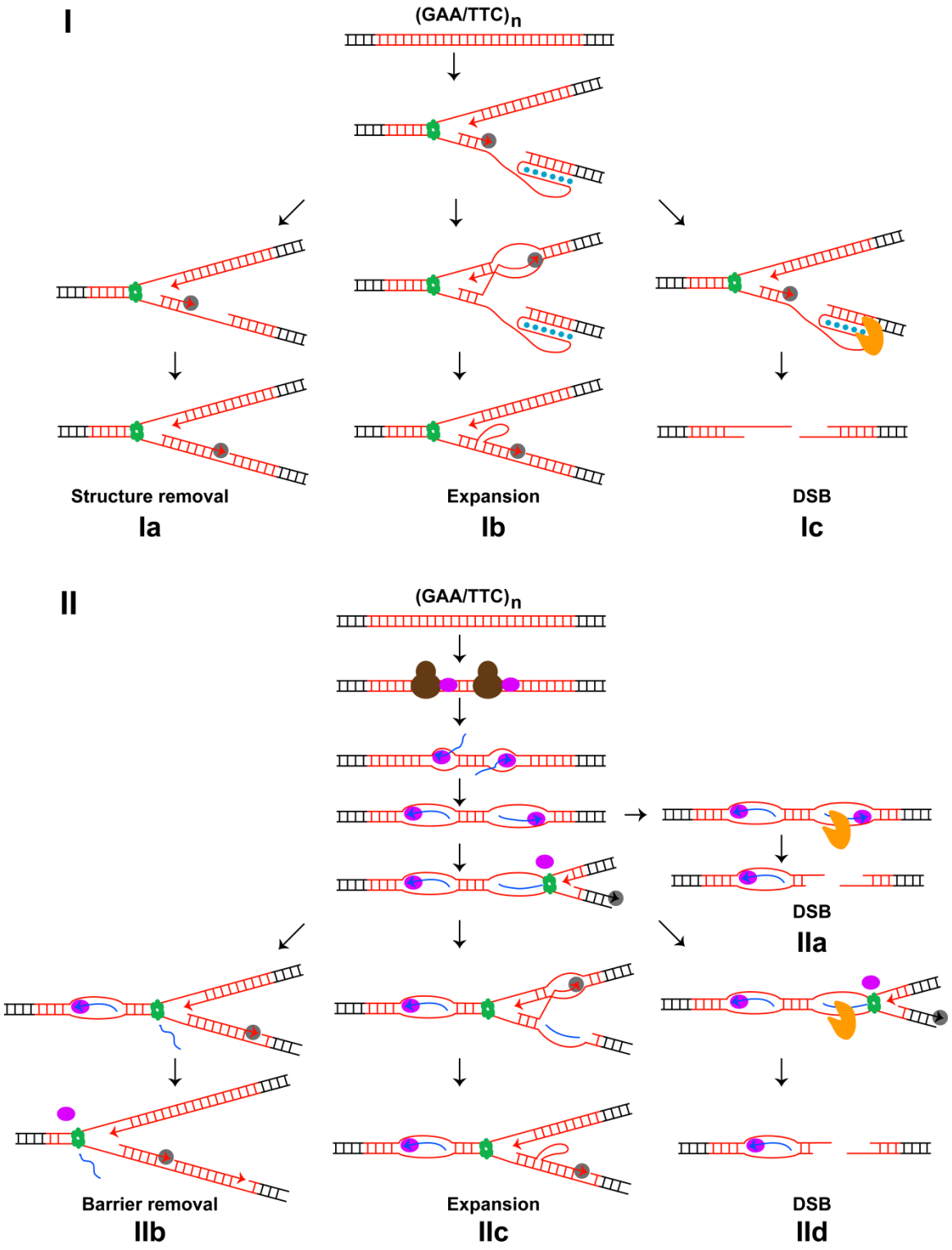

Figure 2.5. Model for GAA/TTC fragility and expansion in dividing and non-dividing cells.

I. Replication-associated pathway for GAA/TTC (red line) instability. Hoogsteen base pairing in the triplex is indicated as blue dots. DNA replication helicase (green ring),

DNA polymerase (solid grey oval) and attacking nuclease (solid orange pacman) are shown. II. Transcription-associated pathway for GAA/TTC instability. Transcription initiation factors (solid brown ovals), the newly synthesized RNA (blue line) forming abnormal RNA-DNA hybrid (R-loop) and RNA polymerase II (solid purple ovals) are depicted. A detailed description is presented in the text.

2.4.4 Conclusions and implications for the human genome stability

We found several pathways that determine GAA/TTC stability and operate in dividing and non-dividing yeast cells. Intact replication machinery is required to prevent increased breakage and expansion of the repeat tracts in dividing cells. In humans, it would indicate that carriers of the hypomorphic alleles of replication-associated genes should be highly predisposed for acquiring GAA/TTC expansions as well as for repeat mediated-chromosomal breakage and rearrangements. Remarkably, we found that proper origin firing and telomere metabolism are also important for maintaining stability of the repeats. Dysfunctions of these processes can contribute to tissue and cell-dependent instability in human cells. It is becoming evident that cells from different tissue types differ in replication initiation patterns and replication timing (Hansen et al., 2010; Ryba et al., 2010). Therefore, the potential of GAA/TTC tracts for instability can be dependent on the location of the repeat tract in the particular tissue-specific replicon. We also propose that GAA/TTC stability should be compromised in cells experiencing telomere crisis such as aging or cancer cells. Our finding that GAA/TTC tracts attract transcription factors and promote mRNA synthesis has dual implications for the metabolism of repeat tracts in human cells. First, R-loop formation in the expanded tracts might not require GAA/TTC region to be transcribed from an outside promoter. This conjecture also suggests that if tracts are located inside an ORF, GAA/TTC repeats can

produce antisense RNA. As proposed by Tufarelli *et al.,* 2006 and McIvor *et al.*, 2010, R-loops and/or antisense transcripts can trigger silencing and might explain accumulation of repressing chromatin markers in regions that flank expanded GAA/TTC repeats (reviewed in Kumari and Usdin, 2012). Second, unlike yeast, in human cells, transcription initiation factors such as Taf proteins can be coded by several genes where particular alleles are expressed in different tissues during differentiation and development (D'Alessio et al., 2009). If these alternate complexes have a different efficiency of transcription initiation in GAA/TTC tracts, this can also contribute to tissue-dependent stability of the repeats.

In humans, repeats can expand in non-dividing neuronal cells (De Biase et al., 2007a) and in confluent embryonic kidney cell lines where it was demonstrated that transcription promotes expansions (Ditch et al., 2009). We found that although fragility of repeats is increased in non-dividing yeast cells and involves transcription process, large-scale expansions require cell proliferation. The question that remains to be answered is whether transcription-associated fragility in non-dividing cells can result in large-scale expansions. The obvious difference between the yeast system used in this study and human cells is ploidy. Whether fragility in arrested yeast cells promotes expansions when a homologous chromosome carrying an additional copy of GAA/TTC tracts is present remains to be tested. The identification of the nuclease that creates breaks in repeats in non-dividing cells is also an important avenue for future studies.

2.5 Materials and Methods

2.5.1 Yeast Strains

All strains used in this study are derivatives of BY4742 (Open Biosystems). The genotype of the query strain, HMK246, used in the screen is: MATα, *ura3*-Δ, *leu2*-Δ, *his3*-Δ, *lys2*-Δ, *rpl28-Q38K*, *mfa1Δ::MFA1pr-HIS3, V34205::lys2::*$(GAA)_{230}$, *V29617::hphMX*. The strain was constructed in several steps. First, in BY4742 *LYS2* was inserted in the left arm of chromosome V telomere distal to *CAN1*. Second, the mating type of this strain was switched to MATa using pJH132 plasmid (gift from Dr. James Haber). The *MFA1pr-HIS3* cassette was PCR amplified from y2454 (Smith et al., 2004) and inserted into the *MFA1* locus. The mating type was switched back to MATα. Third, the *hphMX* cassette was inserted telomere proximal from *CAN1* on chromosome V. Fourth, the *rpl28- Q38K* mutation was obtained by selecting resistant colonies on YPD media supplemented with 5 mg/L cycloheximide. Finally, the $(GAA)_{230}$ tract was moved to the *Bam*HI site of *LYS2* via *delitto perfetto* (Storici et al., 2001).

The three collections of tester strains: yTHC, DAmP, YKO strains, were purchased from Open Biosystems. The strains where fragility was assessed have the following genotype: MATa, *bar1*-Δ, *trp1*-Δ, *his3*-Δ, *ura3*-Δ, *leu2*-Δ, *ade2*-Δ, *lys2*-Δ, *V34205*::*ADE2*, *lys2*::(GAA)n. *LYS2* contains either 120 or 230 of GAA/TTC repeats. As a control, the strain with no repeats was used. To verify the effect of mutants uncovered from the screen, TET-alleles were created by replacing the natural promoters of the essential genes with tetracycline repressible promoters (Belli et al., 1998). pCM225 (Euroscarf) was used as a template for PCR amplification to generate a replacement cassette containg *kanMX*-TETp flanked by 50 bp homology to the region of

integration. Non-essential genes were knocked out by the *kanMX* or *hphMX* cassettes using one-step integration.

For testing GAA repeats' ability to serve as promoter elements, *TRP1* ORF amplified from pFL35 (Bonneaud et al., 1991) was used to replace the promoter part and 3246 bps of *LYS2* 5' region using *delitto perfetto* technique. As a result *TRP1* ORF was positioned 56 bps away from 0, 5, 20, 120 or 230 repeats where GAA strand serves as the sense strand for transcription. The isolate with 400 GAA/TTC repeats in front of *TRP1* ORF was the result of spontaneous expansion of 230 repeats. The size of the expanded repeats is approximate and was estimated by PCR amplification of the region and comparison against a known DNA ladder. *TRP1* ORF was also brought adjacent to 20, and 120 copies of repeats with TTC strand serving as the sense strand for transcription. As a positive control, *TRP1* with the natural promoter was used to replace the promoter and 5' part of *LYS2* in the strain carrying no repeats. After integration, *TRP1* ORF and *TRP1* were sequenced to confirm that no mutations were introduced into *TRP1* during transformation.

The strains for studying large-scale repeat expansions were derived from the strains described in Shishkin et al., 2009. Importantly, it was modified from the originally published cassette by increasing the length of the intron by 269 bps, making its overall length 974 bps. As a result, even relatively small expansions would drive the overall length of the intron above the splicing threshold of ~1.1 kb. The adjusted selectable cassette was created as follows. First, a DNA segment containing the multiple cloning site of pYES3 (Shishkin et al., 2009) was removed by the digestion with *Pdi*I and *Pme*I followed by the vector's re-ligation. Second, a plasmid pTRP1-ISR was generated

by cloning a downstream integration sequence, corresponding to the positions 75641-75895 of chromosome III (*Saccharomyces* genome database) into the *Cla*I site of the modified pYES3. Third, the plasmid pISL-TRP1-ISR was generated by cloning an upstream integration sequence, corresponding to the positions 75227-75594 of chromosome III into the *Zra*I site of the pTRP1-ISR. Fourth, an artificially split *URA3* cassette was amplified from the pYES-Int (Shishkin et al., 2009) and cloned between the *Bgl*II and *Sal*I sites of pISL-TRP1-ISR. (GAA)100 repeat was then cloned between the *Nco*I and *Cla*I fragment of the resultant pISL-URA-TRP1-ISR plasmid. To balance the overall lengths of the intron in the repeat-containing cassettes, a unique 269 bp-long sequence from the coding part of the *tet*R gene of the pACYC184 plasmid (NEB) was amplified and cloned into *Sph*I located in the intron of the split *URA3* gene of the pISL-URA-TRP1-ISR plasmid. The plasmid was digested by *Smi*I and followed by the integration into chromosome III using a selection for tryptophan prototrophy.

2.5.2 Schematics of the genome-wide screen

The scheme was adapted from Tong *et al* (2001) with modifications. MATα query strains marked with *hphMX* and MATa tester strains carrying *kanMX* were patched on YPD plates overnight. Query strains were patched in duplicates. Strains were crossed on YPD overnight and selected on medium containing 300 mg/L G418 and 300 mg/L hygromycin B for diploids. Sporulation was induced by replica plating cells to sporulation medium for 5 days. Yeast were transferred to histidine drop-out plates supplemented with 5 mg/L cycloheximide for 2 days. Since the query strain carries a mating type-specific reporter MFA1pr-*HIS3* and a recessive mutation *rpl28-Q38K* that provides resistance to cycloheximide, this procedure allows selecting for MATa haploid

progeny. This step was repeated to assure the selection of haploids. Haploids harboring both GAA/TTC tracts and mutated genes were selected on medium containing G418 and hygromycin B. Cells were then either replica plated to canavanine plates (for YKO, yTHC and DAmP libraries), or first to medium containing 2 μg/ml doxycyline and then to canavanine plates (for yTHC library). After incubation for 3 days, the effect of the mutants was evaluated based on the amount of revertants on canavanine plates compared with the wild-type level.

2.5.3 Measurement of GAA/TTC fragility and expansion rates

Strains were grown on YPD plates at 30^{o}C for 3 days; a minimum of 14 independent colonies for each strain were taken for fluctuation tests to calculate the rate of fragility or expansions. Each colony was diluted in 250 μl of water and 10-fold serial dilutions were applied to obtain 100-200 colonies on YPD and 50-100 colonies on either canavanine-containing plates for the fragility assay or 5-FOA-containing plates for the expansion assay. Canavanine plates contain 60 mg/L L-canavanine and 5 mg/L adenine; the low amount of adenine is to differentiate mutations at *CAN1* (white colonies) from arm loss events (red colonies). The formula $\mu = f/ln(N\mu)$ was used to calculate the rate of fragility or expansion (Drake, 1991). 95% confidence intervals were calculated as described in Dixon, 1969. Since GAA repeats are prone to size variations, only the colonies with the correct GAA tract sizes (pre-screened by PCR) were used for tests.

For complementation tests, wild-type and TET-*TAF4* strains were transformed with YCp50 or pKL155 containing wild-type *TAF4*, while TET-*POL3* strain was transformed with YCp50 or pBL304 carrying wild-type *POL3* (Morrison et al., 1993). Strains were grown on uracil drop-out plates for 4 days and appropriate dilutions were plated on uracil drop-out media for measuring total number of cells and on uracil drop-out media containing canavanine for estimation of arm loss events.

To estimate rates of fragility when the expression of essential genes is down-regulated, we first determined the optimal concentration of doxycycline that does not cause cell death (data not shown). This concentration, depending on the allele, varied from 0.1 μg/ml to 2 μg/ml of doxycycline in YPD media. Strains were grown on doxycycline plates for 4 days before fluctuation tests were carried out.

For monitoring GAA fragility and expansion in actively dividing and non-dividing cells, ~ 5 yeast cells were inoculated in 5 ml YPD and grown at 30^{o}C. Samples were then taken at indicated time-points and plated on plates for fluctuation tests as mentioned above. A minimum of 8 cultures were used in this experiment for each strain. Wild-type and TET-*TAF4* strains reached saturation in YPD at around 43 hours and 49 hours after inoculation, respectively. Frequencies instead of rates were compared between the actively dividing and arrested cells.

2.5.4 Analysis of yeast growth dynamics

Yeast strains were grown overnight in 5ml of YPD medium. Cultures were washed with distilled water once and resuspended in 10 ml of distilled water. 2 μl cultures were inoculated into 150 μl of SDC, -Trp or 0.75 g/L 5-FAA medium in the 96-well Costar flat bottom plate. The optical density of the culture was measured every 30min using a 600nm filter in the Biotek Synergy H4 plate reader at 30^{o}C with constant shaking. For building the growth curve, the mean of the OD values for each strain was plotted against the time-points used. Four technical and two biological replicates were used for each strain.

2.5.5 DSB detection

Cells were embedded into agarose plugs (0.8% agarose) at a concentration of 24 x 10^{8} cells/ml. Contour-clamped homogeneous electric field (CHEF) gel was used to separate the broken left arm of chromosome V from the intact chromosome V. Chromosomes

were separated at 14^{o}C in a 1% agarose gel in 0.5X TBE for 28 hours at 6V/cm (Bio-Rad CHEF Mapper XA). The included angle was 120^{o}, the initial and final switching times were 12.56 seconds and 17.35 seconds, respectively. The gel was transferred to a nylon membrane at 6 V in 0.5X TBE for 5 hours using the Genie electrophoresis blotter (Idea Scientific). *HPA3*-specific probe was used for Southern hybridization at 67^{o}C overnight. The membrane was washed twice in washing buffer (0.1X SSC, 0.1% SDS) at 70^{o}C and was exposed to Kodak BioMax film. The hybridization signals were quantified using the Carestream Molecular Imaging System software.

2.5.6 Analysis of the ability of GAA/TTC repeats to serve as promoter elements

Strains were grown on YPD for 3 days. The sizes of the repeat tracts were verified before ten-fold serial dilutions were spotted on plates. The plates were incubated at 30^{o}C for 1 day on SDC and -Trp media and for 4 days on 0.75 g/L 5-FAA-containing media.

2.5.7 ChIP and qPCR

TAP-tagged *SUA7* was PCR amplified from *YPR086W* (Open Biosystems) and brought into strains containing no repeats, 120 copies of GAA/TTC repeats, or the natural *TRP1* promoter through one-step integration. Exponentially grown (OD600 = 0.8) cells for each strain were taken for ChIP procedure as described by Aparicio *et al*., 2005. Briefly, cells were cross-linked using 1% formaldehyde followed by 5 min incubation with 2.5M glycine. Cells were then harvested and lysed in a freezer mill. DNA was sheared using a sonicator to give fragments between 500 bp to 1 kb. Chromatin was then immunoprecipitated using Protein A-Sepharose beads. After centrifugation, the supernatants were taken as input samples and the proteins bound to the beads were eluted. The cross-linking was reversed in the eluate and the samples were kept as IP.

The binding of Sua7p with the GAA/TTC repeats and the *TRP1* promoter was measured by qPCR. The samples for PCR were prepared using qPCR kit (Clonetech) in accordance to the manufacturer's instructions. PCR reactions were carried out on Applied Biosystems StepOne™ Real-Time PCR System. Each reaction was done in triplicates. 3' region of *POL5* was amplified and used as an endogenous control for normalization. Comparative CT method using ΔΔCt values was applied for relative quantification. The ratio of PCR amount from IP samples to input genomic DNA was calculated using the formula ratio=$2^{(-\Delta\Delta Ct)}$. Primers information used for strain construction and qPCR are available upon request.

2.5.8 2D analysis of replication intermediates

A colony prescreened for the full size GAA 120 repeats was inoculated into 800 ml YPD and grown overnight. Cells were arrested with alpha factor (50 ng/10^7 cells) at OD600 = 0.8. 2 μg/ml doxycycline was added into YPD during alpha factor arrest for TET-*RFA2* and *TET-TAF4* strains. ~90% cells were arrested in the G1 stage in the case of wild-type and TET-*RFA2* strains. In TET-*TAF4* strains only 70% cells could be arrested. Cells were washed and released into fresh YPD supplemented with pronase. Cells were harvested at 50 min, 50 min or 70 min after release for wild-type, TET-*RFA2* and TET-*TAF4* strains, respectively. Genomic DNA was then extracted as previously described (Friedman and Brewer, 1995). Genomic DNA was digested with *Afl*II such that GAA/TTC tract was positioned on the long arm of the replication arc. DNA was run in 0.4% agarose gel for 22 hours at room temperature in the first dimension and in a 1.2% agarose gel containing 0.3 μg/ml ethidium bromide for 10.5 hours at 4^{o}C in the second dimension. Southern hybridization was carried out using a *LYS2*-specific probe at 64^{o}C. Quantification of 2D gels was carried out as previously described (Krasilnikova et al., 2004).

2.6 Acknowledgements

We thank Dr. N. Degtyareva, Dr. B. Hammer and Z. Sheng for critical reading of the manuscript and helpful discussions. This work was supported by award number R01GM0825950 from NIGMS/NIH and MCB-0818122 from NSF to K.S.L. and R01GM60987 from NIGMS/NIH to S.M.M. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIGMS/NIH or the NSF.

2.7 References

Aguilera, A., and Gomez-Gonzalez, B. (2008). Genome instability: a mechanistic view of its causes and consequences. Nat Rev Genet *9*, 204-217.

Aparicio, O., Geisberg, J.V., Sekinger, E., Yang, A., Moqtaderi, Z., and Struhl, K. (2005). Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. Curr Protoc Mol Biol *Chapter 21*, Unit 21 23.

Belli, G., Gari, E., Aldea, M., and Herrero, E. (1998). Functional analysis of yeast essential genes using a promoter-substitution cassette and the tetracycline-regulatable dual expression system. Yeast *14*, 1127-1138.

Belotserkovskii BP, Liu R, Tornaletti S, Krasilnikova MM, Mirkin SM, Hanawalt PC. (2010). Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. Proc Natl Acad Sci USA 107:12816-12821.

Bonneaud, N., Ozier-Kalogeropoulos, O., Li, G.Y., Labouesse, M., Minvielle-Sebastia, L., and Lacroute, F. (1991). A family of low and high copy replicative, integrative and single-stranded S. cerevisiae/E. coli shuttle vectors. Yeast *7*, 609-615.

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

Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., *et al.* (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science *271*, 1423-1427.

Chandok GS, Patel MP, Mirkin SM, Krasilnikova MM. (2012). Effects of Friedreich's ataxia GAA repeats on DNA replication in mammalian cells. Nucleic Acids Res. *40*, 3964-3974.

Clark, R.M., Bhaskar, S.S., Miyahara, M., Dalgliesh, G.L., and Bidichandani, S.I. (2006). Expansion of GAA trinucleotide repeats in mammals. Genomics *87*, 57-67.

Clark, R.M., Dalgliesh, G.L., Endres, D., Gomez, M., Taylor, J., and Bidichandani, S.I. (2004). Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. Genomics *83*, 373-383.

Clark, R.M., De Biase, I., Malykhina, A.P., Al-Mahdawi, S., Pook, M., and Bidichandani, S.I. (2007). The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. Human genetics *120*, 633-640.

D'Alessio, J.A., Wright, K.J., and Tjian, R. (2009). Shifting players and paradigms in cell-specific transcription. Mol Cell *36*, 924-931.

De Biase, I., Rasmussen, A., Endres, D., Al-Mahdawi, S., Monticelli, A., Cocozza, S., Pook, M., and Bidichandani, S.I. (2007a). Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients. Ann Neurol *61*, 55-60.

De Biase, I., Rasmussen, A., Monticelli, A., Al-Mahdawi, S., Pook, M., Cocozza, S., and Bidichandani, S.I. (2007b). Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. Genomics *90*, 1-5.

De Biase I, Chutake YK, Rindler PM, Bidichandani SI. (2009). Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription. PLoS One. 4, 11, e7914.

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

Dixon, W.J., and Massey, F.J. Jr. (1969). Introduction to statistical analysis. . ed. (New York: McGraw-Hill, Inc.), p. 349.

Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci U S A *88*, 7160-7164.

Frank-Kamenetskii, M.D., and Mirkin, S.M. (1995). Triplex DNA structures. Annual review of biochemistry *64*, 65-95.

Friedman, K.L., and Brewer, B.J. (1995). Analysis of replication intermediates by two-dimensional agarose gel electrophoresis. Methods in enzymology *262*, 613-627.

Giraud-Panis, M.J., Teixeira, M.T., Geli, V., and Gilson, E. CST meets shelterin to keep telomeres in check. (2010) Mol Cell *39*, 665-676.

Glew, M.D., Browning, G.F., Markham, P.F., and Walker, I.D. (2000). pMGA phenotypic variation in Mycoplasma gallisepticum occurs in vivo and is mediated by trinucleotide repeat length variation. Infect Immun *68*, 6027-6033.

Grabczyk, E., Mancuso, M., and Sammarco, M.C. (2007). A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. Nucleic Acids Res *35*, 5351-5359.

Grabczyk, E., and Usdin, K. (2000). The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. Nucleic acids research *28*, 2815-2822.

Grandin, N., and Charbonneau, M. (2007). Mrc1, a non-essential DNA replication protein, is required for telomere end protection following loss of capping by Cdc13, Yku or telomerase. Mol Genet Genomics *277*, 685-699.

Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705-736.

Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M., and Stamatoyannopoulos, J.A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A *107*, 139-144.

Kassai-Jager, E., Ortutay, C., Toth, G., Vellai, T., and Gaspari, Z. (2008). Distribution and evolution of short tandem repeats in closely related bacterial genomes. Gene *410*, 18-25.

Kim, H.M., Narayanan, V., Mieczkowski, P.A., Petes, T.D., Krasilnikova, M.M., Mirkin, S.M., and Lobachev, K.S. (2008). Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. EMBO J *27*, 2896-2906.

Krasilnikova, M.M., and Mirkin, S.M. (2004). Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. Molecular and cellular biology *24*, 2286-2295.

Krasilnikova MM, Kireeva ML, Petrovic V, Knijnikova N, Kashlev M, Mirkin SM. (2007). Effects of Friedreich's ataxia (GAA)n*(TTC)n repeats on RNA synthesis and stability. Nucleic Acids Res. *35*, 1075-1084.

Kumari D, Usdin K. (2012). Is Friedreich ataxia an epigenetic disorder? Clin Epigenetics. *4*, 2.

Letessier, A., Millot, G.A., Koundrioukoff, S., Lachages, A.M., Vogt, N., Hansen, R.S., Malfoy, B., Brison, O., and Debatisse, M. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. Nature *470*, 120-123.

Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., 3rd, and Hartzog, G.A. (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. Mol Cell Biol *23*, 1368-1378.

Longhese, M.P. (2008). DNA damage response at functional and dysfunctional telomeres. Genes Dev *22*, 125-140.

McIvor, E.I., Polak, U., and Napierala, M. (2010). New insights into repeat instability: role of RNA*DNA hybrids. RNA Biol *7*, 551-558.

Mimitou, E.P., and Symington, L.S. (2009). DNA end resection: many nucleases make light work. DNA Repair (Amst) *8*, 983-995.

Mirkin EV, Castro Roa D, Nudler E, Mirkin SM. (2006) Transcription regulatory elements are punctuation marks for DNA replication. Proc Natl Acad Sci USA 103:7276-7281.

Morrison, A., Johnson, A.L., Johnston, L.H., and Sugino, A. (1993). Pathway correcting DNA replication errors in Saccharomyces cerevisiae. The EMBO journal *12*, 1467-1473.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344-1349.

Pollard, L.M., Sharma, R., Gomez, M., Shah, S., Delatycki, M.B., Pianese, L., Monticelli, A., Keats, B.J., and Bidichandani, S.I. (2004). Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. Nucleic acids research *32*, 5962-5971.

Prasanth, S.G., Mendez, J., Prasanth, K.V., and Stillman, B. (2004). Dynamics of pre-replication complex proteins during the cell division cycle. Philos Trans R Soc Lond B Biol Sci *359*, 7-16.

Reddy, K., Tam, M., Bowater, R.P., Barber, M., Tomlinson, M., Nichol Edamura, K., Wang, Y.H., and Pearson, C.E. (2011). Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. Nucleic Acids Res *39*, 1749-1762.

Rindler, P.M., and Bidichandani, S.I. (2011). Role of transcript and interplay between transcription and replication in triplet-repeat instability in mammalian cells. Nucleic Acids Res *39*, 526-535.

Ruan, H., and Wang, Y.H. (2008). Friedreich's ataxia GAA.TTC duplex and GAA.GAA.TTC triplex structures exclude nucleosome assembly. J Mol Biol *383*, 292-300.

Russell, D.W., Smith, M., Cox, D., Williamson, V.M., and Young, E.T. (1983). DNA sequences of two yeast promoter-up mutants. Nature *304*, 652-654.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res *20*, 761-770.

Shishkin, A.A., Voineagu, I., Matera, R., Cherng, N., Chernet, B.T., Krasilnikova, M.M., Narayanan, V., Lobachev, K.S., and Mirkin, S.M. (2009). Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol Cell *35*, 82-92.

Smith, S., Hwang, J.Y., Banerjee, S., Majeed, A., Gupta, A., and Myung, K. (2004). Mutator genes for suppression of gross chromosomal rearrangements identified by a genome-wide screening in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *101*, 9039-9044.

Soragni, E., Herman, D., Dent, S.Y., Gottesfeld, J.M., Wells, R.D., and Napierala, M. (2008). Long intronic GAA*TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia. Nucleic Acids Res *36*, 6056-6065.

Storici, F., Lewis, L.K., and Resnick, M.A. (2001). In vivo site-directed mutagenesis using oligonucleotides. Nat Biotechnol *19*, 773-776.

Struhl, K. (1985). Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. Proc Natl Acad Sci U S A *82*, 8419-8423.

Tang, W., Dominska, M., Greenwell, P.W., Harvanek, Z., Lobachev, K.S., Kim, H.M., Narayanan, V., Mirkin, S.M., and Petes, T.D. (2011). Friedreich's ataxia (GAA)n*(TTC)n repeats strongly stimulate mitotic crossovers in Saccharomyces cerevisae. PLoS Genet *7*, e1001270.

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., *et al.* (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364-2368.

Tourriere, H., and Pasero, P. (2007). Maintenance of fork integrity at damaged DNA and natural pause sites. DNA Repair (Amst) *6*, 900-913.

Tsolou, A., and Lydall, D. (2007). Mrc1 protects uncapped budding yeast telomeres from exonuclease EXO1. DNA Repair (Amst) *6*, 1607-1617.

Tufarelli, C. (2006). The silence RNA keeps: cis mechanisms of RNA mediated epigenetic silencing in mammals. Philos Trans R Soc Lond B Biol Sci *361*, 67-79.

Vetcher, A.A., and Wells, R.D. (2004). Sticky DNA formation in vivo alters the plasmid dimer/monomer ratio. The Journal of biological chemistry *279*, 6434-6443.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033-1037.

CHAPTER 3

GENOME-WIDE SCREEN REVEALS REPLICATION PATHWAY FOR QUASI-PALINDROME FRAGILITY DEPENDENT ON HOMOLOGOUS RECOMBINATION

3.1 Summary

Inverted repeats capable of forming hairpin and cruciform structures present a threat to chromosomal integrity. They induce double strand breaks, which lead to gross chromosomal rearrangements, the hallmarks of cancers and hereditary diseases. Secondary structure formation at this motif has been proposed to be the driving force for the instability, albeit the mechanisms leading to the fragility are not well-understood. We carried out a genome-wide screen to uncover the genetic players that govern fragility of homologous and homeologous *Alu* quasi-palindromes in the yeast *Saccharomyces cerevisiae*. We found that depletion or lack of components of the DNA replication machinery, proteins involved in Fe-S cluster biogenesis, the replication-pausing checkpoint pathway, the telomere maintenance complex or the Sgs1-Top3-Rmi1 dissolvasome augment fragility at *Alu*-IRs. Rad51, a component of the homologous recombination pathway, was found to be required for replication arrest and breakage at the repeats specifically in replication-deficient strains. These data demonstrate that Rad51 is required for the formation of breakage-prone secondary structures in situations when replication is compromised while another mechanism operates in DSB formation in replication-proficient strains.

3.2 Introduction

Long palindromic sequences (inverted repeats without a spacer or with a short spacer) present a threat to both prokaryotic and eukaryotic genome stability. In *E. coli*, long palindromes placed on plasmids are frequently excised and cause cell inviability when introduced to chromosome by phage lambda (Leach, 1994). In yeast, they have been shown to drastically induce ectopic and allelic recombination and a variety of gross chromosomal rearrangements (GCRs) including deletions, translocations and gene amplification (Farah *et al*, 2002; Gordenin *et al*, 1993; Lemoine *et al*, 2005; Lobachev *et al*, 2002; Lobachev *et al*, 1998; Lobachev *et al*, 2000; Narayanan *et al*, 2006; Ruskin & Fink, 1993). Long inverted repeats were demonstrated to undergo frequent deletions and induce gene conversion and intra-chromosomal recombination in mice (Akgun *et al*, 1997; Collick *et al*, 1996; Waldman *et al*, 1999). Palindromic sequences have been found in the vicinity of chromosomal breakpoints of translocations in humans and are implicated in the pathogenesis of diseases. For example, palindromic AT-rich repeats (PATRRs) have been shown to induce both non-recurrent and recurrent translocations; the latter could result into Emanuel syndrome (Gotter *et al*, 2007; Gotter *et al*, 2004; Kehrer-Sawatzki *et al*, 1997; Kurahashi *et al*, 2003; Nimmakayalu *et al*, 2003; Sheridan *et al*, 2010). Palindrome-mediated large deletions and interchromosomal insertions are causative factors of several types of εγδβ thalassemia (Rooks *et al*, 2012) and X-linked congenital hypertrichosis syndrome, respectively (Zhu *et al*, 2011). Also, palindromes are abundant in cancer cells and are associated with DNA amplification in colon and breast cancer, medulloblastoma and lymphoma (Ford & Fried, 1986; Guenthoer *et al*,

2012; Mangano *et al*, 1998; Neiman *et al*, 2008; Neiman *et al*, 2006; Tanaka *et al*, 2005; Tanaka *et al*, 2006; Tanaka *et al*, 2007).

Palindromic sequences can form hairpin and cruciform structures due to their intrinsic symmetry (Leach, 1994). Formation of these aberrant structures has been considered to be responsible for the genetic instability associated with this sequence motif. Hairpins occurring on the lagging strand can interfere with DNA replication and be attacked by structure-specific nucleases leading to DSBs. In *E. coli*, hairpins formed during DNA replication at long palindromic repeats are cleaved by the SbcDC nuclease (Connelly & Leach, 1996; Cromie *et al*, 2000; Darmon *et al*, 2010; Eykelenboom *et al*, 2008; Leach *et al*, 1997). Similarly, in *S. pombe*, the nuclease activity of the Mre11/Rad50/Nbs1complex (Mre11/Rad50 is the homolog of SbcDC) was implicated in the generation of breaks at palindromes (Farah *et al*, 2005; Farah *et al*, 2002). However, Casper et al. (2009) showed that in *S. cerevisiae*, the Mre11 complex is not involved in breakage at a large inverted repeat consisting of two Ty1 elements with a ~280 bp spacer in strains where DNA polymerase α was down-regulated. We previously demonstrated that in *S. cerevisiae*, the Mre11/Rad50/Xrs2 complex does not initiate DSBs at closely spaced *Alu* inverted repeats (*Alu*-IRs) but is required along with Sae2p for processing breaks that have hairpin termini (Lobachev *et al*, 2002). This disparity in the Mre11 complex's effect on DSB generation at palindromic sequences might be attributed to the difference in the formation of stable hairpins during replication and the inability of this complex to cleave hairpins with large loops. This conjecture, however, remains to be experimentally proven. These observations also point out the existence of an Mre11-independent pathway in generating DSBs at palindromic sequences. We proposed that in

yeast, *Alu*-IR-mediated hairpin-capped breaks can result from the resolution of cruciform structures in which a putative nuclease cleaves symmetrically at the base of the two hairpins (Lobachev *et al*, 2002). Cruciform resolution on plasmid in yeast was shown to be dependent on the structure-specific endonuclease Mus81/Mms4 (Cote & Lewis, 2008), although chromosomal fragility at inverted repeats was not influenced by this complex (Lobachev *et al*, 2002). Cruciform formation and resolution were also proposed to be the triggering events for translocations at PATRRs in human sperm cells (Kogo *et al*, 2007; Kurahashi & Emanuel, 2001; Kurahashi *et al*, 2006). Recently, in a plasmid transfection assay, the GEN1 nuclease was implicated in cruciform resolution in HEK293 cells, and the resultant hairpin-capped breaks were further processed by Artemis for DSB repair (Inagaki *et al*, 2013). Whether this mechanism operates in PATRR-mediated chromosomal translocations remains to be established.

Although the formation of hairpin and cruciform structures is deemed the key initiation event for fragility at inverted repeats, the pathways that predispose eukaryotic cells to or provide protection against chromosomal breaks are still not well defined. Previously, deficiencies in Pol1, Pol3 and Rad27 proteins responsible for synthesis of the lagging strand during DNA replication were found to augment instability at inverted repeats (Lemoine *et al*, 2005; Lobachev *et al*, 2000; Ruskin & Fink, 1993). However, it is unknown if fragility is exclusively confined to deficiencies in lagging strand synthesis. In addition, it is important to identify mechanisms that facilitate or prevent instability of imperfect IRs that contain a spacer (quasi-palindrome) and are not fully homologous to each other, since these repeats prevail over perfect palindromes in the human genome (Lobachev *et al*, 2000; Stenger *et al*, 2001).

In this study, we carried out an unbiased genome-wide screen aimed at identifying the genetic factors controlling fragility of homologous and divergent *Alu*-quasi-palindromes in yeast. Using 12 bp-spaced *Alu*-IRs with either 100% or 94% homology between the two repeats, we analyzed the effects of deletions of around 4800 non-essential genes and downregulation of 800 essential genes on quasi-palindrome-mediated GCRs. In addition to defects in lagging strand synthesis, we found that deficiencies in proteins involved in replication initiation and leading strand synthesis, replication pausing checkpoint pathway, the Sgs1-Top3-Rmi1 dissolvasome, proteins involved in Fe-S cluster biogenesis or telomere maintenance augment breakage and GCRs induced by *Alu*-IRs. Replication block and fragility at inverted repeats in replication-deficient strains were abrogated upon deletion of *RAD51*, indicating an unexpected role for homologous recombination in the formation of cruciform structure at palindromic repeats when replication is compromised.

3.3 Results

3.3.1 Experimental systems used in the genome-wide screen

We systematically analyzed the effect of more than 6000 mutations on *Alu*-IR-mediated fragility using a genome-wide screen in the yeast *S. cerevisiae* (Figure 3.1 and Supplemental Figure B.1). The screen's scheme is based on the approach developed in Tong et al. (2001) with modifications. In the query strains, a quasi-palindrome consisting of two 320 bp *Alu* elements in inverted orientation with a 12 bp spacer was placed telomere-distal to the counterselectable marker *CAN1* on the left arm of chromosome V. The two *Alu* elements were either 100% or 94% homologous (100% *Alu*-IRs or 94% *Alu*-IRs). Breakage at the *Alu*-IRs and loss of the 40 kb telomere-proximal fragment results in canavanine-resistant colonies. The tester strains included a complete set of 4786

deletion mutations for non-essential genes (YKO strains) and two sets of 842 essential genes whose expression is either regulated by the doxycycline-repressible promoter (yTHC strains) or decreased due to mRNA perturbation (DAmP strains). An *hphMX* cassette was positioned telomere-proximal to the *Alu*-IRs, providing a marker for selecting the presence of the repeats during the screening and the testers were marked by a *kanMX* cassette. The schematics for combining the left arm of chromosome V containing the fragile motifs and the mutations have been previously applied to study instability of the trinucleotide GAA/TTC repeats and are described in detail in Zhang et al. (2012). Briefly, the query strains were crossed with each tester strain to get diploids, which then underwent sporulation. Haploids containing both the *Alu*-IRs and the mutation of interest were replica plated to canavanine-containing medium. Mutants with augmented repeat-induced GCRs exhibited increased number of canavanine-resistant papillae compared to the wild-type strains. Since the rate of canavanine-resistant colonies occurring due to GCR in the wild-type strain carrying 100% *Alu*-IRs is 10-fold higher ($5x10^{-5}$) than in the strongest mutator Δ*msh2* ($6x10^{-6}$), the screen specifically identified hyper-fragility mutants.

We verified the effect of the identified mutants by recreating the hyper-fragile alleles in strains with the *ADE2* gene inserted between *CAN1* and *Alu*-IRs that allows differentiation of GCRs from mutations based on the color of canavanine-resistant clones (Narayanan *et al*, 2006) (Figure 3.1). To create the mutant alleles, the *kanMX* cassette was used to knockout non-essential genes and a $tetO_7$ repressible promoter was used to replace the natural promoters of essential genes and regulate their expression (Belli *et al*,

1998). The essential genes under the control of *tetO*$_7$ promoter will be referred to as TET-*ORF*s in the following text.

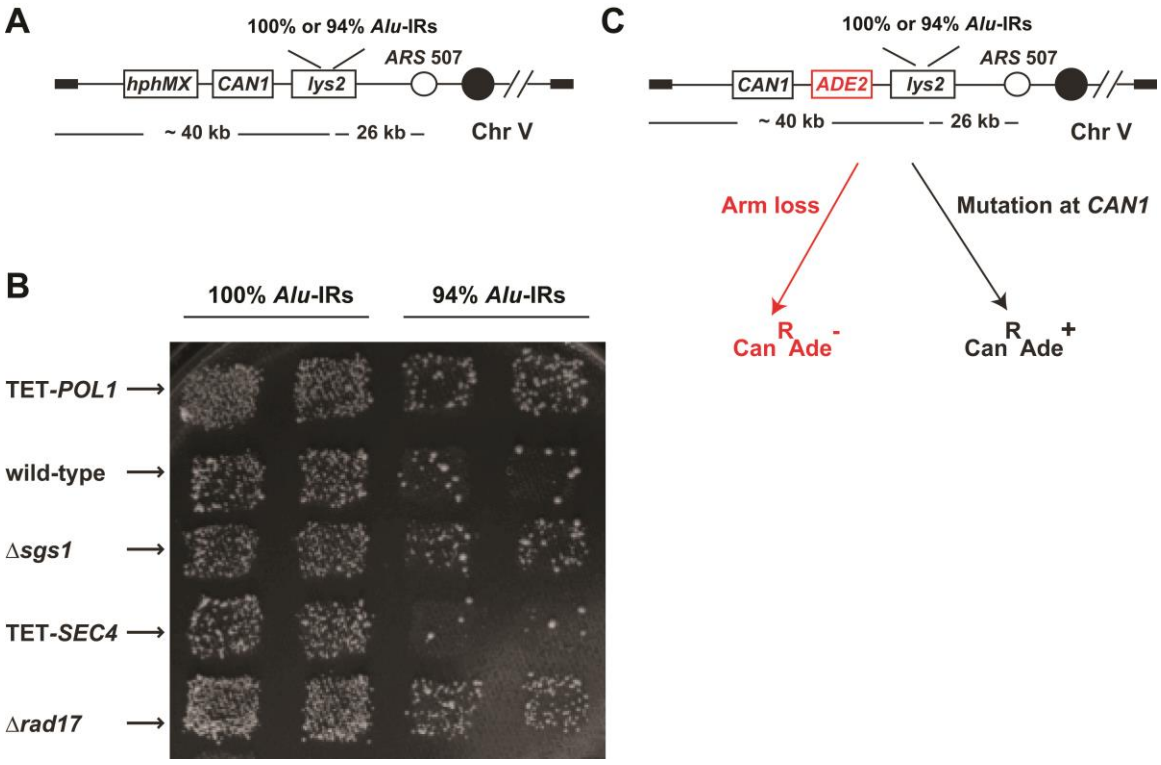

Figure 3.1. The genome-wide screen to identify hyper-GCR mutants.

(**A**) Experimental construct in the query strains. 100% homologous or 94% homeologous *Alu*-IRs were inserted into the left arm of chromosome V. The selectable marker *hphMX* for the presence of the chromosomal fragment containing the repeats and the counterselectable marker *CAN1* used to assay GCR events are depicted. (**B**) Representative plate showing papillae on canavanine-containing medium reflecting the levels of *Alu*-IR-induced GCRs in wild-type and mutants. Columns are duplicates of query strains containing 100% homologous or 94% homeologous *Alu*-IRs. Each row is a tester strain containing the corresponding mutation. (**C**) Experimental construct for verifying the effect of hyper-GCR mutants obtained from the screen. *ADE2* was inserted between *CAN1* and the repeats. As a result, GCR events appear as red Can^R colonies and mutations at *CAN1* give rise to white Can^R colonies on canavanine-containing medium with low amounts of adenine.

3.3.2 Mutants with increased fragility at Alu-IRs

38 mutants that exhibit a hyper-fragility phenotype in strains containing either 100% or 94% homologous *Alu*-IRs were identified from the screen. 17 mutants belonged to the YKO collection, 17 mutants were uncovered from the yTHC collection and 4 mutants were identified from the DAmP collection. The mutants could be grouped into six classes of genes coding for the dissolvasome and proteins involved in replication, Fe-S cluster biogenesis, checkpoint response, telomere maintenance and DSB repair.

Previously, it has been shown that downregulation of or mutation in the DNA polymerases α and δ causes increased instability of inverted repeats (Lemoine *et al*, 2005; Lobachev *et al*, 2002; Ruskin & Fink, 1993). Consistently, we found that TET-*POL1* and TET-*POL3* strains destabilize both 100% and 94% *Alu*-IRs and exhibit 11- to 20-fold higher fragility than the wild-type strains. This screen also revealed that downregulation or deletion of other key components of the DNA replication pathway, namely, the origin recognition complex ORC, the DNA helicase Mcm2-7, the DNA primase complex, the leading strand synthesis polymerase ε, the single-strand binding protein RPA, the polymerase sliding clamp PCNA, the clamp loader RFCs or the endonucleases Dna2 and Rad27 participating in Okazaki fragment maturation, induce fragility at *Alu*-IRs. Deficiencies in these proteins caused a 3- to 12-fold and a 3- to 34-fold increase in GCR rates for 100% *Alu*-IRs and 94% *Alu*-IRs, respectively. We also observed a 4- to 9-fold elevation of GCRs in strains carrying the defective replication checkpoint surveillance complex, Mrc1-Tof1-Csm3 . These data demonstrate that intact replication machinery and replication checkpoint are required to prevent palindrome

instability. Moreover, secondary structure formation and breakage are not only restricted to defects in lagging strand synthesis since fragility is also increased in strains where Polε and Mcm2-7 complex were downregulated.

Besides the replication checkpoint surveillance mutants, the screen also revealed that GCRs mildly increase (2- to 3-fold) in *Δrad17*, *Δmec3*, *Δddc1* and *Δrad24* mutants deficient in DNA damage checkpoint signaling (Navadgi-Patil & Burgers, 2009). As discussed below, this effect could be explained by the improved recovery of the broken chromosome when checkpoint activation is impaired.

The third group of mutants that amplify *Alu*-IRs fragility included members of the cytosolic iron-sulfur protein assembly targeting complex. TET-*YHR122W* led to a 3- and 8-fold increase in GCRs in 100% and 94% *Alu*-IRs, respectively. Yhr122w was shown to physically interact with Cia1 and Mms19 in the biogenesis of Fe-S clusters in various DNA repair and replication proteins (Gari *et al*, 2012; Stehling *et al*, 2012) . We found that disruption of *MMS19* led to an 18- and 14-fold increase in GCRs in strains containing 100% and 94% *Alu*-IRs, respectively. This is also consistent with our previous finding that *Δmms19* causes an increase in *Alu*-IR-induced homologous recombination (Lobachev *et al*, 2000).

The screen revealed that deletion of *SGS1*, the RecQ helicase homolog implicated in the dissolution of branched DNA structures and unwinding of CTG/CAG hairpins (Ashton & Hickson, 2010; Kerrest *et al*, 2009), caused a 6- and 7-fold elevation in GCRs in 100% and 94% repeats-containing strains. Sgs1 interacts with Rmi1 and Top3 to form the dissolvasome complex (Mankouri & Hickson, 2007). Consistently, we found that

deletion of *RMI1* and of *YLR235C* that partially overlaps with *TOP3* also led to hyper-fragility (Supplemental Table B.1). Our data suggest potential roles of Sgs1-Rmi1-Top3 in influencing palindrome stability through unwinding the hairpin or cruciform structures formed by the repeats.

The fifth group of hyper-fragile mutants consisted of TET-*TEN1*, TET-*STN1* and TET-*CDC13*. The Ten1-Stn1-Cdc13 complex is involved in telomere maintenance and protection (Grandin *et al*, 2001). Downregulation of Ten1 resulted in a 3-fold elevation of fragility (Table 3.1). The TET-*CDC13* strain demonstrated a similar increase in the level of arm loss. Notably, the closest telomere is about 40 kb away from the location of the inverted repeats. In another study, we found that downregulation of Ten1-Stn1-Cdc13 also predisposes the triplex-forming GAA/TTC repeats to breakage and expansions (Zhang *et al*, 2012). Taken together, these data suggest among other possibilities that this complex plays a role in helping replication machinery to move through difficult regions.

Previously, we demonstrated that the Mre11-Rad50-Xrs2 complex and the Sae2 protein are required to open hairpins to initiate DSB repair at inverted repeats (Lobachev *et al*, 2002). We also showed that in *Δmre11* mutants, GCR rates increased likely due to the inability of mutants to hold DSB ends together and open the hairpin termini, which therefore increase the probability of formation of dicentric chromosomes (Narayanan *et al*, 2006). Predictably, the screen identified *Δmre11* and *Δrad50* as hyper-fragile mutants with a 10- and ~44-fold increase in GCRs induced by homologous and homeologous *Alu*-IRs, correspondingly. This group therefore encompasses mutants that do not impact

secondary structure formation and breakage, but rather increase probability of arm loss and recovery of the broken chromosome.

Table 3.1. Mutants with increased Alu-IR-induced GCR rate

Genetic background	GCR rate ($X10^{-6}$)			
	100% homologous		94% homologous	
wild-type	41 (30-52)[a]		5 (4-6)	
Replication mutants				
TET-*RFA2*	250 (100-280)	6[b]	170 (80-180)	34
TET-*POL2*	240 (210-270)	6	130 (90-150)	26
TET-*POL1*	470 (380-500)	11	100 (80-110)	20
TET-*POL3*	460 (390-640)	11	82 (72-102)	16
TET-*POL30*	370 (290-390)	9	69 (60-73)	14
TET-*RFC2*	280 (170-380)	7	34 (21-44)	6
TET-*YHR122W*	140 (110-160)	3	38 (23-47)	8
Δ*mms19*	720 (370-820)	18	72 (61-85)	14
TET-*PRI2*	340 (260-470)	8	170 (130-200)	34
TET-*MCM2*	150 (140-240)	4	41 (16-62)	8
TET-*ORC4*	110 (80-230)	3	62 (31-76)	12
TET-*DNA2*	120 (80-200)	3	9 (9-12)	3
Δ*rad27*	600 (450-920)	15	90 (60-240)	18
Δ*pol32*	240 (190-300)	6	32 (28-36)	6
Checkpoint response genes				
Δ*tof1*	250 (170-300)	6	39 (31-48)	8
Δ*csm3*	370 (270-530)	9	27 (22-37)	5
Δ*rad17*	180 (160-250)	4	14 (13-16)	3
Δ*rad24*	140 (130-190)	3	12 (11-17)	2
Helicase				
Δ*sgs1*	260 (250-350)	6	35 (26-44)	7
Telomere protection genes				
TET-*TEN1*	140 (120-230)	3	13 (9-16)	3
TET-*CDC13*	120 (70-140)	3	14 (11-18)	3
Double strand breaks repair genes				
Δ*mre11*	420 (370-440)	10	210 (170-230)	42
Δ*rad50*	400 (370-430)	10	220 (200-250)	44

[a] Numbers in the brackets are 95% confidence intervals

[b] Fold increase in GCR rates in mutants compared to wild-type strains

3.3.3 DSB formation is increased in replication-deficient and Δsgs1 mutants

In the wild-type strain, DSBs induced by *Alu*-IRs have covalently-closed hairpin termini. To determine if the nature of breaks in the identified hyper-GCR mutants was similar to the wild-type strain, we characterized DSB intermediates in a subset of mutants. In addition, estimation of the level of breaks provides a way to distinguish between mutants that facilitate formation or enhance stability of the secondary structures and mutants that increase the loss of the acentric DSB fragment (e.g. *mrx* mutants) or improve the recovery of the broken chromosome.

We compared the levels of chromosomal breaks in the wild-type strain containing 100% *Alu*-IRs with a subset of mutants from each group described in the previous section (Figure 3.2 and Supplemental Figure B.2). DSB detection was carried out in *Δsae2* strains to prevent the opening of the hairpins and the resection of the broken fragments (Lobachev et al, 2002). The lethality of Δ*sgs1*Δ*sae2* can be resucued by the deletion of *HDF1* (Mimitou & Symington, 2010). Therefore, the effect of Δ*sgs1* on DSB formation was assessed in the Δ*sgs1*Δ*sae2*Δ*hdf1* triple mutant. DSBs were analyzed with a telomere-distal probe upon AflII digestion or a telomere-proximal probe using BglII digestion of chromosomal DNA embedded in agarose plugs. Upon AflII or BglII digestion, DSBs occurring inside the repeats were expected to be 1.3 kb or 3.3 kb, respectively. We also anticipated the appearance of inverted dimers that are double the size of the DSB intermediates (2.6 kb or 6.6 kb, correspondingly). These molecules

resulting from replication of hairpin-capped breaks were previously detected in the wild-type strains (Lobachev *et al*, 2002).

No DSBs were observed in the presence of Sae2 in both wild-type and mutant strains, likely due to hairpin opening and robust resection of the breaks. However, DSBs were readily detected in Δ*sae2* background. In TET-*POL3,* TET-*POL2,* Δ*csm3*, Δ*sgs1*Δ*hdf1* (Figure 3.2), Δ*mms19*, and TET-*TEN1* (Supplemental Figure B**.**2) mutants, there was a 2- to 15- fold increase in breaks in comparison with wild-type strains when the telomere-proximal or the telomere-distal fragments were probed, indicating that these mutations increase fragility at *Alu*-IRs by either facilitating secondary structure formation or stabilizing the structures. Conversely, in Δ*rad17*, the amount of breaks was comparable to the wild-type strain, suggesting that DNA damage checkpoint-deficient mutants provide conditions for better recovery of the broken chromosomes, rather than affecting the formation and/or stability of the secondary structures. It is important to note that besides DSBs we could also detect dimers and no other intermediates were observed. The dependence of DSB detection on *Δsae2* and the existence of dimers suggest that breaks in hyper-fragile mutants might contain hairpin termini similar to those in wild-type strains.

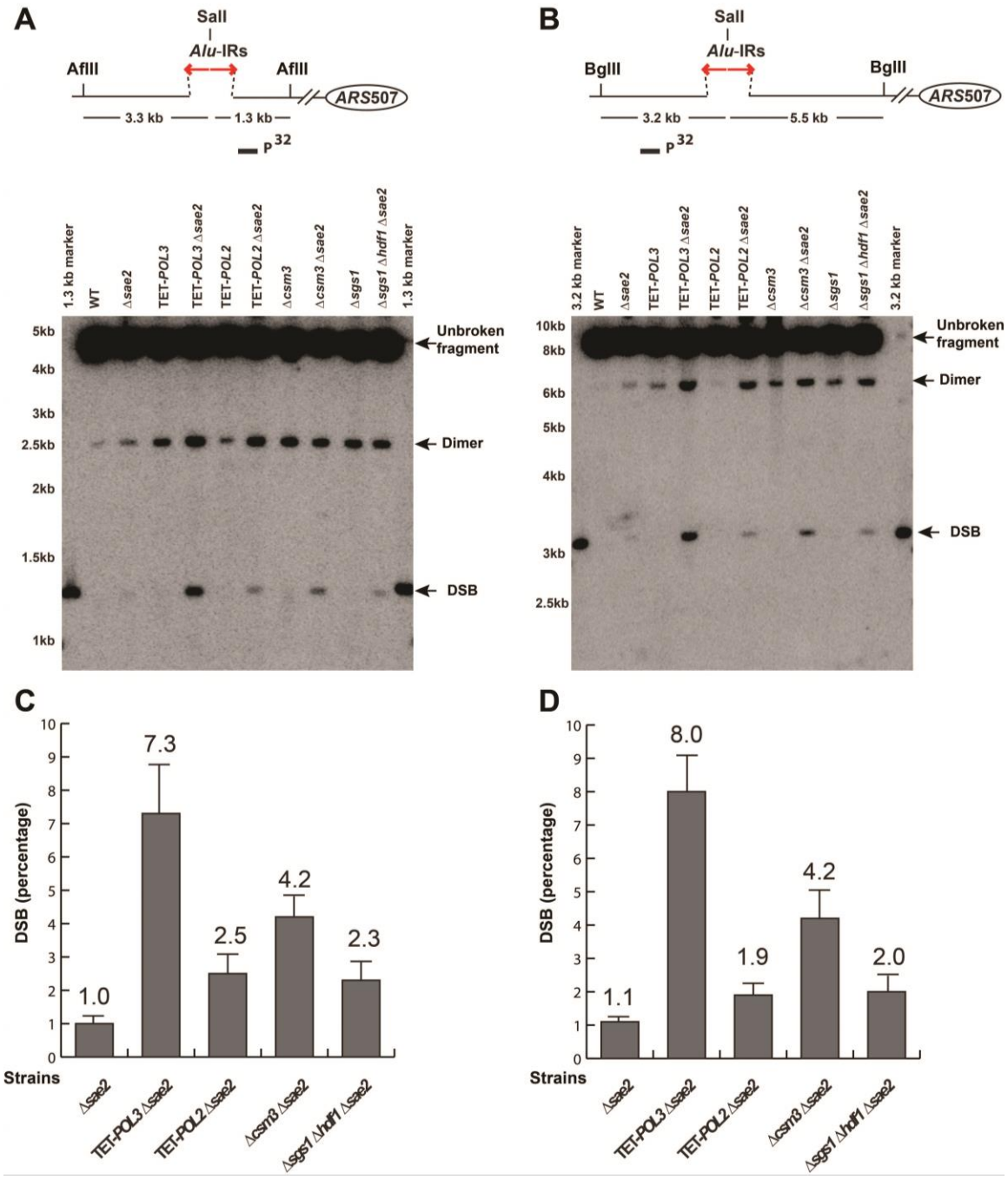

Figure 3.2. Physical detection of breakage intermediates in the wild-type and mutant strains carrying 100% *Alu*-IRs.

Yeast genomic DNA embedded in agarose plugs was digested with either AflII (**A**) or BglII (**B**). The relative positions of the repeats, the restriction sites and the replication origin *ARS*507 are illustrated. Digested DNA was separated by gel electrophoresis. Southern hybridization using *LYS2*-specific probes (solid rectangles) was carried out to detect the chromosomal fragments centromere-proximal (**A**) or distal (**B**) to the breaks

induced by *Alu*-IRs. For the centromere-proximal intermediates (**A**), the size of unbroken fragment, dimer and DSB fragment are 4.6 kb, 2.6 kb and 1.3 kb, respectively. For the centromere-distal intermediates (**B**), the size of unbroken fragment, dimer and DSB fragment are 8.7 kb, 6.4 kb and 3.2 kb, respectively. Bands corresponding to the unbroken fragment, dimer and DSB fragment are indicated by arrows. The 1.3 kb marker and 3.2 kb marker were generated by digesting genomic DNA from the wild-type *Alu*-IRs strain with SalI + AflII (**A**) or SalI + BglII (**B**), where SalI cuts inside the 12 bp spacer of the *Alu*-IRs. The strains used for analysis are: wild-type, Δ*sae2*, TET-*POL3*, TET-*POL3*Δ*sae2*, TET-*POL2*, TET-*POL2*Δ*sae2*, Δ*csm3*, Δ*csm3*Δ*sae2*, Δ*sgs1*, Δ*sgs1*Δ*hdf1*Δ*sae2*. (**C**) and (**D**) Densitometry analysis of the broken fragments normalized to the intact chromosome V in Δ*sae2* strains in (**A**) and (**B**), respectively. Values are shown as mean (shown on the top of the bars) with standard deviation obtained from at least three independent experiments.

3.3.4 DSBs in replication-deficient strains have hairpin-capped termini

To test the premise of hairpin-capped breaks in the mutants experimentally, the DSB fragments in TET-*POL3*Δ*sae2* were analyzed via neutral/alkaline two-dimensional (2D) gel electrophoresis (Figure 3.3). We found that the 1.3 kb telomere-distal DSB fragment migrated as a 2.6 kb single-stranded DNA (ssDNA) fragment in the alkaline gel. Similarly, the 3.3 kb telomere-proximal DSB fragment migrated as a 6.6 kb ssDNA fragment under denaturing conditions. No additional bands (e.g. those corresponding to nicked hairpins) were seen, indicating that *Alu*-IRs generate covalently-closed hairpin-capped breaks in both TET-*POL3* and wild-type strains.

The symmetry of the breaks and the presence of covalently-closed hairpins at the DSB termini suggest that the final steps in breakage in mutants and wild-type are the same and include cruciform formation and resolution.

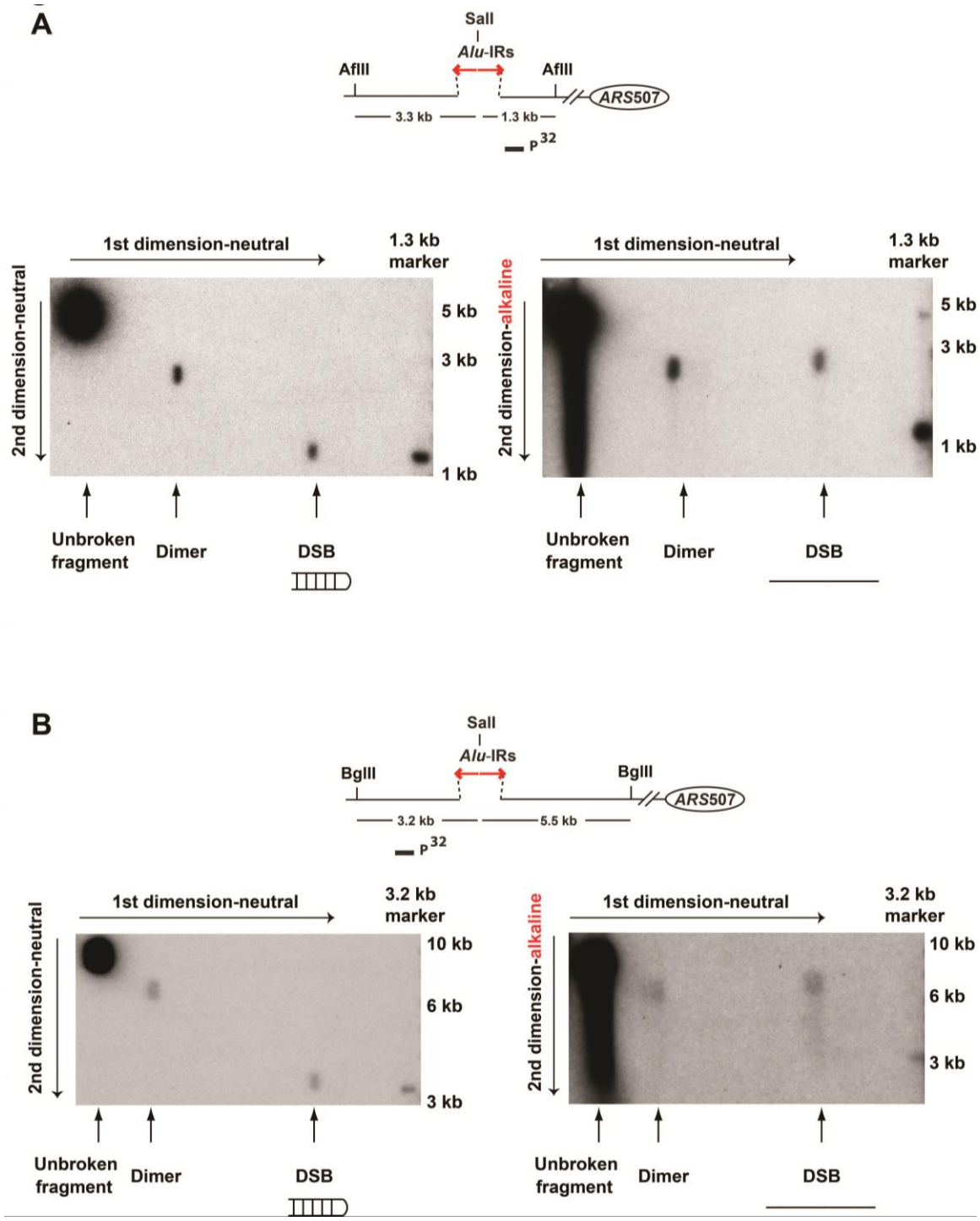

Figure 3.3. 2D neutral/alkaline gel analysis of *Alu*-IR-induced DSBs in theTET-*POL3*Δ*sae2* strain.

Yeast DNA embedded in agarose plugs was digested with AflII (**A**) or BglII (**B**). Digested DNA was separated in neutral conditions in the first dimension. In the second dimension, DNA was run in either neutral (left gel) or alkaline (right gel) conditions. Southern hybridization was done as described in Figure 3.2.

3.3.5 Alu-IR-mediated fragility and fork arrest are Rad51-dependent in replication-deficient strains

The screen revealed that mutants deficient in the DNA replication pathway comprise the major group that augments fragility at *Alu*-IRs. Analysis of DSB intermediates indicated that cruciform resolution is the likely scenario for fragility in these mutants (Figure 3.2, 3.3). Generation of ssDNA due to replication defects in the leading or lagging strands might provide optimal conditions for the formation of hairpins, not cruciforms. We hypothesized that a deficiency in the DNA replication can lead to formation of the cruciform structure through template switching when the fork stalls at a hairpin (Figure 3.6). In a screen for factors that channel replication stress into fragility, we identified Rad51, a key protein in homologous recombination. In the Δ*rad51* background, the GCR rates of both TET-*POL3* and TET-*POL2* mutants decreased almost to the wild-type level (Table 3.2). Consistent with the reduction in GCRs, the amount of DSBs and inverted dimers in TET-*POL3*Δ*sae2* and TET-*POL2*Δ*sae2* significantly decreased upon deletion of *RAD51*. Notably, lack of Rad51 does not affect GCR rates or DSB formation in the wild-type strains, indicating that the involvement of homologous recombination in the induction of fragility is specific to conditions when replication is compromised (Table 3.2 and Figure 3.4).

Table 3.2. *Alu*-IR-mediated fragility in TET-*POL3* and TET-*POL2* mutants is Rad51-dependent

Genetic background	GCR rate ($X10^{-6}$)	Fold increase over wild-type
WT	41 (30-52)[a]	1
Δ*rad51*	37 (27-50)	1
TET-*POL3*	460 (390-640)	11
TET-*POL3*Δ*rad51*	88 (58-108)	2
TET-*POL2*	240 (210-270)	5
TET-*POL2*Δ*rad51*	63 (46-79)	1

[a] Numbers in brackets are the 95% confidence interval.

To gain better insight into the mechanism underlying *Alu*-IR-induced fragility, we monitored replication progression through 100% homologous repeats in the wild-type strain and the replication-deficient mutant TET-*POL3* using 2D gel electrophoresis and Southern hybridization. While replication progression was not hampered at the quasi-palindrome in the wild-type strain, the TET-*POL3* mutant demonstrated a robust fork arrest at the repeats. The fact that the replication block in TET-*POL3* is completely removed upon deletion of *RAD51* (Figure 3.5) argues for Rad51-mediated template switching as the signal for replication pausing. These data, along with the observation that Δ*rad51* suppresses DSB formation in replication deficient strains, support the scenario where an attempt to bypass hairpin structures during compromised replication via Rad51-dependent template switching promotes the formation of cruciform structures behind the replication fork. These structures are further attacked by nucleases, resulting in DSBs (Figure 3.6). Although DSB formation in other hyper-fragile mutants in Δ*rad51* background was not analyzed, the fact that the GCR levels in these strains decreased as

compared to their *RAD51* counterparts strongly suggests that the same mechanism of break formation operates in these mutants (Supplemental Table B.2).

Overall, these data reveal an important role of homologous recombination in promoting DSB formation at inverted repeats, specifically in replication-deficient mutants.

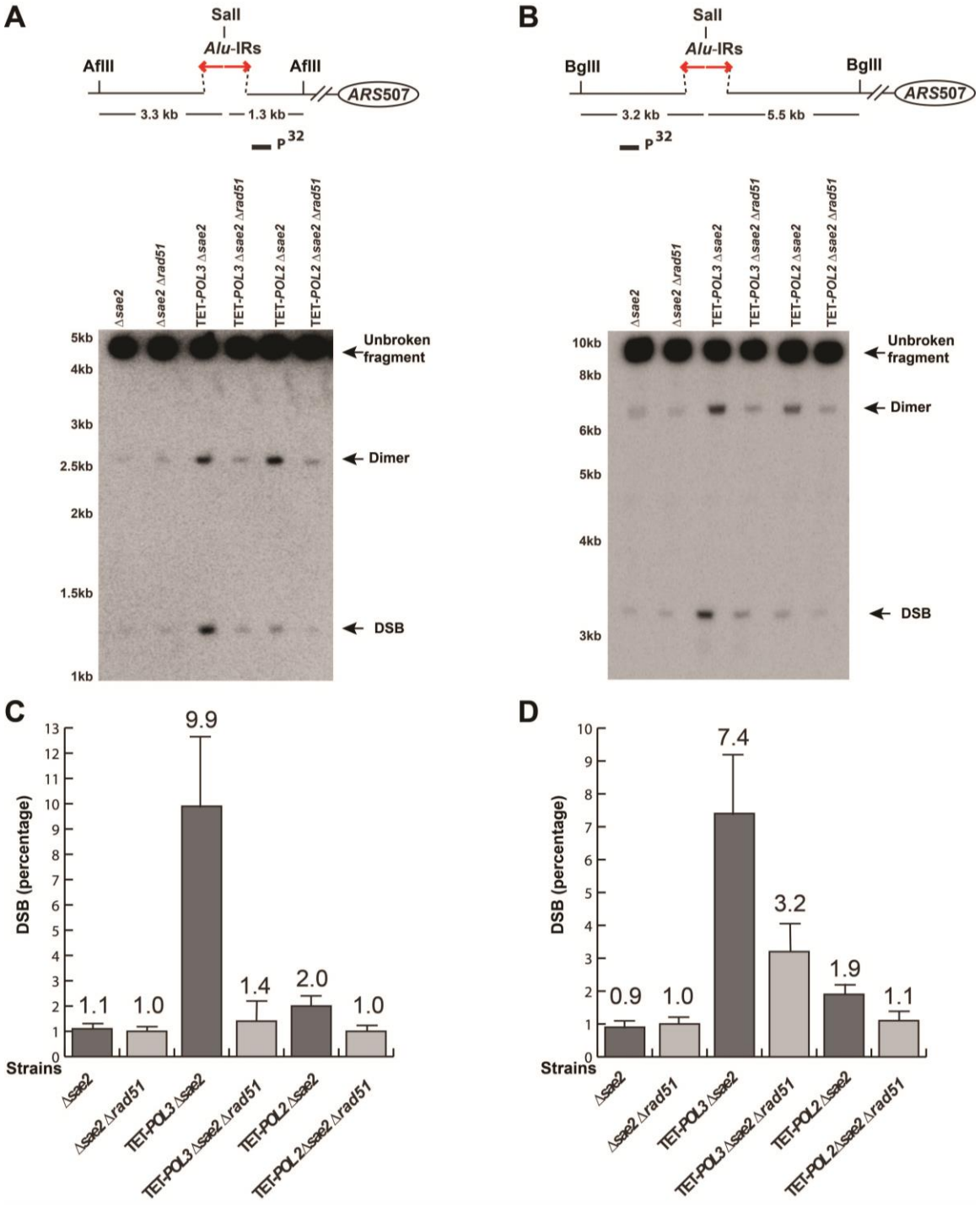

Figure 3.4. Detection of DSB accumulation in wild-type and mutant strains upon deletion of *RAD51*.

DSB detection was carried out as described in Figure 3.2. The strains used in this analysis are: Δ*sae2*, Δ*sae2*Δ*rad51*, TET-*POL3*Δ*sae2,* TET-*POL3*Δ*sae2*Δ*rad51*, TET-*POL2*Δ*sae2,* TET-*POL2*Δ*sae2*Δ*rad51*. (**C**) and (**D**) Densitometry analysis of the broken

fragments normalized to the intact chromosome V in Δ*sae2* strains in (**A**) and (**B**), respectively. Values are shown as mean (shown on the top of the bars) with standard deviation obtained from at least three independent experiments.

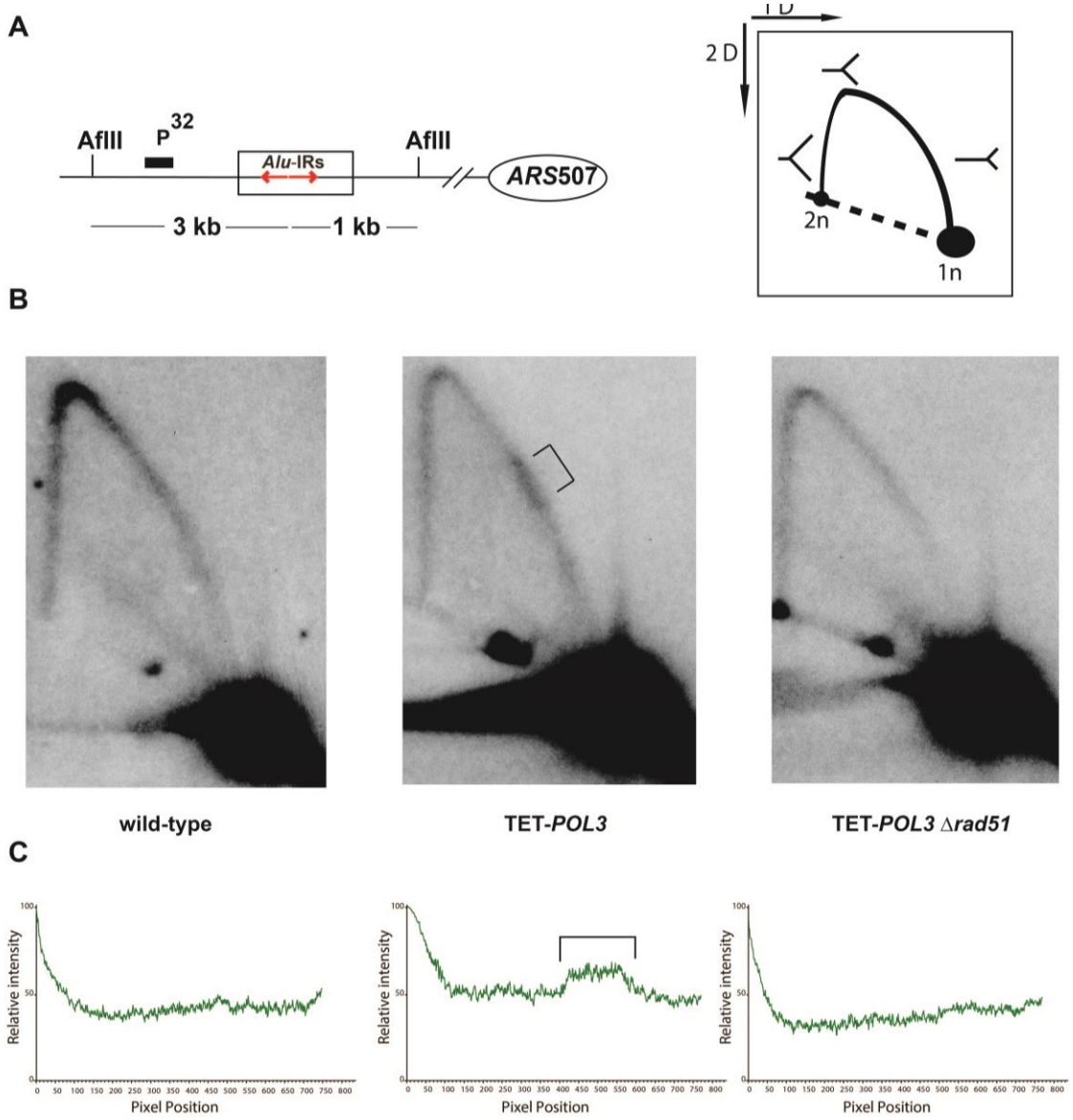

Figure 3.5. Analysis of replication fork progression through *Alu*-IRs in the wild-type and mutant strains.

DNA samples from the wild-type, TET-*POL3* or TET-*POL3*Δ*rad51* strains were digested with AflII and processed for 2D gel analysis. (A) Illustration of restriction digestion and

2D gel analysis. The solid rectangle indicates the position of the probe used for Southern hybridization. (B) 2D gel analysis of replication intermediates in the wild-type and mutant strains. The zone of replication arrest in the TET-*POL3* strain is indicated by the bracket. (C) Densitometry analysis of the Y arc's long arm in the corresponding strains in (B). The relative radioactive counts along the long arm of the Y-arc that starts from the monomer are shown. The bracket depicts the zone of replication arrest.

3.4 Discussion

Palindromic sequences are strong inducers of DSBs and rearrangements in both prokaryotes and eukaryotes. The two distinct events that trigger fragility are considered to be the formation of either hairpin or cruciform structures at the repeats. In this study we found that when replication is compromised, replication delay imposed by inverted repeats is channeled into cruciform resolution via the action of homologous recombination pathway. These data led us to propose that the transition from hairpin to cruciform formation through Rad51-mediated template switching is the mechanism for fragility operating in cells under replication stress.

3.4.1 Genome-wide screen identifies intact replication as a major guardian of quasi-palindrome stability

Inverted repeat-induced GCRs can be augmented in mutants that either influence secondary structure metabolism or alter repair of the broken chromosome. Previous studies from our lab have demonstrated that *Alu*-IRs-induced DSBs have hairpin termini that are opened by the Mre11 complex and Sae2 to initiate resection (Lobachev *et al*, 2002). Unprocessed hairpin-capped molecules lead to the formation of acentric and dicentric inverted chromosomes. Detailed analysis of GCR events showed that dicentric chromosomes are stabilized as a result of breakage in anaphase, followed by resection

and repair preferentially via break-induced replication with non-homologous chromosomes. It is important to note that DSB resection that precedes the healing of the broken chromosome activates checkpoint signaling and is manifested as cells arrested in G2/M (Narayanan *et al*, 2006). Previously, we found that GCR rates are elevated in *mrx* mutants. This increase is not due to frequent DSBs at *Alu*-IRs, but rather a result of more efficient formation of dicentric chromosomes and loss of the broken acentric fragments. Consistently, *Δmre11, Δrad50,* and Δ*xrs2* were identified in this genome-wide screen as hyper-fragile mutants (Table 3.1 and Supplemental Table B.1). Another group of mutants that do not increase breakage but amplify GCR rates are those defective in DNA damage checkpoint signaling (Δ*rad17*, Δ*mec3*, Δ*ddc1*, Δ*rad24*) (Table 3.1, Supplemental Table B.1 and Supplemental Figure B.1). It is conceivable that in the absence of checkpoint activation after dicentric breakage, the rate of resection is decreased (Aylon & Kupiec, 2003) and the broken chromosomes are replicated and segregated together to the daughter cells for several generations (Lee *et al*, 1998; Sandell & Zakian, 1993), which improves their chances for repair.

The mutants identified in the screen that increase DSB formation and GCRs at *Alu*-IRs are deficient in DNA replication, replication-pausing checkpoint surveillance, Fe-S cluster biogenesis, telomere maintenance and protection, or the function of the Sgs1-Rmi1-Top3 dissolvasome. As discussed below, the impact of deficiencies in these different processes on fragility can be explained by an increase in the probability of formation or stability of secondary structures during replication.

The screen revealed that depletion of the major components of the replication fork responsible for synthesis of both leading and lagging strands increases *Alu*-IR-induced

fragility. Our results are consistent with previous findings that mutations in the DNA polymerases α and δ promote excision of IRs and IRs-induced recombination and rearrangements (Farah *et al*, 2005; Gordenin *et al*, 1993; Gordenin *et al*, 1992; Kogo *et al*, 2007; Ruskin & Fink, 1993). It is possible that deficiencies in the synthesis of either the leading or lagging strand can lead to the generation of extensive single-stranded regions, thereby creating ideal conditions for the formation of hairpin structure, the initial event in *Alu*-IRs fragility (Figure 3.6). Interestingly, downregulation of the helicase Mcm2-7 and ORC also led to hyper-fragility at the repeats. Although the MCM helicase, is a part of the replication machinery, it travels ahead of the fork, therefore generation of ssDNA due to depletion of this helicase is unlikely. The effect of deficiencies in MCMs and ORC on fragility might be the consequence of the inability of the closest origin (*ARS*507) to fire since amounts of both protein complexes are important for regulating the timing of origin activation (Wu & Nurse, 2009). Replication forks traveling longer distances from the remote origins might be less processive and more prone to collapse upon encountering replication barriers. Downregulation of MCMs and ORC also increases instability at another fragile motif in yeast, the triplex-forming GAA/TTC repeats (Zhang *et al*, 2012), indicating that this phenomenon might be universal in situations when the replication fork passes through difficult regions. Consistent with this assertion, in human cell lines that have different replication landscapes, fragility at FRA3B and FRA16D sites depends on the distance the replication fork travels (Letessier *et al*, 2011). Alternatively, increased fragility in mutants for MCMs and ORC might be due to the assembly of a hampered replisome that lacks components required for leading or lagging strand synthesis.

Deletion of *MMS19* and downregulation of *YHR122W*, genes encoding proteins involved in Fe-S cluster biogenesis (Gari *et al*, 2012; Stehling *et al*, 2012), were also found to induce hyper-fragility at *Alu*-IRs. Recently, it has been shown that Mms19 and Yhr122W along with Cia1, are required for the transfer of Fe-S clusters to various proteins including polymerase δ, DNA primase and Dna2 (Gari et al, 2012; Stehling et al, 2012), deficiencies in which were identified to augment fragility in the screen . The presence of the Fe-S clusters in the polymerases α and ε (Netz *et al*, 2012) and the fact that these proteins interact with Mms19 (Stehling *et al*, 2012) also makes them likely substrates for the CIA targeting complex. The effect of mutation in this pathway on *Alu*-IRs-mediated fragility can therefore be attributed to the impaired maturation of the DNA replication machinery.

The deficiencies described above are expected to create optimal conditions for the formation of a hairpin that impedes replication progression. The hairpin might be formed at lower frequencies in replication-proficient cells as well. In both replication-proficient and -deficient strains, the secondary structure or the arrested fork might trigger the activation of checkpoint response required to recruit proteins to remove the hairpin and promote replication restart (Figure 3.6). The fact that disruption of the Mrc1-Tof1-Csm3 complex leads to hyper-fragility implicates this replication-pausing surveillance complex as a possible sensor of secondary-structure-imposed replication arrest. However, this complex is also required to coordinate the Mcm2-7 helicase and DNA polymerase activities (Grandin *et al*, 2001; Katou *et al*, 2003; Tourriere *et al*, 2005; Zegerman & Diffley, 2003), therefore, we can not completely rule out that deficiencies in this complex affect the integrity of the replisome as well.

It seemed reasonable to suggest the existence of helicases recruited to remove hairpins at the arrested fork. Indeed, the screen identified the Sgs1-Top3-Rmi1dissolvasome. Although Δ*sgs1* does not affect the stability of short CAG/CTG repeats (less than 25 repeats), it increases the contraction and fragility rate of long CAG/CTG repeats (70 repeats), indicating that longer hairpins might be better substrates for Sgs1 activity (Bhattacharyya & Lahue, 2004; Kerrest *et al*, 2009). In addition, the Sgs1-Top3-Rmi1 complex is involved in the dissolution of double Holliday junctions (Cejka *et al*, 2010). Hence, it is probable that this complex also irons out long hairpins formed by *Alu*-IRs during replication.

An interesting group of mutants that destabilize *Alu*-IRs include TET-*TEN1*, TET-*CDC13*, and TET-*STN1*. The Cdc13-Stn1-Ten1 (CST) complex is involved in protection of chromosome ends, telomerase recruitment and telomere replication. Hyper-fragility at inverted repeats due to deficiencies in this complex can be explained by the sequestration of the Tof1-Mrc1-Csm3 complex from the replisome to the single-stranded regions at uncapped telomeres (Grandin *et al*, 2005; Tsolou & Lydall, 2007). Alternatively, this complex which is structurally similar to RPA (Sun *et al*, 2009) may facilitate replication progression through the hairpin. Dewar and Lydall, (2012) proposed that in mammalian cells the CST complex which is distributed throughout the genome (Miyake *et al*, 2009), aside from its role in telomere metabolism, facilitates replication through difficult regions. Taking into account that downregulation of the CST complex also increases GAA/TTC-mediated fragility and expansions (Zhang *et al*, 2012) and the physical interaction of this complex with Polα (Grossi et al, 2004; Qi & Zakian, 2000), it

is reasonable to suggest that the role of CST in DNA replication might be evolutionarily conserved.

3.4.2 Analysis of DSB intermediates in replication-deficient mutants points towards cruciform-resolution mechanism of fragility

In wild-type strains carrying inverted repeats, the deduced mechanism of breakage is cruciform-resolution by a putative nuclease that cuts symmetrically at the base of the two hairpins. This generates two hairpin-capped molecules that are present in equimolar ratios (Lobachev *et al*, 2002). Since replication is a polar process, in replication-deficient strains, a nuclease attack on the accumulated hairpins or stalled replication fork would be expected to produce DSB intermediates different from those induced in the wild-type strains. Anticipated intermediates would include nicked hairpins, branched structures, or asymmetrical hairpin-capped breaks. Somewhat unexpectedly, we found that in the TET-*POL3* strain in the absence of Sae2, the DSB intermediates were structurally identical to the replication-proficient strains: only covalently-closed hairpin-capped breaks and inverted dimers resulting from replication of the DSBs were detected. Accumulation of hairpin-capped intermediates on both sides of the break indicates that cruciform-resolution is the predominant pathway for fragility under replication stress. Since deletion of *SAE2* leads to stabilization of hairpin-capped breaks in all mutants analyzed, we propose that this mechanism operates not only in the TET-*POL3* strain, but also in other hyper-GCR mutants identified in the screen.

3.4.3 Rad51 is a key mediator of fragility in replication-deficient mutants

Based on our finding that deletion of *RAD51* strongly decreases GCRs and breaks in replication-deficient strains (Table 3.2 and Figure 3.4), we proposed that cruciform

formation and resolution can result from the action of the homologous recombination machinery on intermediates present at the stalled replication fork. Consistent with this conjecture, replication arrest observed in TET-*POL3* was also dependent on Rad51. We can not completely rule out the possibility that Rad51 facilitates the hairpin formation. However, taking into account that Rad51 forms nucleoprotein filaments that are essential for the invasion step of homologous recombination (Krogh & Symington, 2004), we favor the explanation that Rad51 promotes template switching when the replication fork encounters the hairpin structure. Synthesis of the haipin-forming sequence on the unperturbed strand and reannealing of this newly synthesized DNA might allow formation of a cruciform structure which is resolved by a putative nuclease to give rise to hairpin-capped DSBs (Figure 3.6). In this case, the replication stalling observed in TET-*POL3* would reflect the accumulation of arrested forks in response to template switching rather than inhibition of DNA synthesis by the hairpin structure. Rad51 was found to be present at unperturbed and stalled replication forks (Gonzalez-Prieto *et al*, 2013; Hashimoto *et al*, 2010; Schlacher *et al*, 2011; Sirbu *et al*, 2011), and the involvement of recombination proteins in the fork restart and bypass of DNA lesions via template switching has been demonstrated in several studies (Bugreev *et al*, 2011; Gangavarapu *et al*, 2007; Mizuno *et al*, 2009; Mizuno *et al*, 2013; Petermann *et al*, 2010; Schlacher *et al*, 2012; Zhang & Lawrence, 2005). Here, we show that the attempt of homologous recombination proteins to bypass the secondary-structure barrier may be detrimental and culminate in breaks and GCRs.

It is important to note that the Rad51 effect is specific in situations where replication is compromised. In replication-proficient strains, breaks and GCRs are not

affected by Rad51 status, indicating that another mechanism for cruciform-formation exists. It is possible that in wild-type strains a homologous recombination-independent template switching mechanism leading to fragility operates, or that the cruciform formation is unrelated to replication. The latter hypothesis is supported by our recent finding that hairpin-capped breaks in the wild-type strain preferentially occur in G2 phase of the cell cycle (Sheng *et al*, in preparation).

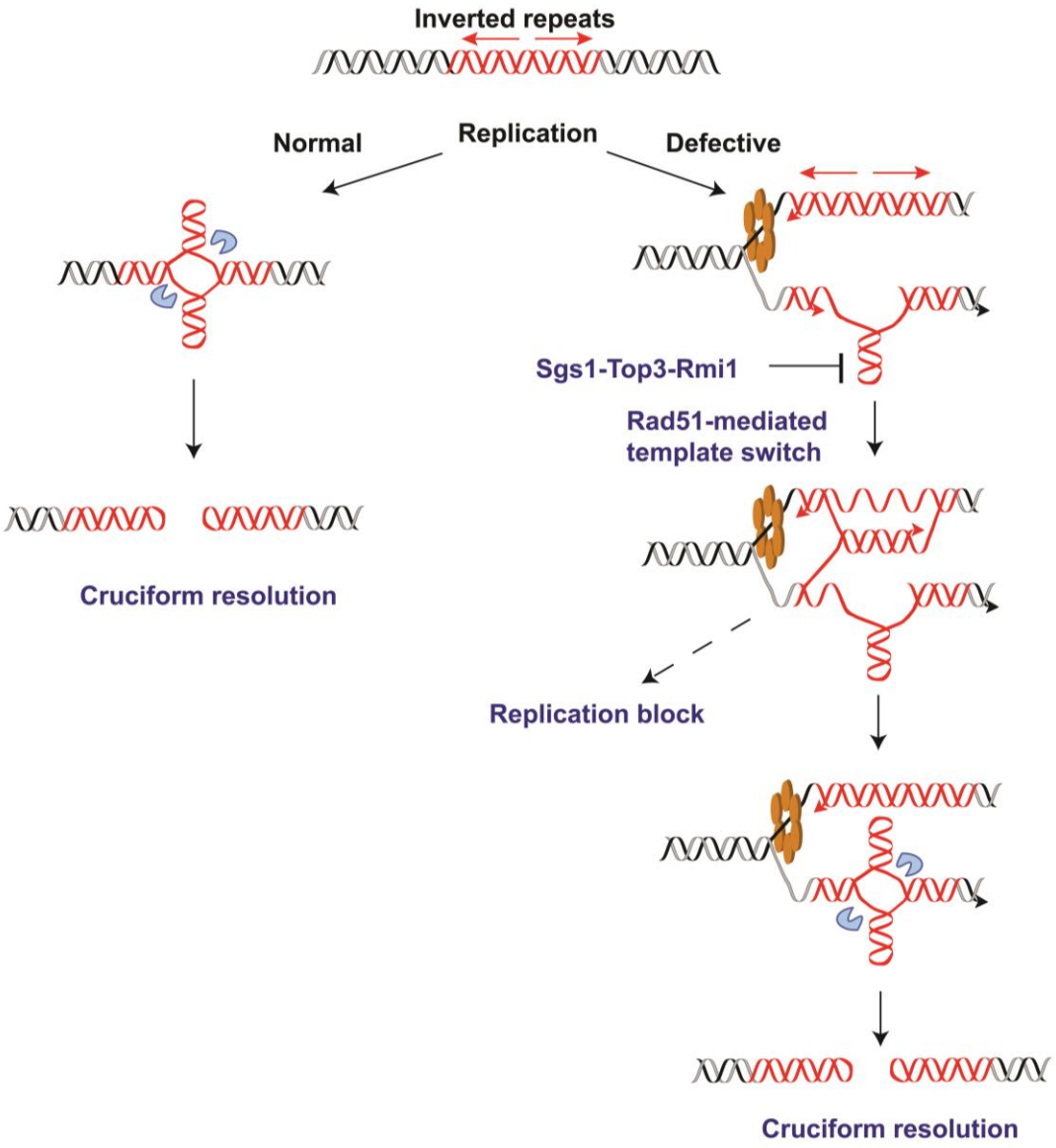

Figure 3.6. Model for *Alu*-IRs-mediated fragility under conditions of replication proficiency and deficiency.

The red helixes, blue pacman and orange hexamer depict the inverted repeats, the putative nuclease and the DNA replication helicase, respectively. In the case of normal replication, cruciform structure might form outside of S-phase as a result of chromatin packing or remodeling. On the other hand, long single-stranded DNA exposed due to compromised replication would facilitate the formation of a hairpin, which could further be converted into cruciform structure via template switch by Rad51. The intermediates of template switching present a strong obstacle for the replication machinery that is manifested as replication block in the replication-deficient strains. The Sgs1-Top3-Rmi1 dissolvosome might participate in unwinding the hairpin. Once formed, the cruciform structure might be attacked by the putative nuclease, leading to DSBs at the IRs.

Based on this study, we propose that in the human population, the carriers of hypomorphic alleles for the BLM-hTOPOIIIα-hRMI1-hRMI2 dissolvasome and proteins involved in DNA replication, replication-pausing checkpoint surveillance, Fe-S cluster biogenesis, telomere maintenance and protection might be susceptible to inverted repeat-induced breaks and carcinogenic GCRs. Importantly, the status of these proteins determines the stability of imperfect repeats with a spacer and divergent arms that are present in the human genome. At the same time, it is likely that homologous recombination can trigger chromosomal breakage at secondary structure-forming fragile sites and AT-rich palindromic sequences under conditions of replication stress. This detrimental role of homologous recombination in promoting chromosomal instability might contribute towards the development of diseases associated with fragile motifs. Homologous recombination-mediated chromosomal breakage and rearrangements might operate at secondary structure-forming fragile sites and AT-rich palindromic sequences under replication stress. This detrimental role of homologous recombination in promoting genome instability might contribute towards the development of diseases.

3.5 Materials and Methods

3.5.1 Yeast strains

yTHC, DAmP and YKO collections were purchased from Open Biosystems. All other strains in this study are derivatives of BY4742 (Open Biosystems). The genotype of the query strains for the screen is: MATα, *ura3*Δ, *leu2*Δ, *his3*Δ, *lys2*Δ, *rpl28-Q38K*, *mfa1*Δ*::MFA1pr-HIS3, V34205::lys2::Alu*-IRs, *V29617::hphMX*. The 100% or 94% homologous inverted *Alus* were inserted into *LYS2* gene via the pop-in and pop-out

method as previously described (Lobachev *et al*, 2000). The detailed construction of the query strain can be found in Zhang et al. (2012).

The effect of mutant alleles identified from the screen was verified in derivatives of YKL36 that carries the GCR assay and has the following genotype: MATa, *bar1*Δ*, trp1*Δ*, his3*Δ*, ura3*Δ*, leu2*Δ*, ade2*Δ*, lys2*Δ*, V34205*::*ADE2, lys2*::*Alu*-IRs. To create the mutant strains, in the case of non-essential genes, the target gene was disrupted by the *kanMX4* cassette (Wach *et al*, 1994); in the case of essential genes, the repressible $tetO_7$ promoter construct was PCR-amplified (Belli *et al*, 1998) from pCM225 (Euroscarf) and was used to replace the natural promoter of the gene to create the TET-alleles.

In strains used for DSB analysis, *SAE2* was disrupted by *TRP1*. For construction of the Δ*sgs1*Δ*hdf1*Δ*sae2* triple mutant, first, *SGS1* was disrupted by the *kanMX4 cassette*, and *HDF1* was knocked out by the *hphMX* cassette (Goldstein & McCusker, 1999). To study the effect of *RAD51* on *Alu*-IRs-mediated fragility, *RAD51* was replaced by a hisG-*URA3*-hisG cassette (Alani *et al*, 1987).

All primer sequence information for strain construction is available upon request.

3.5.2 Genome-wide screen scheme

The screen was carried out as described in Zhang et al. (2012).

3.5.3 Measurement of GCR rates

Yeast cells were grown on YPD plates for 3 days. For each strain, a minimum of 14 independent colonies were taken to perform fluctuation test to estimate GCR rates. Appropriate dilutions of cells were plated on YPD and canavanine-containing plates to

determine the GCR frequency. The GCR rates were calculated using the formula $\mu = f/\ln(N\mu)$ as described in Drake, 1991. 95% confidence intervals were calculated as described in Dixon, 1969. The canavanine-containing plates used for tests were made from arginine-drop out medium with low amount of adenine (5 mg/L) and supplemented with L-canavanine (60 mg/L).

3.5.4 DSB detection

Yeast cells from overnight cultures were embedded into 0.8% low-melting agarose plugs at a concentration of 24 X 10^8 cells/ml. The plugs were treated with 1.5 mg/ml lyticase for 3 hr., followed by overnight 1mg /ml proteinase K treatment. For restriction digestion of the DNA, the plugs were washed twice with 1 X TE buffer (10 mM Tris-Cl [pH 8.0], 0.1 mM EDTA) for 30min, treated with 1 mM PMSF for 1hr, washed with distilled water for 1hr and equilibrated with restriction buffer for 20min. Each plug (~40 µl) was digested with 50 units of AflII or BglII for 16 hr. Digested plugs were loaded in a 1% (AflII digestion) or 0.7% (BglII digestion) agarose gel, respectively, and run in 1 X TBE for 18 hr. The gels were treated with 0.25 N HCl for 20min, alkaline buffer (1.5 M NaCl, 0.5 M NaOH) for 30min and neutralization buffer (1.5 M NaCl, 1 M Tris [pH 7.5]) for 30min. The gels were then transferred in 10X SSC to charged nylon membrane for 2 hr. through a Posiblotter (Stratagen). Southern hybridization was carried out using P^{32}-labeled *LYS2*-specific probes at 67 ^oC overnight. DNA membranes were washed twice for 15 minutes each in buffer containing 0.1% SDS and 1% SSC and the signals were detected by the typhoon phosphoimager (GE Healthcare Life Sciences). The hybridization signals were quantified using ImageJ software (NIH).

3.5.5 2D neutral/neutral or neutral/alkaline gels for analyzing the structure of the broken ends

Yeast plugs were prepared and digested as described above. Neutral/neutral and neutral/alkaline gel analysis was performed as previously described with small modifications (Lobachev et al, 2002; Oh et al, 2009). In the first dimensional gel electrophoresis, the plugs were loaded in a 1% (AflII digestion) or 0.7% agarose (BglII digestion) gel, respectively, and run for 18 hr. in 1 X TBE. The gel slices containing the bands of interest were then cut out for the second dimensional gel electrophoresis. For neutral/neutral gel analysis, the gel slices were loaded in 1% (AflII digestion) or 0.7% (BglII digestion) agarose gel made in 1 X TBE, run in 1 X TBE for 18 hr. at 1.7 V/cm and then processed for Southern hybridization. For neutral/alkaline gel, the gel slices were treated with 10 mM EDTA for 30 minutes, 5 mM EDTA for 30 minutes and embedded in agarose gels made in buffer containing 50 mM NaCl, 1 mM EDTA. Next, the gels were soaked in 5X alkaline buffer for 30 minutes, 1 X alkaline buffer (50 mM NaOH, 1 mM EDTA) for 30 minutes and cooled down in 1 X alkaline buffer at 4 ^{o}C for 15 minutes. The gels were then run in 1 X alkaline buffer at 0.7 V/cm for 40 hr. at 4 ^{o}C and processed for Southern hybridization.

3.5.6 2D neutral/neutral gel analysis for replication fork progression

2D gel analysis was carried out as previously described in Brewer and Fangman, 1987. Overnight yeast cultures were synchronized in G1 with alpha factor (50 μg/10^7 cells) at OD600 = 0.8. 2 μg/ml doxycycline was added to the cultures to downregulate Pol δ in the case of TET-*POL3* and TET-*POL3*Δ*rad51* strains. Cells were then released into fresh YPD. 50min after release, wild-type, TET-*POL3*, TET-*POL3*Δ*rad51* strains were

harvested and their genomic DNA samples were prepared as described in Friedman and Brewer, 1995. For the first dimensional gel electrophoresis, AflII digested DNA samples were loaded in a 0.4% agarose gel and run in 1 X TBE at 1.7 V/cm for 22 hr. For the second dimensional gel electrophoresis, gel slices containing bands of interest were cut out and loaded into a 1.2% agarose gel supplemented with 0.3 mg/ml ethidium bromide. The gels were run in 1 X TBE at 6 V/cm for 11 hr. Gels were then processed for Southern hybridization. Images were quantified using ImageQuant TL software (GE Healthcare Life Sciences).

3.6 Acknowledgements

We thank Dr. Natasha Degtyareva, Dr. Matthew Torres, and Anastasiya Lobacheva for critical reading of the manuscript and helpful discussions. This work was supported by award number MCB-0818122 from NSF and R01GM0825950 from NIGMS/NIH to K.S.L. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIGMS/NIH or the NSF.

3.7 References

Akgun E, Zahn J, Baumes S, Brown G, Liang F, Romanienko PJ, Lewis S, Jasin M (1997) Palindrome resolution and recombination in the mammalian germ line. *Molecular and cellular biology* **17:** 5559-5570

Alani E, Cao L, Kleckner N (1987) A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains. *Genetics* **116:** 541-545

Ashton TM, Hickson ID (2010) Yeast as a model system to study RecQ helicase function. *DNA Repair (Amst)* **9:** 303-314

Aylon Y, Kupiec M (2003) The checkpoint protein Rad24 of Saccharomyces cerevisiae is involved in processing double-strand break ends and in recombination partner choice. *Molecular and cellular biology* **23:** 6585-6596

Belli G, Gari E, Aldea M, Herrero E (1998) Functional analysis of yeast essential genes using a promoter-substitution cassette and the tetracycline-regulatable dual expression system. *Yeast* **14:** 1127-1138

Bhattacharyya S, Lahue RS (2004) Saccharomyces cerevisiae Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. *Molecular and cellular biology* **24:** 7324-7330

Brewer BJ, Fangman WL (1987) The localization of replication origins on ARS plasmids in S. cerevisiae. *Cell* **51:** 463-471

Bugreev DV, Rossi MJ, Mazin AV (2011) Cooperation of RAD51 and RAD54 in regression of a model replication fork. *Nucleic Acids Res* **39:** 2153-2164

Casper AM, Greenwell PW, Tang W, Petes TD (2009) Chromosome aberrations resulting from double-strand DNA breaks at a naturally occurring yeast fragile site composed of inverted ty elements are independent of Mre11p and Sae2p. *Genetics* **183:** 423-439, 421SI-426SI

Cejka P, Plank JL, Bachrati CZ, Hickson ID, Kowalczykowski SC (2010) Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3. *Nat Struct Mol Biol* **17:** 1377-1382

Collick A, Drew J, Penberth J, Bois P, Luckett J, Scaerou F, Jeffreys A, Reik W (1996) Instability of long inverted repeats within mouse transgenes. *EMBO J* **15:** 1163-1171

Connelly JC, Leach DR (1996) The sbcC and sbcD genes of Escherichia coli encode a nuclease involved in palindrome inviability and genetic recombination. *Genes Cells* **1:** 285-291

Cote AG, Lewis SM (2008) Mus81-dependent double-strand DNA breaks at in vivo-generated cruciform structures in S. cerevisiae. *Mol Cell* **31:** 800-812

Cromie GA, Millar CB, Schmidt KH, Leach DR (2000) Palindromes as substrates for multiple pathways of recombination in Escherichia coli. *Genetics* **154:** 513-522

Darmon E, Eykelenboom JK, Lincker F, Jones LH, White M, Okely E, Blackwood JK, Leach DR (2010) E. coli SbcCD and RecA control chromosomal rearrangement induced by an interrupted palindrome. *Mol Cell* **39:** 59-70

Dewar JM, Lydall D (2012) Similarities and differences between "uncapped" telomeres and DNA double-strand breaks. *Chromosoma* **121:** 117-130

Dixon, W.J., and Massey, F.J. Jr. (1969). Introduction to statistical analysis. . ed. (New York: McGraw-Hill, Inc.), p. 349.

Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences of the United States of America* **88:** 7160-7164

Eykelenboom JK, Blackwood JK, Okely E, Leach DR (2008) SbcCD causes a double-strand break at a DNA palindrome in the Escherichia coli chromosome. *Mol Cell* **29:** 644-651

Farah JA, Cromie G, Steiner WW, Smith GR (2005) A novel recombination pathway initiated by the Mre11/Rad50/Nbs1 complex eliminates palindromes during meiosis in Schizosaccharomyces pombe. *Genetics* **169:** 1261-1274

Farah JA, Hartsuiker E, Mizuno K, Ohta K, Smith GR (2002) A 160-bp palindrome is a Rad50.Rad32-dependent mitotic recombination hotspot in Schizosaccharomyces pombe. *Genetics* **161:** 461-468

Ford M, Fried M (1986) Large inverted duplications are associated with gene amplification. *Cell* **45:** 425-430

Gangavarapu V, Prakash S, Prakash L (2007) Requirement of RAD52 group genes for postreplication repair of UV-damaged DNA in Saccharomyces cerevisiae. *Molecular and cellular biology* **27:** 7758-7764

Gari K, Leon Ortiz AM, Borel V, Flynn H, Skehel JM, Boulton SJ (2012) MMS19 links cytoplasmic iron-sulfur cluster assembly to DNA metabolism. *Science* **337:** 243-245

Goldstein AL, McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast* **15:** 1541-1553

Gonzalez-Prieto R, Munoz-Cabello AM, Cabello-Lobato MJ, Prado F (2013) Rad51 replication fork recruitment is required for DNA damage tolerance. *EMBO J*

Gordenin DA, Lobachev KS, Degtyareva NP, Malkova AL, Perkins E, Resnick MA (1993) Inverted DNA repeats: a source of eukaryotic genomic instability. *Molecular and cellular biology* **13:** 5315-5322

Gordenin DA, Malkova AL, Peterzen A, Kulikov VN, Pavlov YI, Perkins E, Resnick MA (1992) Transposon Tn5 excision in yeast: influence of DNA polymerases alpha, delta, and epsilon and repair genes. *Proc Natl Acad Sci U S A* **89:** 3785-3789

Gotter AL, Nimmakayalu MA, Jalali GR, Hacker AM, Vorstman J, Conforto Duffy D, Medne L, Emanuel BS (2007) A palindrome-driven complex rearrangement of 22q11.2 and 8q24.1 elucidated using novel technologies. *Genome research* **17:** 470-481

Gotter AL, Shaikh TH, Budarf ML, Rhodes CH, Emanuel BS (2004) A palindrome-mediated mechanism distinguishes translocations involving LCR-B of chromosome 22q11.2. *Hum Mol Genet* **13:** 103-115

Grandin N, Bailly A, Charbonneau M (2005) Activation of Mrc1, a mediator of the replication checkpoint, by telomere erosion. *Biol Cell* **97:** 799-814

Grandin N, Damon C, Charbonneau M (2001) Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13. *EMBO J* **20:** 1173-1183

Grossi S, Puglisi A, Dmitriev PV, Lopes M, Shore D (2004) Pol12, the B subunit of DNA polymerase alpha, functions in both telomere capping and length regulation. *Genes & development* **18:** 992-1006

Guenthoer J, Diede SJ, Tanaka H, Chai X, Hsu L, Tapscott SJ, Porter PL (2012) Assessment of palindromes as platforms for DNA amplification in breast cancer. *Genome research* **22:** 232-245

Hashimoto Y, Ray Chaudhuri A, Lopes M, Costanzo V (2010) Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. *Nat Struct Mol Biol* **17:** 1305-1311

Inagaki H, Ohye T, Kogo H, Tsutsumi M, Kato T, Tong M, Emanuel BS, Kurahashi H (2013) Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations. *Nat Commun* **4:** 1592

Katou Y, Kanoh Y, Bando M, Noguchi H, Tanaka H, Ashikari T, Sugimoto K, Shirahige K (2003) S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature* **424:** 1078-1083

Kehrer-Sawatzki H, Haussler J, Krone W, Bode H, Jenne DE, Mehnert KU, Tummers U, Assum G (1997) The second case of a t(17;22) in a family with neurofibromatosis type 1: sequence analysis of the breakpoint regions. *Hum Genet* **99:** 237-247

Kerrest A, Anand RP, Sundararajan R, Bermejo R, Liberi G, Dujon B, Freudenreich CH, Richard GF (2009) SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. *Nat Struct Mol Biol* **16:** 159-167

Kogo H, Inagaki H, Ohye T, Kato T, Emanuel BS, Kurahashi H (2007) Cruciform extrusion propensity of human translocation-mediating palindromic AT-rich repeats. *Nucleic Acids Res* **35:** 1198-1208

Krogh BO, Symington LS (2004) Recombination proteins in yeast. *Annual review of genetics* **38:** 233-271

Kurahashi H, Emanuel BS (2001) Unexpectedly high rate of de novo constitutional t(11;22) translocations in sperm from normal males. *Nat Genet* **29:** 139-140

Kurahashi H, Inagaki H, Ohye T, Kogo H, Kato T, Emanuel BS (2006) Palindrome-mediated chromosomal translocations in humans. *DNA Repair (Amst)* **5:** 1136-1145

Kurahashi H, Shaikh T, Takata M, Toda T, Emanuel BS (2003) The constitutional t(17;22): another translocation mediated by palindromic AT-rich repeats. *Am J Hum Genet* **72:** 733-738

Leach DR (1994) Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *Bioessays* **16:** 893-900

Leach DR, Okely EA, Pinder DJ (1997) Repair by recombination of DNA containing a palindromic sequence. *Mol Microbiol* **26:** 597-606

Lee SE, Moore JK, Holmes A, Umezu K, Kolodner RD, Haber JE (1998) Saccharomyces Ku70, mre11/rad50 and RPA proteins regulate adaptation to G2/M arrest after DNA damage. *Cell* **94:** 399-409

Lemoine FJ, Degtyareva NP, Lobachev K, Petes TD (2005) Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell* **120:** 587-598

Letessier A, Millot GA, Koundrioukoff S, Lachages AM, Vogt N, Hansen RS, Malfoy B, Brison O, Debatisse M (2011) Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* **470:** 120-123

Lobachev KS, Gordenin DA, Resnick MA (2002) The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* **108:** 183-193

Lobachev KS, Shor BM, Tran HT, Taylor W, Keen JD, Resnick MA, Gordenin DA (1998) Factors affecting inverted repeat stimulation of recombination and deletion in Saccharomyces cerevisiae. *Genetics* **148:** 1507-1524

Lobachev KS, Stenger JE, Kozyreva OG, Jurka J, Gordenin DA, Resnick MA (2000) Inverted Alu repeats unstable in yeast are excluded from the human genome. *EMBO J* **19:** 3822-3830

Mangano R, Piddini E, Carramusa L, Duhig T, Feo S, Fried M (1998) Chimeric amplicons containing the c-myc gene in HL60 cells. *Oncogene* **17:** 2771-2777

Mankouri HW, Hickson ID (2007) The RecQ helicase-topoisomerase III-Rmi1 complex: a DNA structure-specific 'dissolvasome'? *Trends Biochem Sci* **32:** 538-546

Mimitou EP, Symington LS (2010) Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *EMBO J* **29:** 3358-3369

Miyake Y, Nakamura M, Nabetani A, Shimamura S, Tamura M, Yonehara S, Saito M, Ishikawa F (2009) RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. *Mol Cell* **36:** 193-206

Mizuno K, Lambert S, Baldacci G, Murray JM, Carr AM (2009) Nearby inverted repeats fuse to generate acentric and dicentric palindromic chromosomes by a replication template exchange mechanism. *Genes & development* **23:** 2876-2886

Mizuno K, Miyabe I, Schalbetter SA, Carr AM, Murray JM (2013) Recombination-restarted replication makes inverted chromosome fusions at inverted repeats. *Nature* **493:** 246-249

Narayanan V, Mieczkowski PA, Kim HM, Petes TD, Lobachev KS (2006) The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell* **125:** 1283-1296

Navadgi-Patil VM, Burgers PM (2009) A tale of two tails: activation of DNA damage checkpoint kinase Mec1/ATR by the 9-1-1 clamp and by Dpb11/TopBP1. *DNA Repair (Amst)* **8:** 996-1003

Neiman PE, Elsaesser K, Loring G, Kimmel R (2008) Myc oncogene-induced genomic instability: DNA palindromes in bursal lymphomagenesis. *PLoS Genet* **4:** e1000132

Neiman PE, Kimmel R, Icreverzi A, Elsaesser K, Bowers SJ, Burnside J, Delrow J (2006) Genomic instability during Myc-induced lymphomagenesis in the bursa of Fabricius. *Oncogene* **25:** 6325-6335

Netz DJ, Stith CM, Stumpfig M, Kopf G, Vogel D, Genau HM, Stodola JL, Lill R, Burgers PM, Pierik AJ (2012) Eukaryotic DNA polymerases require an iron-sulfur cluster for the formation of active complexes. *Nat Chem Biol* **8:** 125-132

Nimmakayalu MA, Gotter AL, Shaikh TH, Emanuel BS (2003) A novel sequence-based approach to localize translocation breakpoints identifies the molecular basis of a t(4;22). *Hum Mol Genet* **12:** 2817-2825

Oh SD, Jessop L, Lao JP, Allers T, Lichten M, Hunter N (2009) Stabilization and electrophoretic analysis of meiotic recombination intermediates in Saccharomyces cerevisiae. *Methods Mol Biol* **557:** 209-234

Petermann E, Orta ML, Issaeva N, Schultz N, Helleday T (2010) Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. *Mol Cell* **37:** 492-502

Qi H, Zakian VA (2000) The Saccharomyces telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein. *Genes & development* **14:** 1777-1788

Rooks H, Clark B, Best S, Rushton P, Oakley M, Thein OS, Cuthbert AC, Britland A, Ruf A, Thein SL (2012) A novel 506kb deletion causing epsilongammadeltabeta thalassemia. *Blood Cells Mol Dis* **49:** 121-127

Ruskin B, Fink GR (1993) Mutations in POL1 increase the mitotic instability of tandem inverted repeats in Saccharomyces cerevisiae. *Genetics* **134:** 43-56

Sandell LL, Zakian VA (1993) Loss of a yeast telomere: arrest, recovery, and chromosome loss. *Cell* **75:** 729-739

Schlacher K, Christ N, Siaud N, Egashira A, Wu H, Jasin M (2011) Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145:** 529-542

Schlacher K, Wu H, Jasin M (2012) A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22:** 106-116

Sheridan MB, Kato T, Haldeman-Englert C, Jalali GR, Milunsky JM, Zou Y, Klaes R, Gimelli G, Gimelli S, Gemmill RM, Drabkin HA, Hacker AM, Brown J, Tomkins D, Shaikh TH, Kurahashi H, Zackai EH, Emanuel BS (2010) A palindrome-mediated recurrent translocation with 3:1 meiotic nondisjunction: the t(8;22)(q24.13;q11.21). *Am J Hum Genet* **87:** 209-218

Sirbu BM, Couch FB, Feigerle JT, Bhaskara S, Hiebert SW, Cortez D (2011) Analysis of protein dynamics at active, stalled, and collapsed replication forks. *Genes & development* **25:** 1320-1327

Stehling O, Vashisht AA, Mascarenhas J, Jonsson ZO, Sharma T, Netz DJ, Pierik AJ, Wohlschlegel JA, Lill R (2012) MMS19 assembles iron-sulfur proteins required for DNA metabolism and genomic integrity. *Science* **337:** 195-199

Stenger JE, Lobachev KS, Gordenin D, Darden TA, Jurka J, Resnick MA (2001) Biased distribution of inverted and direct Alus in the human genome: implications for insertion, exclusion, and genome stability. *Genome research* **11:** 12-27

Sun J, Yu EY, Yang YT, Confer LA, Sun SH, Wan K, Lue NF, Lei M (2009) Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres. *Genes & development* **23:** 2900-2914

Tanaka H, Bergstrom DA, Yao MC, Tapscott SJ (2005) Widespread and nonrandom distribution of DNA palindromes in cancer cells provides a structural platform for subsequent gene amplification. *Nat Genet* **37:** 320-327

Tanaka H, Bergstrom DA, Yao MC, Tapscott SJ (2006) Large DNA palindromes as a common form of structural chromosome aberrations in human cancers. *Hum Cell* **19:** 17-23

Tanaka H, Cao Y, Bergstrom DA, Kooperberg C, Tapscott SJ, Yao MC (2007) Intrastrand annealing leads to the formation of a large DNA palindrome and determines the boundaries of genomic amplification in human cancer. *Molecular and cellular biology* **27:** 1993-2002

Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294:** 2364-2368

Tourriere H, Versini G, Cordon-Preciado V, Alabert C, Pasero P (2005) Mrc1 and Tof1 promote replication fork progression and recovery independently of Rad53. *Mol Cell* **19:** 699-706

Tsolou A, Lydall D (2007) Mrc1 protects uncapped budding yeast telomeres from exonuclease EXO1. *DNA Repair (Amst)* **6:** 1607-1617

Wach A, Brachat A, Pohlmann R, Philippsen P (1994) New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. *Yeast* **10:** 1793-1808

Waldman AS, Tran H, Goldsmith EC, Resnick MA (1999) Long inverted repeats are an at-risk motif for recombination in mammalian cells. *Genetics* **153:** 1873-1883

Wu PY, Nurse P (2009) Establishing the program of origin firing during S phase in fission Yeast. *Cell* **136:** 852-864

Zegerman P, Diffley JF (2003) Lessons in how to hold a fork. *Nat Struct Biol* **10:** 778-779

Zhang H, Lawrence CW (2005) The error-free component of the RAD6/RAD18 DNA damage tolerance pathway of budding yeast employs sister-strand recombination. *Proc Natl Acad Sci U S A* **102:** 15954-15959

Zhang Y, Shishkin AA, Nishida Y, Marcinkowski-Desmond D, Saini N, Volkov KV, Mirkin SM, Lobachev KS (2012) Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. *Mol Cell* **48:** 254-265

Zhu H, Shang D, Sun M, Choi S, Liu Q, Hao J, Figuera LE, Zhang F, Choy KW, Ao Y, Liu Y, Zhang XL, Yue F, Wang MR, Jin L, Patel PI, Jing T, Zhang X (2011) X-linked congenital hypertrichosis syndrome is associated with interchromosomal insertions mediated by a human-specific palindrome near SOX3. *Am J Hum Genet* **88:** 819-826

CHAPTER 4

CONCLUSIONS

This dissertation examined the mechanism of chromosomal instability induced by unstable DNA repeats in yeast *S. cerevisiae* by using GAA repeats and palindromic sequences as representative motifs. The systematic genome-wide screen allowed us to unbiasely identify genetic players that predispose the repeats to instability. Remarkably, the parallel screens revealed both global and specific mechanisms driving the instability of two repeats.

4.1 Global mechanisms driving repeats instability

With regards to the global mechanisms, we found that defects in 21 proteins that constitute the core of the replisome and replication-pausing checkpoint surveillance largely increase the instability of both GAA and inverted repeats. These data strongly suggests that intact replication is essential for governing the stability of unstable repetitive sequences in dividing cells. Moreover, the fragility and expansion potential of GAA tracts as well as the fragility of palindromic sequences are augmented by depletion of the telomere maintenance proteins. This observation brings forth an unorthodox role of these proteins in the bypass of replication barriers.

4.2 Specific mechanisms driving repeats instability

Notably, although both GAA and inverted repeats are repetitive sequences, they differ in the sequence compositions and the ability to form non-B structures. As a result, we also uncovered distinct mechanisms underlying their instability.

4.2.1 Transcription initiation promotes GAA repeats fragility and expansion

We found that GAA repeats can serve as promoters and recruit transcription initiation factors, which is likely due to the AT richness of the sequences and their ability to exclude nucleosomes. Initiation of aberrant transcription inside the tracts could have detrimental consequences. Consistent with this, defects in transcription initiation strongly induce both fragility and large expansions of the repeats. Also, fragility greatly increases in non-dividing cells where replication is not a player, and transcription initiation deficiency further amplifies the fragility.

4.2.2 Rad51 directs replication deficiency into fragility for palindromic sequences

Palindromic sequences can adopt hairpin or cruciform structured due to the intrinsic symmetry of the sequences. The follow-up analysis of the global screen revealed that DSBs generated at the repeats are likely a result of cruciform resolution in both wild-type and mutant strains. Significantly, we found that the homologous recombination protein Rad51 mediates the structure formation and DSB generation in replication deficient cells. Deletion of *RAD51* removes the replication block and increased breaks in replication mutants specifically. This Rad51-dependent fragility led us to propose that replication deficiency creates optimal conditions for the formation of a hairpin, which could be further converted into a cruciform via template switching mediated by Rad51. These data also predict existence of helicase removing the secondary structures in cells. Consistent with this, we found that, besides being an important player in homologous recombination, the dissolvasome prevents breakage at the palindromic sequences, probably through unwinding the hairpin or cruciform structures.

4.3 Future directions

Despite the exciting progress made by the genome-wide analysis of genetic factors promoting repeat instability in this study, future research is required to gain more in-depth knowledge of the mechanisms. Below are several clear research directions that have been inspired by this study:

4.3.1 How does transcription initiation stimulate GAA repeats instability?

Transcription initiation inside GAA tracts was identified as a major pathway besides replication for the repeat instability in the screen. However, we still lack the knowledge of detailed molecular steps that direct the transcription initiation defects into repeat instability. We hypothesize that the accumulation of triplex or R-loops during transcription mediates the fragility and expansion. It is therefore important to detect these abnormal structures using specific antibodies and compare the abundance of aberrant structures in the wild-type strains with those in the transcription initiation mutants. In addition, the phenomenon of repeat instability in non-dividing cells is still mysterious. We found that transcription initiation deficient strains augment GAA-induced breaks in non-dividing yeast cells. Hence, it is attractive to determine the impact of transcription initiation on GAA fragility in the wild-type strains. This could be achieved by abolishing transcription initiation using genetic silencers in non-dividing cells. Also, this assay would provide an opportunity for uncovering other pathways operating in the non-dividing cells, whose effect might be masked by the strong influence of the transcription initiation pathway.

4.3.2 How does the CST complex suppress the instability of GAA and palindromic repeats?

The screen strongly suggests an unorthodox role of the telomere maintenance complex CST in protecting the repeats from instability events via facilitating replication machinery to bypass difficult regions. It is of extreme interest to investigate this unexpected contribution of the complex in replication. Follow-up experiments would include ChIP analyses to test the association of these proteins with GAA or palindromic repeats; to check whether the complex travels along with replication machinery; and to find separation of function-mutants that are defective in replication maintenance but proficient in telomere regulation and to test their effect on repeat fragility and expansion.

4.3.3 What is the replication-independent pathway promoting palindromic repeats fragility?

Our research revealed that palindromic sequences cause chromosomal breaks through both replication-dependent and -independent pathways. While homologous recombination proteins mediate cruciform-formation and fragility under conditions of replication-deficiency, it is unclear who drives the structure-formation and repeat breakage in the case of replication proficiency. Well-designed genetic screens can be very helpful in searching for these mysterious protein mediators. It is also interesting to find out the responsible nucleases for making the breaks as well as their regulators, as these proteins could be promising targets for medical therapy.

Results of this thesis research provide novel mechanistic insights into the genome instability mediated by unstable repeats and will contribute to a broad spectrum of interests including human polymorphisms, genome integrity and rearrangements.

Follow-up experiments of the study is anticipated to shed more light on the mechanisms, which are important for identifying targets for the drug design for repeat-associated disorders.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Table A.1. Mutants identified in the genome-wide screen

Genetic background	Essential genes		Non-essential genes	Mutants recreated
	yTHC	DAmP	YKO	
DNA repair genes				
mre11			+[a]	+[b]
rad50			+	
xrs2			+	
sae2			+	+
tsa1			+	
Replication genes				
rfa2	+	+		+
pol12	+			+
pri2	+			+
pol1	+	+		
pol2				+
pol3	+			+
pol30	+	+		+
rfc2	+			+
rfc3	+			
rfc4	+			
rfc5	+			
dna2	+			+
mcm4	+			+
mcm5	+			
mcm7	+			
orc2	+			
orc4	+			+
rad27			+	+
rtt101			+	+
mms1			+	+
tof1			+	+
mrc1			+	+
csm3			+	
Telomere protection genes				
ten1		+		+
cdc13				+

Table A.1. (continued).				
Transcription initiation genes				
taf4	+			+
taf11	+			
taf12	+			
taf9				+
toa1				+
sua7				+
tfg1				+
spn1	+			+

[a] + indicates that effect of the mutant allele on GCRs was identified using this library

[b] + indicates mutant alleles that were used to calculate fragility rates and frequencies

Table A.2. Complementation of hyper-GCR phenotype of TET-*TAF4* and TET-*POL3* by centromeric vectors carrying *TAF4* and *POL3* expressed from native promoters

Genetic background	Rate of GCRs ($X10^{-7}$)		
	YCp50[a]	*TAF4* plasmid	*POL3* plasmid
$(GAA)_{230}$	27 (18-41)[b]	33 (19-50)	30 (26-80)
TET-*TAF4* with $(GAA)_{230}$	3500 (2800-4800)	22 (18-33)	ND
TET-*POL3* with $(GAA)_{230}$	150 (110-250)	ND[c]	27 (23-55)

[a] YCp50, centromeric vector carrying *URA3* gene, was used as a control.

[b] Numbers in parentheses correspond to the 95% confidence interval

[c] Not determined

Table A.3. Rates of GAA/TTC-induced GCRs with and without down-regulation of TET-*RFA2* and TET-*TAF4* using doxycycline

Genetic background	Rate of GCRs ($X10^{-7}$)	
	Without down-regulation	After down-regulation[a]
TET-*RFA2* $(GAA)_{120}$	270 (93-380)[b]	510 (210-1300)
TET-*TAF4* $(GAA)_{120}$	130 (110-170)	66 (47-110)

[a] 0.1 μg/ml doxycycline was used for down-regulation

[b] Numbers in parentheses correspond to the 95% confidence interval

Table A.4. Frequency of fragility at $(GAA)_{120}$ in wild-type and TET-*TAF4* strains in dividing and non-dividing cells

Time after inoculation (hours)	Fragility frequency ($x10^{-7}$)	
	Wild-type	TET-*TAF4*
37	5 (4-5)[a]	29 (12-38)
40	4 (2-9)	45 (16-110)
43	7 (2-10)	47 (43-160)
46	10 (3-14)	100 (53-150)
49	12 (6-18)	170 (100-300)
52	15 (10-23)	350 (140-440)
55	22 (14-24)	630 (440-770)
58	23 (17-37)	460 (290-520)
61	34 (31-46)	520 (360-840)
64	45 (30-54)	740 (610-890)
68	51 (34-66)	980 (720-1400)
82	100 (73-140)	1500 (900-2500)
93	150 (100-240)	2300 (1500-3500)
115	350 (160-480)	1900 (1600-3000)

[a] Numbers in parentheses correspond to the 95% confidence interval

Table A.5. Expansion frequencies of $(GAA)_{100}$ in dividing and non-dividing cells

Time after inoculation (hours)	Expansion frequency ($x10^{-5}$)	
	Wild-type	TET-*TAF4*
37	57 (39-76)[a]	210 (130-550)
43	35 (27-54)	700 (340-1400)
49	18 (11-23)	700 (600-1600)
55	12 (9-21)	860 (510-1000)
68	14 (6-21)	540 (380-830)
82	12 (8-23)	400 (320-520)
115	12 (10-19)	580 (350-800)

[a] Numbers in parentheses correspond to the 95% confidence interval

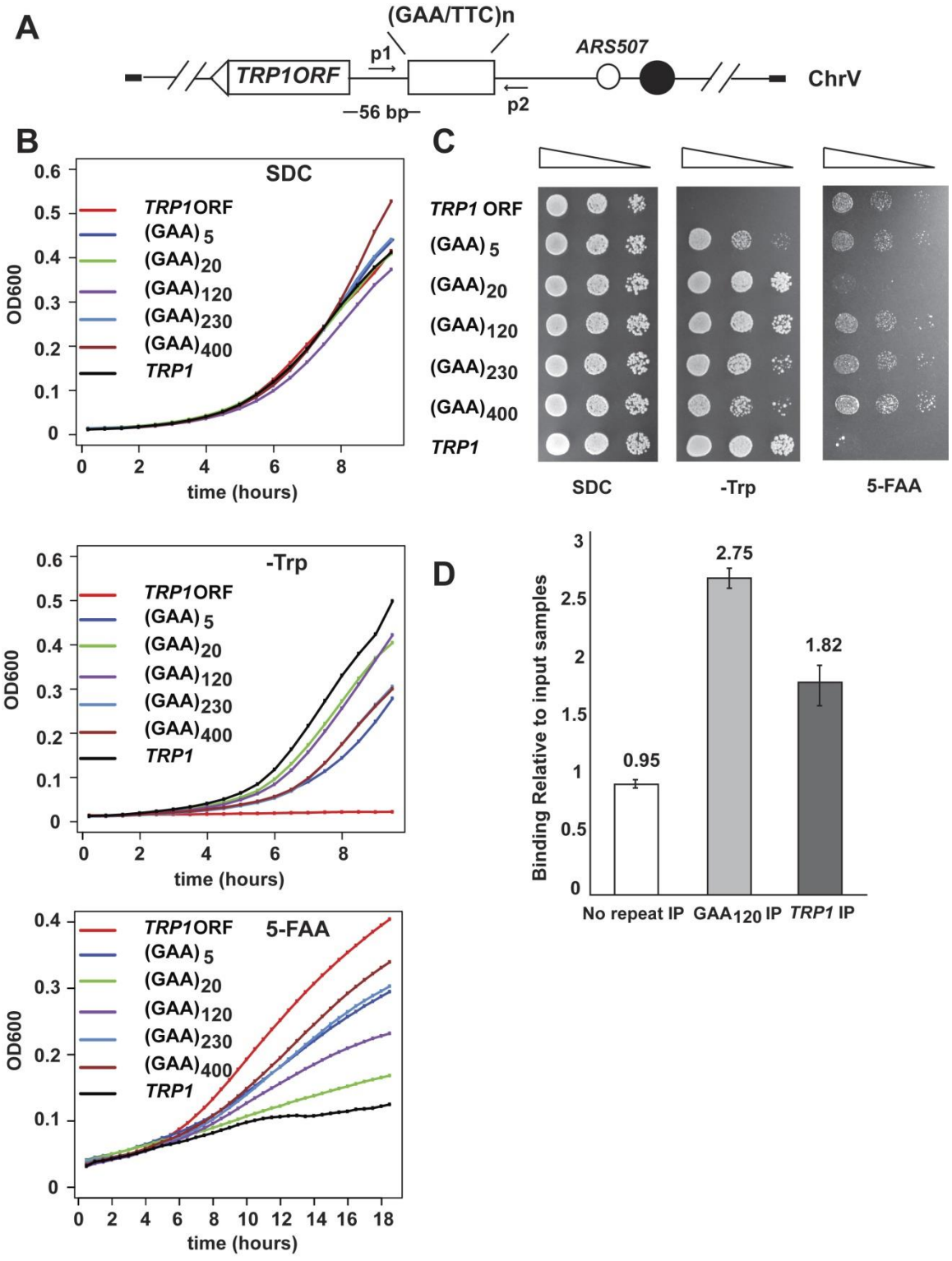

Figure A.1. GAA/TTC repeats can recruit transcription initiation factors and drive gene expression

(A) Diagram depicting the position of the repeats relative to *TRP1* ORF. Primers used for real-time qPCR are shown as black arrows. (B) Growth dynamics of GAA-containing and control strains in synthetic complete media (SDC), tryptophan drop-out media (-Trp) and on 5-FAA-containing media (5-FAA). (C) GAA/TTC repeats promote *TRP1* ORF expression. Strains contain 0, 5, 20, 120, 230, 400 GAA/TTC repeats in front of *TRP1* ORF. Strains with *TRP1* ORF expressed from its natural promoter are used as a positive control. 10-fold serial dilutions of yeast cells were plated on SDC, -Trp or 5-FAA medium. (D) GAA/TTC repeats recruit the transcription initiation factor Sua7. ChIP was carried out using TAP-tagged Sua7 in strains harboring no repeats, 120 GAA/TTC repeats, or the natural *TRP1* promoter. Precipitated DNA was quantified by real-time PCR using primers for the 5'end of *TRP1*. PCR amplifying the 3' region of *POL5* was used for data normalization. The graph shows relative enrichment of Sua7 in antibody treated samples compared with untreated samples. Experiments were done in triplicates; error bars represent the standard error of the mean.

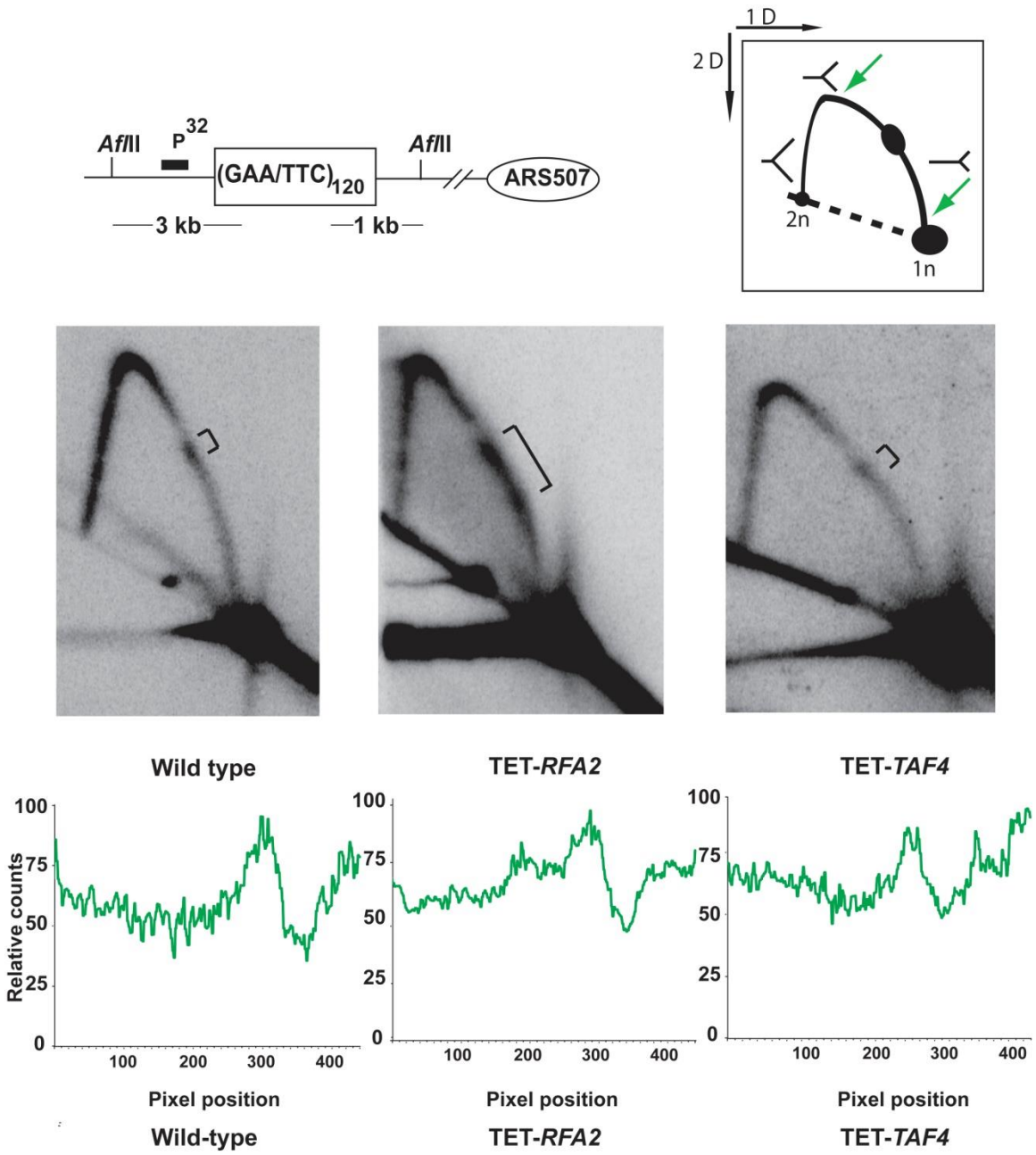

Figure A.2. 2D analysis of replication intermediates in wild-type, TET-*RFA2* and TET-*TAF4* strains

The upper panel shows relative position of repeats, restriction sites and *ARS507* (not to scale). Replication intermediates are highlighted by *LYS2* specific probe (black solid rectangle). Replication arrest at GAA/TTC tract is shown by bracket. Densitometry analysis of the Y-arc's long arm (marked by green arrows) is shown. Relative intensities of the arc are plotted against the distance from the monomers. The peaks depict the zones of replication arrests across the repeats. The TET-*RFA2* strain shows extended replication arrest in comparison with the wild-type and TET-*TAF4* strains.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

Table B.1. Hyper-GCR mutants identified in the genome-wide screen

Genetic background	Essential genes		Non-essential genes	Mutants recreated
	yTHC	DAmP	YKO	
Double strand breaks repair genes				
mre11			+[a]	*[b]
rad50			+	*
xrs2			+	
sae2			+	*
Replication genes				
orc2	+			*
orc4	+			*
mcm4	+			*
mcm5	+			
mcm7	+			
rfa2	+	+		*
pol12	+			
pri2	+			*
pol1	+	+		*
pol2				*
pol3	+		+	*
pol32				*
pol30	+	+		*
rfc2	+			*
rfc3	+			
rfc4	+			
rfc5	+			
dna2	+			*
rad27			+	*
yhr122w	+			*
mms19				*
Checkpoint response genes				
tof1			+	*
mrc1			+	

Table B.1. (continued).				
csm3			+	*
rad17			+	*
mec3			+	
ddc1			+	
rad24			+	*
dun1			+	+
Telomere maintenance genes				
ten1		+		*
cdc13				*
Sgs1-Top3-Rmi1 dissolvosome				
sgs1			+	*
rmi1			+	
ylr235c				

[a]+ shows mutants identified as hyper-GCR alleles from the libraries indicated.

[b]* shows mutant alleles whose effect on *Alu*-IR- mediated GCRs were determined by fluctuation tests.

Table B.2. Effect of *RAD51* deletion on *Alu*-IR-mediated GCR in mutants identified from the screen

Genetic background	GCR rate ($X10^{-6}$)	Fold increase over wild-type
WT (100% *Alu*-IRs)	41 (30-52)[a]	1
Δ*rad51*	37 (27-50)	1
TET-*POL30*	370 (290-390)	9
TET-*POL30*Δ*rad51*	67 (46-145)	1
TET-*POL1*	520 (500-800)	10
TET-*POL1*Δ*rad51*	170 (150-260)	4
TET-*RFA2*	250 (100-280)	6
TET-*RFA2*Δ*rad51*	92 (65-120)	2
TET-*MCM2*	150 (140-240)	4
TET-*MCM2*Δ*rad51*	44 (28-71)	1
TET-*ORC4*	110 (80-230)	3
TET-*ORC4*Δ*rad51*	60 (29-100)	1
TET-*TEN1*	140 (120-230)	3
TET-*TEN1*Δ*rad51*	40 (32-50)	1
TET-*YHR122W*	140 (110-160)	3
TET-*YHR122W*Δ*rad51*	48 (39-53)	1
Δ*csm3*	370 (270-530)	9
Δ*csm3*Δ*rad51*	22 (18-34)	0.5
Δ*sgs1*	260 (250-350)	6
Δ*sgs1*Δ*rad51*	52 (42-53)	1
Δ*rad17*	180 (160-250)	4
Δ*rad17*Δ*rad51*	100 (75-100)	2

[a]. Numbers in the brackets are 95% confidence intervals of the fluctuation tests

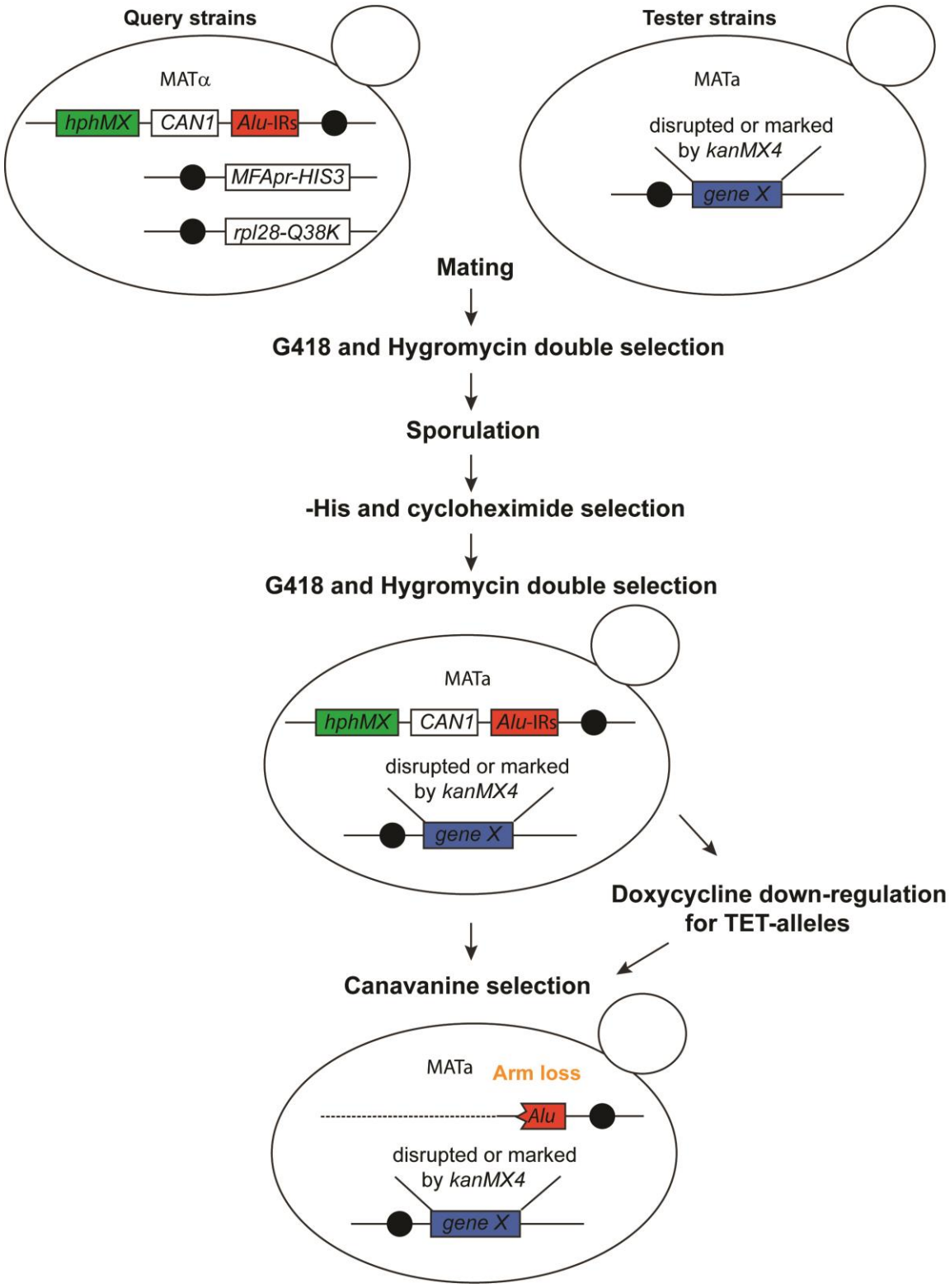

Figure B.1. The genome-wide screen scheme.

In the query strains, the chromosomal arm containing the GCR assay was marked by the *hphMX* cassette. The strains also carried a mating-type-regulated reporter MFApr-*HIS3* and a Q38K mutation in *RPL28* that rendered the strains resistant to cycloheximide. Both modifications serve as selection markers for the haploid strains during the screen. The tester strains were labeled with the *kanMX* cassette and consisted of three libraries: yTHC, DAmP and YKO (Open Biosystems). Each tester strain was crossed with duplicates of the query strains on YPD. The diploids were selected on medium supplemented with G418 and hygromycin and induced for sporulation. Haploid progeny (MATa) were selected on histidine drop-out medium supplemented with cycloheximide. Haploids containing both the repeats and the mutantion of interest were selected by G418- and hygromycin-containing medium. The strains were then replica plated to canavanine-containing medium to select for GCR events. For the yTHC library, doxycycline down-regulation (2 ug/ml) of the mutated alleles was applied prior to canavanine selection.

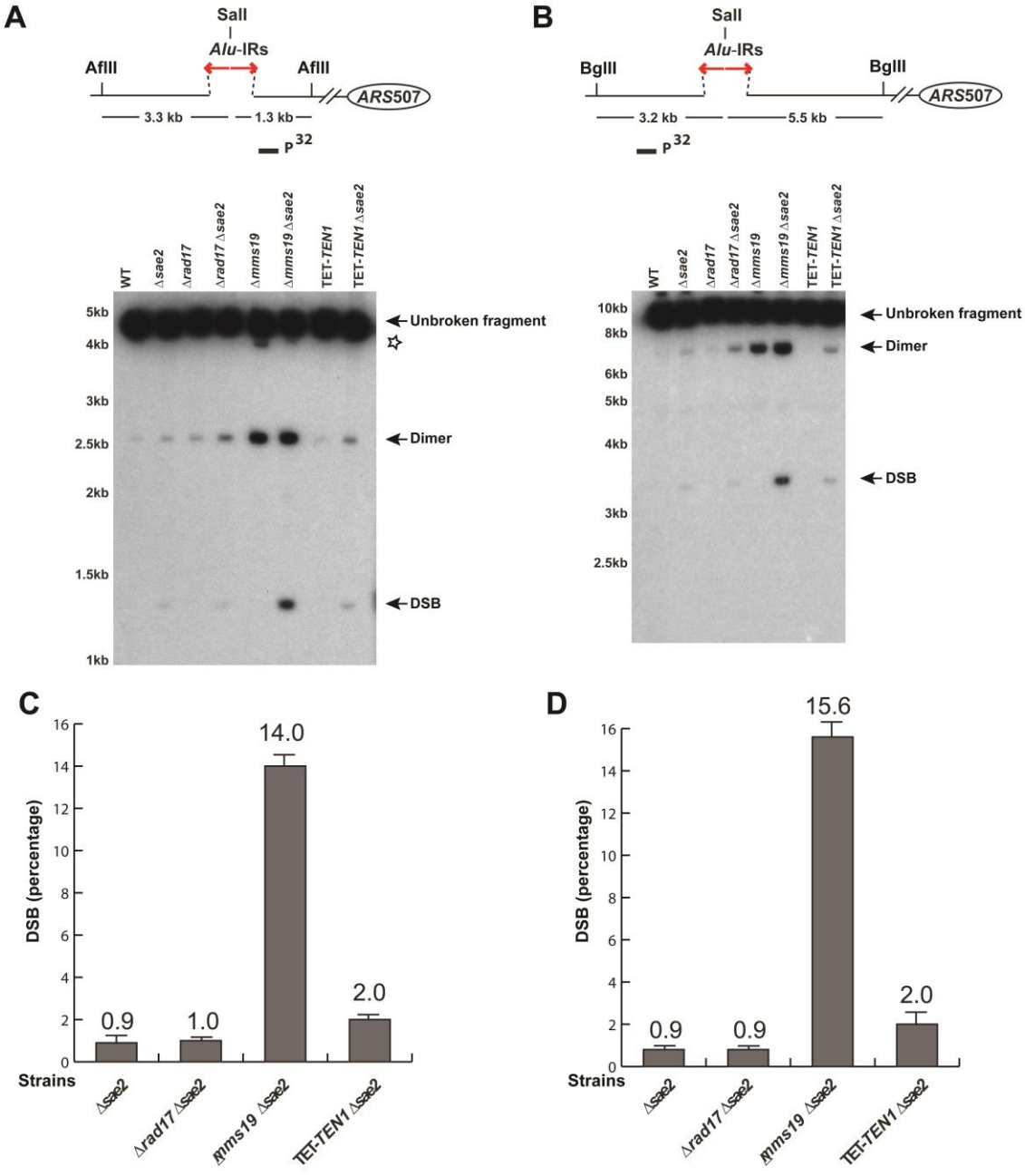

Figure B.2. Detection of breakage intermediates in a subset of hyper-GCR mutants. Genomic DNA embedded in agarose plugs were digested by AflII (**A**) or BglII (**B**) and processed for Southern hybridization as described in Figure 3.1. Strains included in the analysis are: wild-type, Δ*sae2*, Δ*rad17*, Δ*rad17*Δ*sae2*, Δ*mms19*, Δ*mms19*Δ*sae2*, TET-*TEN1*, TET-*TEN1*Δ*sae2*. Bands corresponding to the unbroken fragment, dimer and DSB fragment are indicated by arrows. The star indicates the bands below the unbroken

fragment in the Δ*mms19* and Δ*mms19*Δ*sae2* strains, which likely result from partial excision of the inverted repeats in these strains. (**C**) and (**D**) Densitometry analysis of the broken fragments normalized to the intact chromosome V in Δ*sae2* strains in (**A**) and (**B**), respectively. Values are shown as mean (shown on the top of the bars) with standard deviation obtained from at least three independent experiments.

PUBLICATIONS

Zhang Y, Saini N, Sheng ZW, Lobachev KS. "Genome-wide screen identifies replication-dependent pathway for palindrome fragility that requires homologous recombination". (submitted)

Kadyrova LY, Mertz TM, Zhang Y, Northam MR, Sheng Z, Lobachev KS, Shcherbakova PV, Kadyrov FA. "A reversible histone H3 acetylation cooperates with mismatch repair and replicative polymerases in suppressing genome instability". (submitted)

Saini N, Ramakrishnan S, Ayyar S, Zhang Y, Deem A, Elango R, Haber J, Lobachev KS, Malkova A. "Migrating bubble during break-induced replication drives conservative DNA synthesis". (in revision)

Saini N, Zhang Y, Nishida Y, Sheng ZW, Lobachev KS. (2013) "Fragile DNA motifs trigger mutagenesis at distant chromosomal loci in Saccharomyces cerevisiae" (accepted by *PLoS Genet*)

Saini N, Zhang Y, Usdin K, Lobachev KS. (2013) "When seocondary comes first – The importance of non-canonical DNA structures". *Biochimie*, 95(2):117-23.

Zhang Y, Shishkin AA, Nishida Y, Marcinkowski-Desmond D, Saini N, Volkov KV, Mirkin SM, Lobachev KS. (2012) "Genome-wide Screen Identifies Pathways that Govern GAA/TTC Repeat Fragility and Expansions in Dividing and Nondividing Yeast Cells". *Molecular Cell*, 48(2):254-65.