MECHANISMS OF CHROMOSOMAL INSTABILITY INDUCED BY UNSTABLE DNA REPEATS IN YEAST \textit{S. CEREBISIAE}

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

Copyright © Yu Zhang, 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

ACKNOWLEDGEMENTS ................................................................................................. III

LIST OF TABLES .............................................................................................................. VII

LIST OF FIGURES .............................................................................................................. VIII

LIST OF SYMBOLS AND ABBREVIATIONS ................................................................. IX

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW ........................................... 1

1.1 DNA repeats can adopt non-B DNA structures .................................................. 1

1.2 Non-B structure-adopting sequences and genome instability .............................. 2

1.2.1 Chromosomal fragility ..................................................................................... 3

1.2.2 Size variations ................................................................................................. 4

1.3 Models for non-B DNA-induced genome instability ........................................... 5

1.3.1 DNA replication ............................................................................................. 6

1.3.2 DNA transcription ......................................................................................... 7

1.3.3 DNA repair .................................................................................................... 9

1.4 Overview of the dissertation ................................................................................ 10

1.5 References .......................................................................................................... 12

CHAPTER 2: GENOME-WIDE SCREEN IDENTIFIES PATHWAYS THAT GOVERN GAA/TTC REPEAT FRAGILITY AND EXPANSIONS IN DIVIDING AND NON-DIVIDING YEAST CELLS ................................................................. 23

2.1 Summary .............................................................................................................. 23

2.2 Introduction ......................................................................................................... 24

2.3 Results ............................................................................................................... 28

2.3.1 Genome-wide screen of mutants that augment GAA/TTC fragility ................ 28

2.3.2 Hyper-fragility mutants identified in the screen .............................................. 30

2.3.3 TET-TAF4 and TET-RFA2 strains exhibit increased DSB formation ............ 35

2.3.4 GAA/TTC tracts can promote transcription .................................................... 37

2.3.5 Fragility is increased in non-dividing cells in a time-dependent manner and is amplified in TET-TAF4 strains ................................................................. 38
2.3.6 Mutations conferring hyper-fragility also induce large-scale repeat expansions ............ 41

2.4 Discussion .......................................................................................................................................................... 46

2.4.1 Defect in DNA replication ......................................................................................................................... 46

2.4.2 Defect in telomere maintenance .................................................................................................................. 48

2.4.3 Defect in transcription initiation ................................................................................................................ 49

2.4.4 Conclusions and implications for the human genome stability ................................................................. 53

2.5 Materials and Methods .................................................................................................................................... 55

2.5.1 Yeast Strains .................................................................................................................................................... 55

2.5.2 Schematics of the genome-wide screen .................................................................................................... 57

2.5.3 Measurement of GAA/TTC fragility and expansion rates ........................................................................ 58

2.5.4 Analysis of yeast growth dynamics ........................................................................................................... 59

2.5.5 DSB detection ................................................................................................................................................ 59

2.5.6 Analysis of the ability of GAA/TTC repeats to serve as promoter elements .......................................... 60

2.5.7 ChIP and qPCR ............................................................................................................................................ 60

2.5.8 2D analysis of replication intermediates .................................................................................................. 61

2.6 Acknowledgements .......................................................................................................................................... 62

2.7 References ........................................................................................................................................................... 63

CHAPTER 3: GENOME-WIDE SCREEN REVEALS REPLICATION PATHWAY FOR QUASI-PALINDROME FRAGILITY DEPENDENT ON HOMOLOGOUS RECOMBINATION ......................................................................................................................... 70

3.1 Summary ............................................................................................................................................................. 70

3.3 Results ................................................................................................................................................................ 74

3.3.1 Experimental systems used in the genome-wide screen ............................................................................. 74

3.3.2 Mutants with increased fragility at Alu-IRs ............................................................................................... 77

3.3.3 DSB formation is increased in replication-deficient and Δsgs1 mutants ................................................ 81

3.3.4 DSBs in replication-deficient strains have hairpin-capped termini ......................................................... 84

3.3.5 Alu-IR-mediated fragility and fork arrest are Rad51-dependent in replication-deficient strains................. 86

3.4 Discussion .......................................................................................................................................................... 91

3.4.1 Genome-wide screen identifies intact replication as a major guardian of quasi-palindrome stability ......................................................................................................................... 91

3.4.2 Analysis of DSB intermediates in replication-deficient mutants points towards cruciform-resolution mechanism of fragility ......................................................................................................................... 96
3.4.3 Rad51 is a key mediator of fragility in replication-deficient mutants ............................................ 96
3.5 Materials and Methods ......................................................................................................................... 100
3.5.1 Yeast strains ........................................................................................................................................ 100
3.5.2 Genome-wide screen scheme ............................................................................................................. 101
3.5.3 Measurement of GCR rates ............................................................................................................... 101
3.5.4 DSB detection ...................................................................................................................................... 102
3.5.5 2D neutral/neutral or neutral/alkaline gels for analyzing the structure of the broken ends........................................................................................................................................ 103
3.5.6 2D neutral/neutral gel analysis for replication fork progression ....................................................... 103
3.6 Acknowledgements ................................................................................................................................. 105
3.7 References ............................................................................................................................................... 106
CHAPTER 4: CONCLUSIONS .......................................................................................................................... 116
4.1 Global mechanisms driving repeats instability .......................................................................................... 116
4.2 Specific mechanisms driving repeats instability .......................................................................................... 116
4.2.1 Transcription initiation promotes GAA repeats fragility and expansion ........................................... 117
4.2.2 Rad51 directs replication deficiency into fragility for palindromic sequences .................................... 117
4.3 Future directions ....................................................................................................................................... 118
4.3.1 How does transcription initiation stimulate GAA repeats instability? .............................................. 118
4.3.2 How does the CST complex suppress the instability of GAA and palindromic repeats? ................. 119
4.3.3 What is the replication-independent pathway promoting palindromic repeats fragility? .............. 119
APPENDIX A: SUPPLEMENTARY INFORMATION FOR CHAPTER 2 .................................................. 121
APPENDIX B: SUPPLEMENTARY INFORMATION FOR CHAPTER 3 .................................................. 130
PUBLICATIONS ............................................................................................................................................. 137
## LIST OF TABLES

Table 2.1. Effect of mutants identified in the genome-wide screen on GAA-induced GCRs .................................................................................................................................................................................. 34

Table 2.2. Rates of expansions of (GAA)\textsubscript{100} in wild-type and hyper-GCR mutants ...... 43

Table 3.1. Mutants with increased Alu-IR-induced GCR rate .......................................................................................................................... 80

Table 3.2. Alu-IR-mediated fragility in TET-POL3 and TET-POL2 mutants is Rad51-dependent ........................................................................................................................................................ 87

Table A.1. Mutants identified in the genome-wide screen .......................................................... 121

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............. 123

Table A.3. Rates of GAA/TTC-induced GCRs with and without down-regulation of TET-RFA2 and TET-TAF4 using doxycycline ........................................................................................................ 124

Table A.4. Frequency of fragility at (GAA)\textsubscript{120} in wild-type and TET-TAF4 strains in dividing and non-dividing cells ......................................................................................................................... 125

Table A.5. Expansion frequencies of (GAA)\textsubscript{100} in dividing and non-dividing cells ... 126

Table B.1. Hyper-GCR mutants identified in the genome-wide screen ................................. 130

Table B.2. Effect of RAD51 deletion on Alu-IR-mediated GCR in mutants identified from the screen ............................................................................................................................... 132
LIST OF FIGURES

Figure 1.1  Examples of non-B DNA secondary structures........................................ 2

Figure 2.1.  Genome-wide screen methodology.......................................................... 29

Figure 2.2.  Physical detection of GAA/TTC-induced DSBs.......................................... 36

Figure 2.3.  GAA/TTC fragility increases in non-dividing cells...................................... 40

Figure 2.4. Expansion dynamics of (GAA)$_{100}$ repeats in actively dividing and non-
dividing cells in wild-type and TET-TAF4 strains....................................................... 44

Figure 2.5.  Model for GAA/TTC fragility and expansion in dividing and non-dividing
cells.................................................................................................................................... 52

Figure 3.1.  The genome-wide screen to identify hyper-GCR mutants............................ 76

Figure 3.2.  Physical detection of breakage intermediates in the wild-type and mutant
strains carrying 100% Alu-IRs........................................................................................ 83

Figure 3.3.  2D neutral/alkaline gel analysis of Alu-IR-induced DSBs in the TET-
POL3$\Delta$sae2 strain........................................................................................................ 85

Figure 3.4.  Detection of DSB accumulation in wild-type and mutant strains upon
deletion of RAD51.............................................................................................................. 89

Figure 3.5.  Analysis of replication fork progression through Alu-IRs in the wild-type and
mutant strains.................................................................................................................... 90

Figure 3.6.  Model for Alu-IRs-mediated fragility under conditions of replication
proficiency and deficiency................................................................................................. 99

Figure A.1.  GAA/TTC repeats can recruit transcription initiation factors and drive gene
expression......................................................................................................................... 127

Figure A.2.  2D analysis of replication intermediates in wild-type, TET-RFA2 and TET-
TAF4 strains.................................................................................................................... 129

Figure B.1.  The genome-wide screen scheme............................................................... 133

Figure B.2. Detection of breakage intermediates in a subset of hyper-GCR mutants... 135
# LIST OF SYMBOLS AND ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Alu</em>-IRs</td>
<td>Inverted <em>Alu</em> Repeats</td>
</tr>
<tr>
<td>CHEF</td>
<td>Contour-clamped Homogeneous Electric Field</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin Immunoprecipitation</td>
</tr>
<tr>
<td>CST</td>
<td>Cdc13-Stn1-Ten1</td>
</tr>
<tr>
<td>DSB</td>
<td>Double Strand Break</td>
</tr>
<tr>
<td>FRDA</td>
<td>Friedreich’s ataxia</td>
</tr>
<tr>
<td>GCR</td>
<td>Gross Chromosomal Rearrangement</td>
</tr>
<tr>
<td>MMR</td>
<td>Mismatch Repair</td>
</tr>
<tr>
<td>PATRRs</td>
<td>Palindromic AT-rich Repeats</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>2D gel analysis</td>
<td>Two-dimensional Gel Electrophoresis Analysis</td>
</tr>
<tr>
<td>5-FOA</td>
<td>5-Fluoroorotic Acid</td>
</tr>
<tr>
<td>5-FAA</td>
<td>5-Fluoroanthranilic Acid</td>
</tr>
</tbody>
</table>
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/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.
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-quaquoplex, 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


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


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. \(MAT\alpha\) query strains carrying GAA/TTC tract were crossed with \(MAT\alpha\) 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, \(MAT\alpha\) 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), Δmrc1 (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 CanR Ade+ white colonies, while arm loss events will give rise to CanR 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 nuclease 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 $tetO_7$–
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, TFIIIB and TFIIIF 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

<table>
<thead>
<tr>
<th>Genetic background</th>
<th>Rate of GCR (X10^-7)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(GAA)₀</td>
<td>(GAA)₁₂₀</td>
</tr>
<tr>
<td>wild-type</td>
<td>0.03 (0.02-0.04)ᵃ</td>
<td>3 (3-4)</td>
</tr>
<tr>
<td>Double-strand break repair genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δmre11</td>
<td>36 (27-42)</td>
<td>1200ᵇ</td>
</tr>
<tr>
<td>Δsae2</td>
<td>0.9 (0.2-1)</td>
<td>30</td>
</tr>
<tr>
<td>Replication genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TET-RFA2</td>
<td>4 (3-6)</td>
<td>133</td>
</tr>
<tr>
<td>TET-POL12</td>
<td>4 (1-8)</td>
<td>133</td>
</tr>
<tr>
<td>TET-PRI2</td>
<td>1 (1-2)</td>
<td>33</td>
</tr>
<tr>
<td>TET-POL3</td>
<td>0.5 (0.2-0.7)</td>
<td>17</td>
</tr>
<tr>
<td>TET-POL2</td>
<td>0.2 (0.2-0.5)</td>
<td>7</td>
</tr>
<tr>
<td>TET-POL30</td>
<td>0.6 (0.4-1)</td>
<td>20</td>
</tr>
<tr>
<td>TET-RFC2</td>
<td>0.2 (0.1-0.4)</td>
<td>7</td>
</tr>
<tr>
<td>TET-DNA2</td>
<td>0.5 (0.3-0.7)</td>
<td>17</td>
</tr>
<tr>
<td>TET-MCM4</td>
<td>2 (1-3)</td>
<td>67</td>
</tr>
<tr>
<td>TET-ORC4</td>
<td>3 (2-4)</td>
<td>100</td>
</tr>
<tr>
<td>Δrad27</td>
<td>12 (7-20)</td>
<td>400</td>
</tr>
<tr>
<td>Δrtt101</td>
<td>3 (2-3)</td>
<td>100</td>
</tr>
<tr>
<td>Δmms1</td>
<td>0.14 (0.1-0.4)</td>
<td>5</td>
</tr>
<tr>
<td>Δmrc1</td>
<td>4 (3-8)</td>
<td>133</td>
</tr>
<tr>
<td>Δtof1</td>
<td>2 (0.9-3)</td>
<td>67</td>
</tr>
<tr>
<td>Telomere maintenance genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TET-TEN1</td>
<td>0.2 (0.2-0.5)</td>
<td>7</td>
</tr>
<tr>
<td>TET-CDC13</td>
<td>2 (0.8-2)</td>
<td>67</td>
</tr>
<tr>
<td>Transcription initiation genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TET-TAF4</td>
<td>0.2 (0.07-0.2)</td>
<td>7</td>
</tr>
<tr>
<td>TET-TAF9</td>
<td>0.1 (0.05-0.2)</td>
<td>3</td>
</tr>
<tr>
<td>TET-TOA1</td>
<td>0.6 (0.4-1)</td>
<td>20</td>
</tr>
<tr>
<td>TET-SUA7</td>
<td>0.2 (0.1-0.4)</td>
<td>7</td>
</tr>
<tr>
<td>TET-TFG1</td>
<td>0.5 (0.2-1)</td>
<td>17</td>
</tr>
<tr>
<td>TET-SPNI</td>
<td>0.3 (0.2-0.4)</td>
<td>10</td>
</tr>
</tbody>
</table>

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 \textit{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 \textit{TRP1} gene with GAA/TTC tracts of different lengths. The repeats were placed 56 bps upstream from the ATG codon of \textit{TRP1} in the orientation where during transcription GAA strand will be the sense strand. As expected, strains containing promoter-less \textit{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 \textit{TRP1} expressed from its native promoter. Interestingly, growth was inhibited in strains with (GAA)$_{230}$ and (GAA)$_{400}$. Plating serial dilutions on \textit{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 \textit{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 (~1x10$^8$/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-FOAR).

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-FOAR\textsuperscript{R} colonies arose due to GAA/TTC expansions, but not due to changes in \textit{URA3} expression levels in transcription initiation mutants, expansions at the repeat locus were confirmed by PCR in both wild-type and TET-\textit{TAF4} strains. In addition, analysis of \textit{URA3} expression using reverse transcriptase PCR showed that wild-type and TET-\textit{TAF4} strains are comparable (data not shown).
### Table 2.2. Rates of expansions of (GAA)\textsubscript{100} in wild-type and hyper-GCR mutants

<table>
<thead>
<tr>
<th>Genetic background</th>
<th>Expansion rate(X10\textsuperscript{-5})</th>
<th>Fold increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild-type</td>
<td>1 (0.6-1) \textsuperscript{a}</td>
<td>1 \textsuperscript{b}</td>
</tr>
<tr>
<td>Double-strand break repair genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Δ}sae2</td>
<td>1 (0.6-2)</td>
<td>1</td>
</tr>
<tr>
<td>Replication genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TET-RFA2</td>
<td>171 (84-330)</td>
<td>171</td>
</tr>
<tr>
<td>TET-POL12</td>
<td>55 (49-82)</td>
<td>55</td>
</tr>
<tr>
<td>TET-PRI2</td>
<td>110 (77-190)</td>
<td>110</td>
</tr>
<tr>
<td>TET-POL3</td>
<td>11 (7-25)</td>
<td>11</td>
</tr>
<tr>
<td>TET-POL2</td>
<td>5 (2-12)</td>
<td>5</td>
</tr>
<tr>
<td>TET-POL30</td>
<td>26 (16-30)</td>
<td>26</td>
</tr>
<tr>
<td>TET-RFC2</td>
<td>26 (13-66)</td>
<td>26</td>
</tr>
<tr>
<td>TET-DNA2</td>
<td>8 (4-12)</td>
<td>8</td>
</tr>
<tr>
<td>TET-MCM4</td>
<td>23 (17-36)</td>
<td>23</td>
</tr>
<tr>
<td>TET-ORC4</td>
<td>18 (4-90)</td>
<td>18</td>
</tr>
<tr>
<td>\textit{Δ}rad27</td>
<td>20 (9-35)</td>
<td>20</td>
</tr>
<tr>
<td>\textit{Δ}rtt101</td>
<td>4 (2-31)</td>
<td>4</td>
</tr>
<tr>
<td>\textit{Δ}mms1</td>
<td>3 (2-4)</td>
<td>3</td>
</tr>
<tr>
<td>\textit{Δ}tof1</td>
<td>10 (13-24)</td>
<td>10</td>
</tr>
<tr>
<td>Telomere protection genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TET-TEN1</td>
<td>12 (9-17)</td>
<td>12</td>
</tr>
<tr>
<td>TET-CDC13</td>
<td>14 (10-29)</td>
<td>14</td>
</tr>
<tr>
<td>Transcription initiation genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TET-TAF4</td>
<td>58 (40-68)</td>
<td>58</td>
</tr>
<tr>
<td>TET-TAF9</td>
<td>37 (32-46)</td>
<td>37</td>
</tr>
<tr>
<td>TET-TOAI</td>
<td>14 (8-23)</td>
<td>14</td>
</tr>
<tr>
<td>TET-SUA7</td>
<td>19 (12-23)</td>
<td>19</td>
</tr>
<tr>
<td>TET-TFG1</td>
<td>22 (17-57)</td>
<td>22</td>
</tr>
<tr>
<td>TET-SPN1</td>
<td>2 (1-3)</td>
<td>2</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Numbers in parentheses correspond to the 95% confidence intervals

\textsuperscript{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-\textit{TAF4} strains (Figure 2.4 and Table A.5). Interestingly, the expansion frequency in TET-\textit{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](image1.png)

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

(A) Growth curves for wild-type and TET-\textit{TAF4} strains. Error bars represent standard error of the mean. (B) Frequencies of expansion of \((\text{GAA})_{100}\) 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 2.5IIId).

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 MATα 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 BamHI 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: MATα, 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 PdiI and PmelI 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 ClaI 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 ZraI 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 BgII and SalI sites of pISL-TRP1-ISR. (GAA)100 repeat was then cloned between the NcoI and ClaI 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 tetR gene of the pACYC184 plasmid (NEB) was amplified and cloned into SphI located in the intron of the split URA3 gene of the pISL-URA-TRP1-ISR plasmid. The plasmid was digested by SmI 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 doxycycline 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°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°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°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°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°, 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°C overnight. The membrane was washed twice in washing buffer (0.1X SSC, 0.1% SDS) at 70°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°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 AflII 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 40°C in the second dimension. Southern hybridization was carried out using a LYS2-specific probe at 64°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


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 αγδβ 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.*, 2010).
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/Nbs1 complex (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 ($5 \times 10^{-5}$) than in the strongest mutator $\Delta msh2$ ($6 \times 10^{-6}$), 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 tetO7 repressible promoter was used to replace the natural promoters of essential genes and regulate their expression (Belli et al,
The essential genes under the control of \textit{tetO}7 promoter will be referred to as TET-\textit{ORFs} 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 \textit{Alu}-IRs were inserted into the left arm of chromosome V. The selectable marker \textit{hphMX} for the presence of the chromosomal fragment containing the repeats and the counterselectable marker \textit{CAN1} used to assay GCR events are depicted. (B) Representative plate showing papillae on canavanine-containing medium reflecting the levels of \textit{Alu}-IR-induced GCRs in wild-type and mutants. Columns are duplicates of query strains containing 100\% homologous or 94\% homeologous \textit{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. \textit{ADE2} was inserted between \textit{CAN1} and the repeats. As a result, GCR events appear as red \text{Can}^R colonies and mutations at \textit{CAN1} give rise to white \text{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, Δddcl 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 \textit{RMI1} and of \textit{YLR235C} that partially overlaps with \textit{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-\textit{TEN1}, TET-\textit{STN1} and TET-\textit{CDC13}. The Ten1-Stn1-Cdc13 complex is involved in telomere maintenance and protection (Grandin \textit{et al}, 2001). Downregulation of Ten1 resulted in a 3-fold elevation of fragility (Table 3.1). The TET-\textit{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 \textit{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 \textit{et al}, 2002). We also showed that in \textit{\Delta mre11} 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 \textit{et al}, 2006). Predictably, the screen identified \textit{\Delta mre11} and \textit{\Delta rad50} as hyper-fragile mutants with a 10- and ~44-fold increase in GCRs induced by homologous and homeologous \textit{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

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

$a$ 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 Δsgs1Δsae2 can be rescued by the deletion of HDF1 (Mimitou & Symington, 2010). Therefore, the effect of Δsgs1 on DSB formation was assessed in the Δsgs1Δsae2Δhdf1 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 Δsae2 background. In TET-POL3, TET-POL2, Δcsm3, Δsgs1Δhdf1 (Figure 3.2), Δmms19, 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 Δrad17, 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 ARS507 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 \textit{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 \textit{Alu}-IRs strain with SalI + AflII (A) or SalI + BglII (B), where SalI cuts inside the 12 bp spacer of the \textit{Alu}-IRs. The strains used for analysis are: wild-type, Δsae2, TET-POL3, TET-POL3Δsae2, TET-POL2, TET-POL2Δsae2, Δcsm3, Δcsm3Δsae2, Δsgs1, Δsgs1Δhdf1Δsae2. (C) and (D) Densitometry analysis of the broken fragments normalized to the intact chromosome V in Δsae2 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-POL3Δsae2 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 \textit{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 the TET-POL3Δsae2 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 Δrad51 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-POL3Δsae2 and TET-POL2Δsae2 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).
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 Δrad51 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 Δrad51 background was not analyzed, the fact that the GCR levels in these strains decreased as

<table>
<thead>
<tr>
<th>Genetic background</th>
<th>GCR rate (X10⁻⁶)</th>
<th>Fold increase over wild-type</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>41 (30-52)ᵃ</td>
<td>1</td>
</tr>
<tr>
<td>Δrad51</td>
<td>37 (27-50)</td>
<td>1</td>
</tr>
<tr>
<td>TET-POL3</td>
<td>460 (390-640)</td>
<td>11</td>
</tr>
<tr>
<td>TET-POL3Δrad51</td>
<td>88 (58-108)</td>
<td>2</td>
</tr>
<tr>
<td>TET-POL2</td>
<td>240 (210-270)</td>
<td>5</td>
</tr>
<tr>
<td>TET-POL2Δrad51</td>
<td>63 (46-79)</td>
<td>1</td>
</tr>
</tbody>
</table>

ᵃ Numbers in brackets are the 95% confidence interval.
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: Δsae2, Δsae2Δrad51, TET-POL3Δsae2, TET-POL3Δsae2Δrad51, TET-POL2Δsae2, TET-POL2Δsae2Δrad51. (C) and (D) Densitometry analysis of the broken
fragments normalized to the intact chromosome V in $\Delta sae2$ 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-$POL3\Delta rad51$ strains were digested with AflIII 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 \textit{mrx} mutants. This increase is not due to frequent DSBs at \textit{Alu}-IRs, but rather a result of more efficient formation of dicentric chromosomes and loss of the broken acentric fragments. Consistently, \textit{\Delta mre11}, \textit{\Delta rad50}, and \textit{\Delta 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 (\textit{\Delta rad17}, \textit{\Delta mec3}, \textit{\Delta ddc1}, \textit{\Delta 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 \textit{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 \textit{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 (ARS507) 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 Δsgs1 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 hairpin-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, ura3Δ, leu2Δ, his3Δ, lys2Δ, rpl28-Q38K, mfa1Δ::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, bar1Δ, trp1Δ, his3Δ, ura3Δ, leu2Δ, ade2Δ, lys2Δ, 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 tetO7 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 Δsgs1Δhdf1Δsae2 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 $24 \times 10^8$ 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 AflIII or BglIII for 16 hr. Digested plugs were loaded in a 1% (AflIII digestion) or 0.7% (BglIII 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 °C 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 °C for 15 minutes. The gels were then run in 1 X alkaline buffer at 0.7 V/cm for 40 hr. at 4 °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-POL3Δrad51 strains. Cells were then released into fresh YPD. 50min after release, wild-type, TET-POL3, TET-POL3Δrad51 strains were
harvested and their genomic DNA samples were prepared as described in Friedman and Brewer, 1995. For the first dimensional gel electrophoresis, AflIII 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


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


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

<table>
<thead>
<tr>
<th>Genetic background</th>
<th>Essential genes</th>
<th>Non-essential genes</th>
<th>Mutants recreated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yTHC</td>
<td>DAmP</td>
<td>YKO</td>
</tr>
<tr>
<td>DNA repair genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mre11</td>
<td>+</td>
<td>+^a</td>
<td>+^b</td>
</tr>
<tr>
<td>rad50</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>xrs2</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>sae2</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>tsa1</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Replication genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rfa2</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>pol12</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pri2</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>pol1</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>pol2</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pol3</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>pol30</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>rfc2</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rfc3</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rfc4</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rfc5</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dna2</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mcm4</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mcm5</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mcm7</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orc2</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orc4</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>rad27</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>rtt101</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>mms1</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>tof1</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>mrc1</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>csm3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telomere protection genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ten1</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>cdc13</td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>
Table A.1. (continued).

<table>
<thead>
<tr>
<th>Transcription initiation genes</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>taf4</em></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>taf11</em></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>taf12</em></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>taf9</em></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td><em>toa1</em></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td><em>sua7</em></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td><em>tfg1</em></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td><em>spn1</em></td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th>Genetic background</th>
<th>Rate of GCRs (X10^{-7})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YCp50&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>(GAA)&lt;sub&gt;230&lt;/sub&gt;</td>
<td>27 (18-41)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>TET-TAF4 with (GAA)&lt;sub&gt;230&lt;/sub&gt;</td>
<td>3500 (2800-4800)</td>
</tr>
<tr>
<td>TET-POL3 with (GAA)&lt;sub&gt;230&lt;/sub&gt;</td>
<td>150 (110-250)</td>
</tr>
</tbody>
</table>

<sup>a</sup> YCp50, centromeric vector carrying URA3 gene, was used as a control.

<sup>b</sup> Numbers in parentheses correspond to the 95% confidence interval.

<sup>c</sup> Not determined
Table A.3. Rates of GAA/TTC-induced GCRs with and without down-regulation of TET-RA2 and TET-TAF4 using doxycycline

<table>
<thead>
<tr>
<th>Genetic background</th>
<th>Rate of GCRs (X10⁻⁷)</th>
<th>Without down-regulation</th>
<th>After down-regulation&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>TET-RA2 (GAA)₁₂₀</td>
<td>270 (93-380)&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>510 (210-1300)</td>
</tr>
<tr>
<td>TET-TAF4 (GAA)₁₂₀</td>
<td>130 (110-170)</td>
<td></td>
<td>66 (47-110)</td>
</tr>
</tbody>
</table>

<sup>a</sup> 0.1 µg/ml doxycycline was used for down-regulation
<sup>b</sup> 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

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

$^a$ Numbers in parentheses correspond to the 95% confidence interval
Table A.5. Expansion frequencies of (GAA)\textsubscript{100} in dividing and non-dividing cells

<table>
<thead>
<tr>
<th>Time after inoculation (hours)</th>
<th>Expansion frequency (x10\textsuperscript{-5})</th>
<th>TET-TAF4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wild-type</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>57 (39-76)\textsuperscript{a}</td>
<td>210 (130-550)</td>
</tr>
<tr>
<td>43</td>
<td>35 (27-54)</td>
<td>700 (340-1400)</td>
</tr>
<tr>
<td>49</td>
<td>18 (11-23)</td>
<td>700 (600-1600)</td>
</tr>
<tr>
<td>55</td>
<td>12 (9-21)</td>
<td>860 (510-1000)</td>
</tr>
<tr>
<td>68</td>
<td>14 (6-21)</td>
<td>540 (380-830)</td>
</tr>
<tr>
<td>82</td>
<td>12 (8-23)</td>
<td>400 (320-520)</td>
</tr>
<tr>
<td>115</td>
<td>12 (10-19)</td>
<td>580 (350-800)</td>
</tr>
</tbody>
</table>

\textsuperscript{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 \textit{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 \textit{TRP1 ORF} expression. Strains contain 0, 5, 20, 120, 230, 400 GAA/TTC repeats in front of \textit{TRP1 ORF}. Strains with \textit{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 \textit{TRP1} promoter. Precipitated DNA was quantified by real-time PCR using primers for the 5’end of \textit{TRP1}. PCR amplifying the 3’ region of \textit{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

<table>
<thead>
<tr>
<th>Genetic background</th>
<th>Essential genes</th>
<th>Non-essential genes</th>
<th>Mutants recreated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yTHC</td>
<td>DAmP</td>
<td>YKO</td>
</tr>
<tr>
<td>Double strand breaks repair genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>mre11</em></td>
<td></td>
<td>+^a</td>
<td>*b</td>
</tr>
<tr>
<td><em>rad50</em></td>
<td></td>
<td>+</td>
<td>*</td>
</tr>
<tr>
<td><em>xrs2</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>sae2</em></td>
<td></td>
<td>+</td>
<td>*</td>
</tr>
<tr>
<td>Replication genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>orc2</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>orc4</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>mcm4</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>mcm5</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>mcm7</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>rfa2</em></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>pol12</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>pri2</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>pol1</em></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>pol2</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>pol3</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>pol32</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>pol30</em></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>rfc2</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>rfc3</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>rfc4</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>rfc5</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>dna2</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>rad27</em></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td><em>yhr122w</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>mms19</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Checkpoint response genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>tof1</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>mrc1</em></td>
<td></td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>
Table B.1. (continued).

<table>
<thead>
<tr>
<th>Gene</th>
<th></th>
<th>+</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>csm3</td>
<td></td>
<td>+</td>
<td>*</td>
</tr>
<tr>
<td>rad17</td>
<td></td>
<td>+</td>
<td>*</td>
</tr>
<tr>
<td>mec3</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>ddc1</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>rad24</td>
<td></td>
<td>+</td>
<td>*</td>
</tr>
<tr>
<td>dun1</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Telomere maintenance genes

<table>
<thead>
<tr>
<th>Gene</th>
<th></th>
<th>+</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ten1</td>
<td></td>
<td>+</td>
<td>*</td>
</tr>
<tr>
<td>cdc13</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

Sgs1-Top3-Rmi1 dissolvsome

<table>
<thead>
<tr>
<th>Gene</th>
<th></th>
<th>+</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>sgs1</td>
<td></td>
<td>+</td>
<td>*</td>
</tr>
<tr>
<td>rmi1</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>ylr235c</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{a+}\) 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

<table>
<thead>
<tr>
<th>Genetic background</th>
<th>GCR rate (X10^6)</th>
<th>Fold increase over wild-type</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT (100% Alu-IRs)</td>
<td>41 (30-52)</td>
<td>1</td>
</tr>
<tr>
<td>Δrad51</td>
<td>37 (27-50)</td>
<td>1</td>
</tr>
<tr>
<td>TET-POL30</td>
<td>370 (290-390)</td>
<td>9</td>
</tr>
<tr>
<td>TET-POL30Δrad51</td>
<td>67 (46-145)</td>
<td>1</td>
</tr>
<tr>
<td>TET-POL1</td>
<td>520 (500-800)</td>
<td>10</td>
</tr>
<tr>
<td>TET-POL1Δrad51</td>
<td>170 (150-260)</td>
<td>4</td>
</tr>
<tr>
<td>TET-RFA2</td>
<td>250 (100-280)</td>
<td>6</td>
</tr>
<tr>
<td>TET-RFA2Δrad51</td>
<td>92 (65-120)</td>
<td>2</td>
</tr>
<tr>
<td>TET-MCM2</td>
<td>150 (140-240)</td>
<td>4</td>
</tr>
<tr>
<td>TET-MCM2Δrad51</td>
<td>44 (28-71)</td>
<td>1</td>
</tr>
<tr>
<td>TET-ORC4</td>
<td>110 (80-230)</td>
<td>3</td>
</tr>
<tr>
<td>TET-ORC4Δrad51</td>
<td>60 (29-100)</td>
<td>1</td>
</tr>
<tr>
<td>TET-TEN1</td>
<td>140 (120-230)</td>
<td>3</td>
</tr>
<tr>
<td>TET-TEN1Δrad51</td>
<td>40 (32-50)</td>
<td>1</td>
</tr>
<tr>
<td>TET-YHR122W</td>
<td>140 (110-160)</td>
<td>3</td>
</tr>
<tr>
<td>TET-YHR122WΔrad51</td>
<td>48 (39-53)</td>
<td>1</td>
</tr>
<tr>
<td>Δcsm3</td>
<td>370 (270-530)</td>
<td>9</td>
</tr>
<tr>
<td>Δcsm3Δrad51</td>
<td>22 (18-34)</td>
<td>0.5</td>
</tr>
<tr>
<td>Δsgs1</td>
<td>260 (250-350)</td>
<td>6</td>
</tr>
<tr>
<td>Δsgs1Δrad51</td>
<td>52 (42-53)</td>
<td>1</td>
</tr>
<tr>
<td>Δrad17</td>
<td>180 (160-250)</td>
<td>4</td>
</tr>
<tr>
<td>Δrad17Δrad51</td>
<td>100 (75-100)</td>
<td>2</td>
</tr>
</tbody>
</table>

*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 mutation 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, Δsae2, Δrad17, Δrad17Δsae2, Δmms19, Δmms19Δsae2, TET-TEN1, TET-TEN1Δsae2. 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 Δmms19 and Δmms19Δsae2 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 Δsae2 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.
Zhang Y, Saini N, Sheng ZW, Lobachev KS. “Genome-wide screen identifies replication-dependent pathway for palindrome fragility that requires homologous recombination”. (submitted)


