DISSERTATION

INVESTIGATION OF MECHANISMS OF MITOTIC RECOMBINATION IN YEAST

Submitted by

Lisa Victoria Harcy

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2016

Doctoral Committee:

Advisor: Lucas Argueso

Santiago Di Pietro
Howard Liber
Erica Suchman

ABSTRACT

INVESTIGATION OF MECHANISMS OF MITOTIC RECOMBINATION IN YEAST

At the submicroscopic level within all living cells the workings of a dynamic molecular world attempt to preserve the integrity of DNA - the blueprint of life. This dissertation describes in detail the experimental systems and results from two of our studies conducted in which DNA lesions compromised genomic integrity. The unifying theme of the following chapters revolves around the mechanisms responsible for structural genomic variation, that is, what happens to chromosomes when they break.

In the first phase of research, we examined outcomes associated with double-strand breaks (DSBs) at G-quadruplex DNA sequences. Here, we show that G4 DNA motifs, capable of initiating double-strand breaks, result in the formation of chromosomal aberrations. Our results provide structural context and support to a putative mechanism of homology-directed repair revealed by molecular analysis of these complex rearrangements.

In the second project, we investigated chromosomal translocations that developed from spontaneously occurring DSBs in diploid yeast to ascertain the means by which the DSBs had been repaired. The objective of this study was to examine the intricate interplay between the non-allelic DSB repair (DSBR) processes: canonical reciprocal homologous recombination (CRHR) vs. break-induced replication (BIR).

While numerous assays have previously measured repair pathway efficiency by isolating one of the two pathways from the other, no prior studies to our knowledge, have prospectively examined the contribution of BIR to overall double-strand break repair in diploids, using a non-

inducible experimental system where either mechanism can be used freely by the cells. I designed and constructed a new assay system to study chromosomal translocations in the yeast eukaryotic model to investigate the balance between these two DSB repair processes. I characterized genotypic and phenotypic alterations in wild type background and in mutant yeast strains defective for BIR.

The data obtained from this work provides an additional perspective to the field of DNA repair biology with broad relevance to DSBR regulation in eukaryotes. It provides further understanding about the role of DNA repair to undesired genetic outcomes, thus, leading the way to the design of new and more effective treatments for diseases in which these molecular actions are the instigators of pathogenesis.

ACKNOWLEDGEMENTS

First and foremost, I express my profound gratitude and sincere thanks to my advisor, Dr. Lucas Argueso. I could not imagine having a better mentor to guide me to the completion of my graduate education. I am truly indebted to him for allowing me the opportunity to work in his laboratory, for his unbounded encouragement, and for sharing his encyclopedic knowledge. Without his limitless patience and valuable advice, I would never have fully comprehended the difficult concepts and sorted out the technical details pertinent to my research experiments. On account of the personal attention and training I received, I am a more self-confident and proficient scientist, capable of thinking independently. I am immensely fortunate to have had an advisor who always whole-heartedly supported my career aspirations and professional development. Together, we established a clear path to the completion of the research objectives of my thesis. He reignited the passion for science that brought me to Colorado State University through his contagious enthusiasm for the subject matter and related research. He fostered the same engagement and inspiration in me during my years as his student. He is an excellent supervisor with a strong work ethic, a sincere educator, and an articulate mentor who has a genuine interest in his students' success. For all of these things, I am thankful.

I offer my sincere appreciation to the rest of my graduate advisory committee:
Dr. Erica Suchman, Dr. Howard Liber and Dr. Santiago Di Pietro. They imparted their wisdom, provided constructive criticism and challenged me intellectually. Most importantly, a special thank you to these professors, along with my advisor, for believing in me even when I doubted myself, which invigorated me to overcome many crises and finish my graduate education.

Dr. Suchman, a long time mentor of mine, provided continuous guidance at many points along my academic journey and propelled me forward on my journey beyond CSU. Additionally, I am grateful for Donna Weedman for her tremendous support during my graduate studies. I also thank everyone who scrupulously proofread and made improvements upon countless drafts of this written work.

I acknowledge the insightful questions and suggestions at different stages of my research by departmental colleagues and fellow researchers in the field of DNA damage and repair. Their comments were thought-provoking and focused my ideas and improved my knowledge base. I thank my fellow labmates for enriching my graduate school experience and for creating an overall positive and enjoyable work atmosphere in which to do science. I also appreciate the efforts of the undergraduate students who contributed to various aspects of the research process.

I thank my parents, family members and friends who aided me throughout my endeavors, both directly and indirectly. I express my deepest gratitude to Matthew, for his steadfast devotion, unconditional love and unwavering support throughout the years I have been in graduate school and in life in general. I am excited to see what adventures life takes us on next!

Lastly, I appreciate the financial support from the Boettcher Foundation and the Graduate Assistance in Areas of National Need (GAANN) teaching fellowship program that funded parts of the research discussed in this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2

Chapter 3

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

LIST OF ACRONYMNS

BIR: Break-induced replication

CEN: Centromere

Chr: Chromosome

CNV: Copy number variation

CRHR: Canonical reciprocal homologous recombination

DSB: Double-strand break

DSBR: Double-strand break repair

dsDNA: Double-stranded DNA

G4 DNA: G-quadruplex DNA

GCR: Gross chromosomal rearrangement

HCC: Half-crossover cascade

HDR: Homology-directed repair

HR: Homologous recombination

LOH: Loss of heterozygosity

LTR: Long terminal repeat

MMS: Methyl methanesulfonate

NAHR: Non-allelic homologous recombination

NRT: Non-reciprocal translocation

Pol: Polymerase

ssDNA: Single-stranded DNA

TEL: Telomere

Ty: Transposable element in yeast

WGS: Whole genome sequencing

Chapter One

Investigation of Mechanisms of Mitotic Recombination in Yeast

The topics pertinent to this dissertation concerning DNA double-strand break repair have been extensively described in minute detail by two recent review articles (Donnianni and Symington 2016; Verma and Greenberg 2016). For the purposes of this introduction, I have highlighted some of the aspects that are most relevant to the scope of my two experimental chapters that follow.

Introduction

Consequences of genomic rearrangements

Genomic instability is generally defined as the disruption of the capacity for cells to faithfully maintain and transmit genetic information from cell to cell, parent to child. While genetic damage cannot be entirely avoided, attempts to preserve chromosomal integrity, and thus avoid cell death, can result in introduced mutations and structural deviations. Indisputably, the most deleterious DNA lesion is the double-strand break (DSB) in which both strands of the double-helix are severed (Sakofsky et al. 2012; Morales et al. 2015). DSBs may arise not only from exogenous factors such as ionizing radiation and genotoxic chemicals, but also result from endogenous sources such as oxidative damage and replication stress, to name a few.

Genomes are composed of dynamic architecture. Within cells, there is continuous surveillance and constant repair of genetic material by evolutionarily conserved molecular machinery and mechanisms ongoing (Jackson 2002; Kachroo et al. 2015). Furthermore, since

many of the proteins responsible for DNA maintenance in humans have orthologs in budding yeast, *Saccharomyces cerevisiae*, it provides an excellent model organism for the study of these fundamental processes (Aggarwal and Brosh 2012).

One way that cells preserve genomic stability is through a robust DNA replication system. Normal S-phase replication proceeds by a semi-conservative mechanism with high fidelity resulting in a relatively rare occurrence of nucleotide misincorporation, approximately 10^{-10} mutations per base pair per cell division (Preston et al. 2010). This low frequency, however, cannot fully explain evolution and cancer development, both of which show a significant accumulation of mutations (Malkova and Haber 2012). Many recent whole genome sequencing (WGS) studies demonstrated that within cancer genomes the level of mutagenesis (often associated with chromosomal translocations) cannot be explained by this basal mutation frequency (Chin et al. 2012). This phenomena of massive chromosomal changes underpins the importance of attempting to model particular DNA repair situations due to their significant biological consequences.

The focus of our research links double-strand breaks repair (DSBR) to gross chromosomal rearrangements (GRCs). The modes of recombination often involve DNA synthesis that is potentially different from the type of DNA synthesis occurring during S-phase where initiation occurs at an origin of replication (Kraus et al. 2001). If instead, synthesis initiated at a DSB, this would mean there is a break occurring in a chromosome in such a way that homology within the broken chromosome would be exposed, allowing strand invasion and synthesis from an alternative template.

Break-Induced Replication mechanism

Emerging lines of evidence suggest that break-induced replication (BIR) is an under-recognized method by which genomic rearrangements occur in human cells (Llorente et al. 2008b). BIR is a variant of homologous recombination (HR) in that it involves repair of only one side of a DSB (Figure 1.1). 5' to 3' degradation at the DSB exposes a 3' unpaired end, necessitating invasion into a homologous DNA duplex to serve as a template for repair. Following strand invasion, DNA synthesis proceeds to the end of the chromosome, resolving the DSB (Kolodner 2000; Donnianni and Symington 2015).

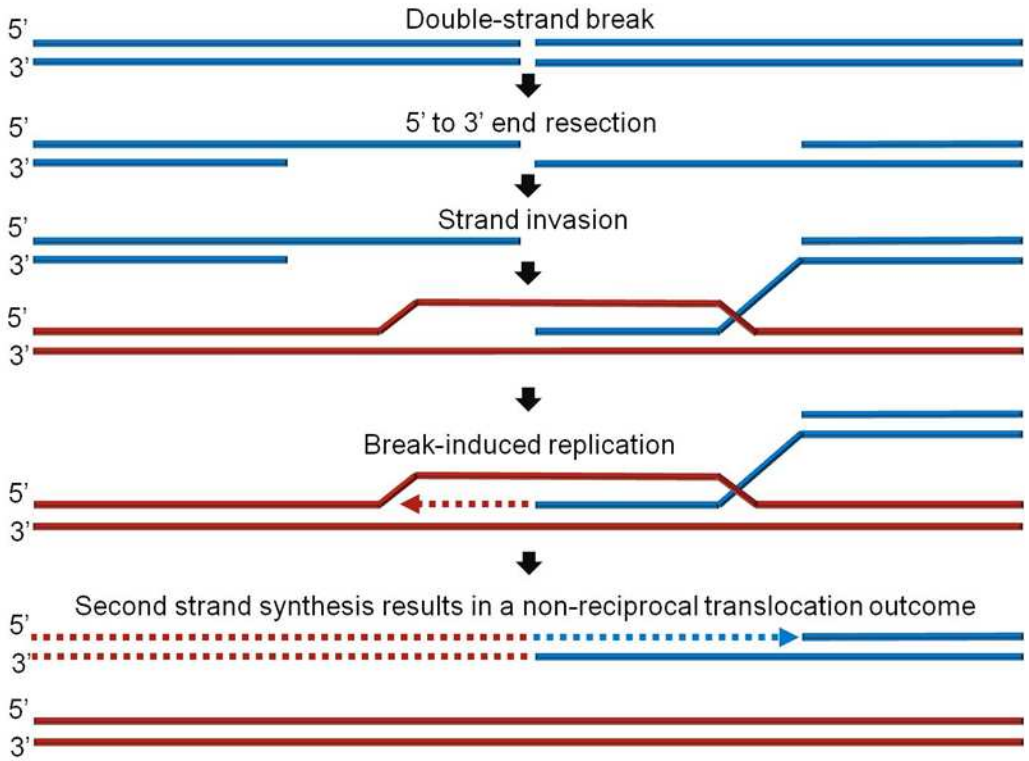


Figure 1.1. Mechanism of break-induced replication.
The repair process is initiated by a DNA double-strand break exposing broken chromosome ends. Exonucleases resect the DNA in a 5' to 3' manner exposing an unpaired 3' end. In this representation, the left side of the blue chromosome is lost. Aided by recombination proteins, the free 3' end invades a non-allelic template based on homology pairing. Extensive DNA synthesis (represented by dashed lines) proceeds in a migrating D-loop primed from the 3' invading end (indicated by arrowheads). Image adapted from (Saini et al. 2013)

While BIR is not the most physiologically predominant means of DSBR, this non-canonical homology-directed repair (HDR) strategy plays a dual role in cells. BIR repairs breaks that cannot be repaired by other means, recovers one-ended break events such as broken replication forks and can be useful for restoring eroded telomeres. Unfortunately, BIR is not without drawbacks. BIR leads to loss of heterozygosity (LOH) by copying entire regions of allelic chromosomes and can alternate between copying many sources or "donor sequences", a process referred to as template switching (Smith et al. 2007; Stafa et al. 2014). Consequently, BIR instigates the formation of non-reciprocal translocations (NRTs), extensively discussed in Chapter 3 of this dissertation (Kraus et al. 2001; Sakofsky et al. 2012). Additionally, multiple instances of BIR can potentiate genomic degeneration into complex genomic rearrangements, including ones termed half-crossover cascades (HCC), described in Future Directions and Studies of Chapter 3 (Vasan et al. 2014). It has also been demonstrated by Malkova *et. al.* that the type of replication in BIR is extremely mutagenic, proceeding via a migrating bubble similar to that formed during transcription (Deem et al. 2011b; Sakofsky et al. 2012; Saini et al. 2013).

Since the DNA repair process itself can be mutagenic, it can lead to a variety of large-scale amplifications and/or deletions, known as copy number variations (CNVs) (Gu et al. 2008; Zarrei et al. 2015). CNVs can introduce neutral polymorphic variation and have been identified as drivers of genetic diversity across organisms, but also can introduce pathologic variation and serve as promoters of premature aging and causes for a myriad of diseases, including cancer (Conover et al. 2015; Mishra and Whetstine 2016). As such, targeting alternative DSBR mechanisms provides a promising prospect for designing clinical therapeutics for use in personalized medicine interventions (Carvalho and Lupski 2016; Verma and Greenberg 2016).

Research project descriptions

Here, I provide an overview of the two yeast-based bioassays conducted to investigate the outcomes of mitotic recombination. Both studies comprising the research results contained within this dissertation revolve around characterizing the nature of gross chromosomal rearrangements (GCRs) (Bell et al. 2014; Putnam et al. 2016). In the first study, we performed several molecular typing techniques to determine the precise genetic makeup of a spectrum of GCRs obtained from CNV assays. We established that G-quadruplex DNA motifs consisting of self-associating guanine-rich tracts of DNA can cause a helical-distortion and serve as a hot spot for recombination through a micro-homology mediated mechanism. We were able to map the breakpoints at these regions where translocations were initiated for most of the clones analyzed. Technically speaking however, a "breakpoint" is a misnomer because the DNA most likely had been resected back some distance from the initial DSB before repair began; making the sites of recombination we detected actually "repair points." Regardless of the nomenclature, the actual sequencing data we obtained allowed us to infer the mechanism of formation for these GCRs, described in Chapter 2.

The objective of the second study was to determine the relative contribution of two competing DNA repair pathways to the production of NRT outcomes. Prior data did not allow distinguishing between the alternative mechanisms that can form NRTs. Ultimately, my intention was to gauge the footprint of BIR in homology-directed repair and to provide a more comprehensive understanding of the balance between non-allelic homologous recombination (NAHR) pathways that involve regions of DNA that have high sequence identity, but are not alleles, or between dispersed repetitive DNA elements. I quantitatively measured mutation rate based on colonies that would only form on selective media following a specific genomic rearrangement. Additionally, I sought to analyze whether the mutation rates were affected by

integrating the donor sequence at different loci. Finally, I investigated how disrupting BIR affected the basal rate of recombination by creating *POL32* deletion strains. *POL32* encodes a non-essential subunit of a replicative polymerase required for BIR (Payen et al. 2008a). The dramatic and unexpected effects observed in these deficient strains speaks to the importance of reliable and processive DNA replication, as previously mentioned.

The discovery of genomic structural variants in humans still has many challenges to overcome. However, the advancements in sequencing technology and bioinformatics approaches have already begun to spotlight the fact that previously identified, but perhaps less appreciated processes discovered in model organisms do indeed have significant relevance for better understanding the nature of human DNA (Onishi-Seebacher and Korbel 2011; Startek et al. 2015a). Continued investigation of these alternative DSB repair mechanisms in experimentally tractable model systems will present new mechanistic and functional insights into the roles they play in genome structure and related diseases (O'Driscoll and Jeggo 2006).

In a broader context

The genome stability experimental model system mentioned above to investigate NRTs and further described in Chapter 3 is an innovative and unique tool for characterizing the contribution of break-induced replication to the DNA repair process. There are some important distinctions to note between our design and the conventional BIR assay systems that have been described previously. Firstly, our experimental CNV-detection assay system involves diploid yeast cells, while most, if not all other BIR-focused studies involve haploid strains. Haploids cannot tolerate large deletions that frequently span essential genes. Therefore, studies that rely on haploid systems are likely limited for the types of clones carrying genome rearrangements that can be recovered. Therefore, our diploid system has the advantage of allowing a more diverse spectrum

of genomic changes to produce viable clones. However, this diversity also makes the interpretation of results more challenging, specifically, because it revealed a large number of clones that did not fit our simplistic model that framed the rationale for our hypothesis about the differences between CRHR and BIR. In any case, we would rather have our vision be clouded by revealing the real complexity of nature, than be selectively blinded by a system that oversimplifies it. Additionally, and as general principle, we believe that the use of diploid cells as a model are preferable over haploids because they allow for a more relevant and applicable translational study of the stability of the diploid human genome.

Second, extensive studies have been conducted involving experimental strains that are constructed such that DNA double-strand breakage is initiated by inducible, site-specific endonucleases (*e.g.* galactose-inducible HO-endonuclease or I-*Sce*I) (Deem et al. 2011a; Donnianni and Symington 2013; Anand et al. 2014). These are robust enzymes that cut chromosomes almost simultaneously in a population of unsynchronized cells, regardless of where the cell is at in the cell cycle; some are in G1 and have not replicated DNA while others are in late S or G2 phase and have already replicated their DNA. A typical experiment investigating BIR isolates the BIR pathway from other DNA repair mechanisms. These studies start with a chromosome that has a recombination substrate, a cut site for a site-specific recombinase, and on the other side of where the cut will occur there is a short DNA sequence with no homology anywhere in the genome attached to a telomere. Therefore, these qualities make it a totally dispensable piece of DNA. Once the cut is made, it cannot recombine because it has no homology elsewhere. If that segment gets lost and degraded it is of no consequence to the cell because it does not contain any genes required for viability. As a result, when the break occurs, only one side can engage in repair, and is thus referred to as a "single-ended break." The cut chromosome resulting

in a "single-ended break" must be repaired by BIR using another chromosome as a template in order to survive. Without a two-sided break, the cells cannot engage the CRHR pathway to repair the lesion. Therefore, the cells must do BIR or they will die because the broken chromosome will trigger permanent cell cycle checkpoint arrest and the cell will not divide – there is no other option.

On account of this experimental design, these types of studies ultimately measure survival percentage. Typically, a culture is placed in galactose to induce the HO-endonuclease, and since the enzyme is extremely efficient, meaning every chromosome gets broken, the amount of recovery is dependent on the efficiency of repair by BIR. Thus, percent of cell survival is measured as the endpoint and is displayed as recombination *frequency* since these studies are not run during the growth of a population of cells. In experiments of this type, they cannot apply the recombination rate calculation because all of their mutations occur instantly, at one cell division, that is, the cell division the cells were at the time the HO was expressed. Most often the cells are placed in solution containing the galactose (not growth medium) so that for a short period of time cells are only repairing their DNA. They are then plated to continue growing. For any recombinant that grows, it is known when that breakage and repair event occurred. It is simply a matter of how many cells there were and out of those how many got the mutation. The frequencies obtained are on the 9×10^{-1} range, meaning a wild type strain will do BIR about 90% of the time. These are very common events that have been experimentally-cornered to undergo BIR. Often times when these cells are plated they still have a broken chromosome and BIR occurs post-plating. Additionally, in these studies, survival measurement typically shows that the wild type rate of recombination is higher than of *pol32*Δ strains. If the pol32 protein is absent, more cells will die overall because of the inability to do BIR. Furthermore, the endonucleases used for this kind of study are so efficient that often times they cut both sister chromatids. So, with this induced

recombination system the cells do not have the option to repair from the sister chromatid because it may also be broken. If a sister chromatid was cut by HO-endonuclease and the other sister chromatid was not cut, if repair occurs using the information from the unbroken sister restores the original configuration this will be a futile repair because the enzyme will simply cut it again. By definition, this is irreparable damage due to this cycle of cutting and re-cutting until the repair product no longer contains the HO cut site. Therefore, some of the survivors recovered by this means could have been generated by a situation where the process of non-homologous end-joining removed a base or two and destroyed the HO cut site.

To interrogate the basal level of contribution of BIR to the overall repair processes available it was imperative that the DSBs in our experimental system be un-induced or naturally occurring and thus arise spontaneously over a period of time. It is likely that the DSBs would occur in a region where there is homology on both sides, not limiting the cells, but rather, allowing them the opportunity to do CRHR. Furthermore, "double-ended breaks" presented a more physiologically realistic repair situation where non-lethal recombination allowed for chromosome breaks to be "healed." There is no requirement for BIR in our assay; we simply observe what happens naturally to see how important BIR anyway is in the process of repair.

Finally, by providing an undefined period of time for DSBs to occur and be repaired by the cells' method of choice the recombination *rate* was determined. This type of calculation incorporates how many times a single cell had to divide to generate that population and factors in time in the form of cellular generations, while frequency measurements do not include this. In our experiment, the recombination events are dispersed, occurring over time during the growth of that culture and they can occur at any time, making it truly a fluctuation determination, without issues about survival.

References

Bell S, Putnam CD, Kolodner RD. 2014. Template homology determines the genetics and mechanisms of gross chromosomal rearrangements in *S. cerevisiae*. *The FASEB Journal* **28**.

Carvalho CM, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nature reviews Genetics* **17**: 224-238.

Conover H, Arugeso J, Kunkel T. 2015. Stimulation of Chromosomal Rearrangements by Ribonucleotides. *Genetics*: 951-961.

Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, Mathur R, Chabes A, Malkova A. 2011. Break-induced replication is highly inaccurate. *PLoS biology* **9**: e1000594.

Gu W, Zhang F, Lupski JR. 2008. Mechanisms for human genomic rearrangements. *PathoGenetics* **1**: 4.

Jackson SP. 2002. Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**: 687-696.

Llorente B, Smith CE, Symington LS. 2008. Break-induced replication: what is it and what is it for? *Cell cycle* **7**: 859-864.

O'Driscoll M, Jeggo PA. 2006. The role of double-strand break repair - insights from human genetics. *Nature reviews Genetics* **7**: 45-54.

Onishi-Seebacher M, Korbel J. 2011. Challenges in studying genomic structural variant formation mechanisms: The short-read dilemma and beyond. *BioEssays* **33**: 840–850.

Payen C, Koszul R, Dujon B, Fischer G. 2008. Segmental Duplications Arise from Pol32-Dependent Repair of Broken Forks through Two Alternative Replication-Based Mechanisms. *Plos Genet* **4**.

Putnam CD, Hayes TK, Kolodner RD. 2009. Specific pathways prevent duplication-mediated genome rearrangements. *Nature* **460**: 984-989.

Saini N, Ramakrishnan S, Elango R, Ayyar S, Zhang Y, Deem A, Ira G, Haber JE, Lobachev KS, Malkova A. 2013. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502**: 389-392.

Sakofsky CJ, Ayyar S, Malkova A. 2012. Break-Induced Replication and Genome Stability. *Biomolecules* **2**: 483-504.

Startek M, Szafranski P, Gambin T, Campbell IM, Hixson P, Shaw CA, Stankiewicz P, Gambin A. 2015. Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination. *Nucleic Acids Res* **43**: 2188-2198.

Verma P, Greenberg RA. 2016. Noncanonical views of homology-directed DNA repair. *Genes & development* **30**: 1138-1154.

Zhang HS, Zeidler AFB, Song W, Puccia CM, Malc E, Greenwell PW, Mieczkowski PA, Petes TD, Argueso JL. 2013. Gene Copy-Number Variation in Haploid and Diploid Strains of the Yeast Saccharomyces cerevisiae. *Genetics* **193**: 785-801.

Chapter Two

Characterization of Chromosomal Rearrangements Resulting from Double-Strand Breaks at G-Quadruplex (G4) DNA

Summary

G-quadruplexes (G4 DNA) are secondary nucleic acid structures formed by guanine-rich DNA sequences that are associated via hydrogen bonding to form stable, guanine-tetrad stacks (Paeschke et al. 2013). The topological characteristics introduced into the double-helix by this four-stranded structure can create physical barriers for the biological processes of transcription and replication, which in turn may lead to the formation of DNA double strand breaks (DSBs) (Murat and Balasubramanian 2014). Therefore, unresolved G4 DNA can result in compromised genome stability (Lemmens et al. 2015b). Here we describe the characterization of spontaneous gross chromosomal rearrangements (GCRs) initiated at G-quadruplex DNA motifs in *top1*Δ *Saccharomyces cerevisiae* strains (Putnam et al. 2009). The results indicated that repair broken chromosomes through the break-induced replication (BIR) mechanism was an important contributor to the formation of GCRs in the clones examined.

This chapter is an adaptation of previously published material, emphasizing the analysis of chromosomal rearrangements derived from G4 DNA instability. The complete published article is reproduced as Supplemental Material to this thesis.

Reference for the full article:

Yadav P, **Harcy V**, Argueso JL, Dominska M, Jinks-Robertson S, Kim N. 2014. Topoisomerase I Plays a Critical Role in Suppressing Genome Instability at a Highly Transcribed G-Quadruplex-Forming Sequence. PLoS Genetics **10**:e1004839.

Introduction

DNA double-strand breaks (DSBs) are a severe form of DNA damage that can result from a variety of insults and actions including, but not limited to, exogenous chemical agents, by-products of cellular metabolism (such as reactive oxygen species), and also by intrinsic DNA structures such as inverted repeats (Narayanan et al. 2006; VanHulle et al. 2007), trinucleotide repeats (Kim et al. 2008) G-quadruplex DNA structures (Lemmens et al. 2015a). DSBs are particularly dangerous to cells because of consequences from non-repair such as cell cycle arrest and/or activation of the apoptotic cascade, and also from improper repair which can result in chromosomal aberrations (Jackson 2002). A variety of genome maintenance mechanisms exist to specifically repair double-strand breaks. These mechanisms include non-homologous end-joining (NHEJ) where the break ends are directly ligated, homologous recombination (HR) which requires a homologous sequence to guide repair, and break-induced replication (BIR), a variant of HR, involving repair of only one side of the break. Despite these DNA repair pathways being in place, chromosome structure can be altered in a number of different ways that may lead to changes in gene copy number. Along with point mutations, mitotic gross chromosomal rearrangements (GCRs) resulting from inappropriate DSB repair are well recognized as a driver of cancer development (*e.g.* gene amplification leading to oncogene activation and gene deletion leading to tumor suppressor loss) (Yu et al. 2000; Aplan 2006; Marshall et al. 2008).

Importantly, G-quadruplex (G4) DNA has been implicated as a structural catalyst for DSBs and is thus associated with genome instability. It has been estimated that there are more than 375,000 potential G4-forming motifs in the human genome and more than 1,400 that occur in the yeast genome (Bochman et al. 2012). G4 DNA is a non-B form, four-stranded, secondary DNA structure composed of guanine-rich nucleic acid sequences associated in a planar orientation. G4

DNA forms primarily during transcription at the single-stranded non-template strand. If G4 DNA remains unresolved, it can become a structural barrier to transcription and replication *in vitro* and likely leads to replication fork stalling and ultimately DNA breakage *in vivo* (Tornaletti et al. 2008). G4 DNA is found highly concentrated at telomeres, ribosomal DNA and immunoglobulin (Ig) heavy-chain switch regions (Dunnick et al. 1993; Willis and Dyer 2000; Maizels 2006). Ig-isotype switching is one such example where recombination would actually be desirable to generate different classes of antibodies. Not surprisingly, abundant chromosomal translocations involving G-rich Ig switch regions have been identified. G4 motifs have also been identified frequently at breakpoints involved in translocations in cancer (Katapadi et al. 2012).

In the present study, *Saccharomyces cerevisiae* served as the model organism for investigation of GCRs triggered by G4 DNA. When the gene encoding Topoisomerase I (*TOP1*) was knocked out, an introduced G4 DNA-forming sequence became the site of mitotic recombination at a frequency higher than normal. G4 DNA motifs were incorporated into the yeast genome on Chromosome 5 (Chr05) using one of two cassettes, each cassette descriptive of the orientation of the guanine-rich sequence relative the direction of transcription. These constructs were named either -*GBTM* or -*GTOP*; (Figure 2.1A). The integrated G4 DNA-forming sequences were derived from a 760 bp G-quadruplex motif from a mouse immunoglobulin switch Mu (Sμ) region (Kim and Jinks-Robertson 2011). On the right side of Figure 2.1A, -*GBTM*: the guanine-rich sequence is in the transcribed bottom strand and is bound in a DNA-RNA hybrid (the red line indicates nascent RNA).

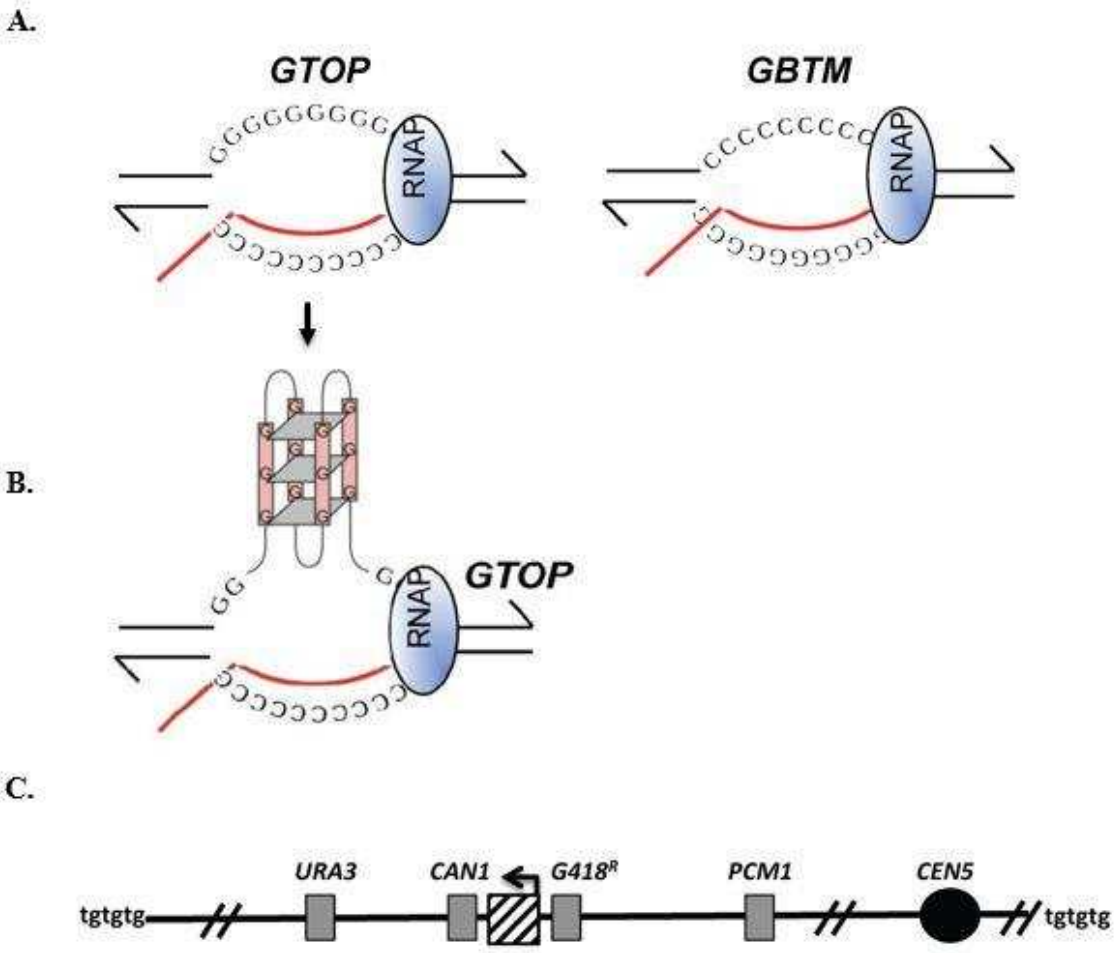


Figure 2.1. G-quadruplex DNA cassette orientations and chromosome map
A. The transcription orientations of the Sμ-containing cassettes.
Guanine-runs are on the non-transcribed strand in a single-stranded state in *pTET-lys2-GTOP* cassette and on the transcribed strand annealed to the nascent RNA (red line) in the *pTET-lys2–GBTM* cassette. RNA polymerase holoenzyme is indicated by the blue oval.
B. Presumed mechanism of G4 DNA folding.
-*GTOP* cassette with depicted formation of the G4 secondary DNA structure from the exposed guanine-rich DNA sequence in the transcription bubble. The intervening DNA sequences are shown as a black line. The grey squares between the G-tracts represent planar hydrogen bonds. –*GBTM* cassette does not form G4 structure as the guanine-run is bound as a DNA-RNA hybrid. (Image adapted from (Capra et al. 2010))
C. Map of Chromosome 5 for Gross Chromosome Rearrangement (GCR) assay.
A schematic representation of the relevant features of Chr05. The location of *pTET-lys2-GTOP* or *–GBTM* cassette is indicated by the hashed box with the arrow above indicating the direction of transcription. The distances (in kb) are approximate and not to scale. Telomeres are represented as "tgtgtg".

On the left side of Figure 2.1A, the string of guanine residues are along the non-transcribed top strand, and have the ability to fold on themselves to form a G4 DNA structure as shown in Figure 2.1B. After the G4 DNA structure forms, transcription proceeds normally beyond it. However, this region becomes unstable when a replication fork reaches the G4 DNA structure impediment, leading to replication fork stalling and resulting in double-strand breakage. Prior to the publication of our work (Yadav et al. 2014), there was no experimental *in vivo* demonstration of genome instability effects at G4 sites.

We sought to understand the role that G4 DNA plays in spontaneous chromosomal rearrangements. The study included use of the well-established Gross Chromosomal Rearrangement (GCR) assay in which *URA3* and *CAN1*, counter-selectable markers, confer sensitivity to 5-Fluorootic Acid (5-FOA) and canavanine, respectively (Figure 2.1C) (Chen and Kolodner 1999). Selection for double-drug resistance primarily yields clones carrying deletions of the non-essential chromosomal segment containing the two markers. The likelihood for point mutations to inactivate these two genes independently is extremely low ($\sim 10^{-12}$), well below the detection limit ($\sim 10^{-10}$) of the selection for 5-FOA and canavanine double resistance. Additional studies described in detail in attached Supplementary Material, demonstrate that the increase in chromosomal rearrangements induced by G4 DNA depended on: Topoisomerase I function, transcription through the G4 substrates, and G4 orientation. This was shown by inserting the *GBTM/GTOP* cassette (hashed box in Figure 2.1C) in opposite orientations relative to the direction of transcription (arrow in Figure 2.1C): *TET-OFF pTET-lys2-GTOP* (G4 in non-template strand) or *TET-OFF pTET-lys2-GBTM* (G4 in template strand). Furthermore, doxycycline was used to regulate transcriptional activity using a *TET* on/off gene expression system (*i.e.* when doxycycline was present, transcription was repressed). Figure 2.2 depicts the measured rate of occurrence of

double-drug resistant colonies based on the orientation of the G4 DNA forming sequence for wild type and in *top1*Δ mutants. The rate of GCR was significantly different between -*GTOP* and -*GBTM* in wild type, but was ~30 fold higher in -*GTOP top1*Δ mutants.

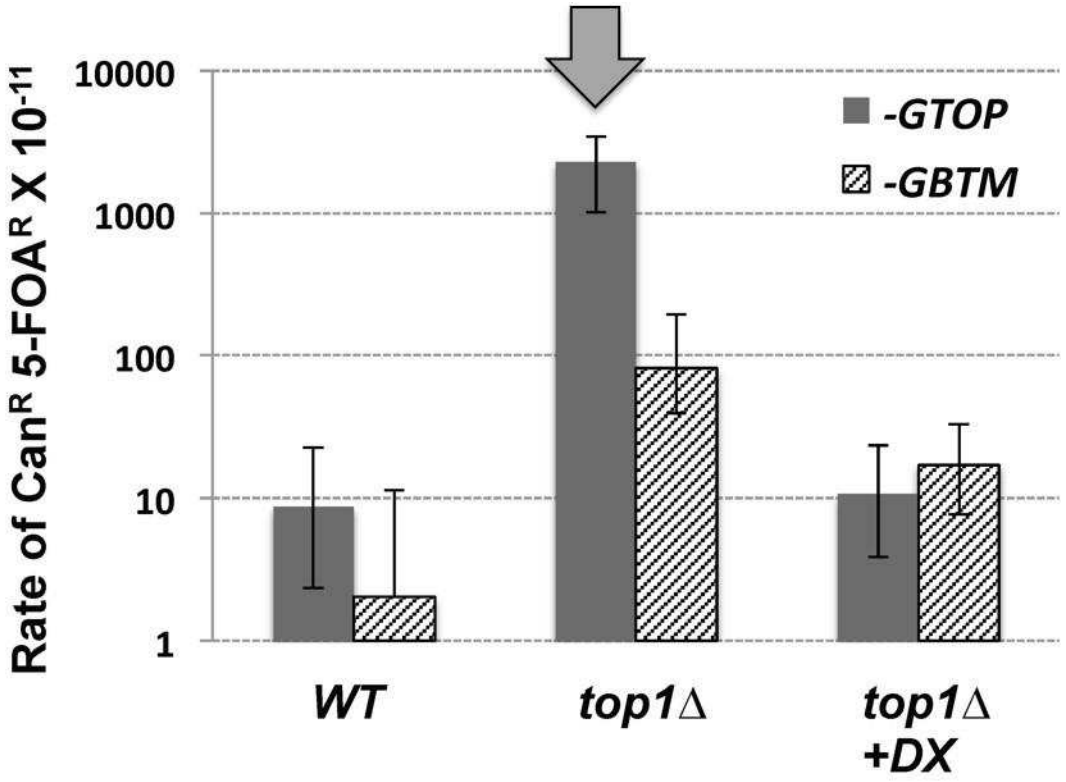


Figure 2.2. Gross Chromosomal Rearrangement (GCR) rates.
GCRs occurring at Chr05 containing the *pTET-lys2-GTOP* (gray bar) or *–GBTM* (hashed bar) cassette. +DX; Doxycycline was added to the final concentration of 2 μg/ml to repress the transcription from *pTET*. 95% confidence intervals are indicated by the error bars. The block arrow indicates the experimental conditions from which the clones analyzed were isolated from: *top1Δ* mutants with active transcription through the -*GTOP* cassette.

This differential rate of GCRs was eliminated by repressing transcription through the addition of doxycycline (right bar pair). Our specific contribution to the study was to qualitatively evaluate the nature of the abundant chromosomal rearrangements isolated from *top1*Δ mutants with the inserted -*GTOP* cassette and active transcription (Block arrow in Figure 2.2). This characterization enabled us to infer the DSB repair pathways likely responsible for the genome instability associated with G4-DNA. To analyze the recombination mutants we employed three methods, used in conjunction, to thoroughly characterize the GCRs: Pulsed-field gel electrophoresis (PFGE), microarray-based comparative genomic hybridization (array-CGH), and Sanger sequencing of chromosomal breakpoints from PCR products or DNA sequences captured via plasmid rescue.

Results & Discussion

In order to characterize the chromosomal rearrangements associated with G4 DNA, we first used pulsed-field gel electrophoresis (PFGE), a technique for size-based separation of large DNA molecules (up to 6 Mb) achieved by applying an electric field that periodically changed direction. This method allowed for separation of whole, intact yeast chromosomes and was used to identify rearrangements based on the migration of new chromosomal bands. In Figure 2.3, the karyotype of the parental strain is in the left-most lane showing the migration pattern of normal-length chromosomes. There were a total of 27 double-drug-resistant clones which had undergone GCRs. Those with similar size rearrangements were then segregated into three different classes: Class I, Class II, and Class III (a & b). Representative clones carrying these classes of rearrangements are shown in Figure 2.3.

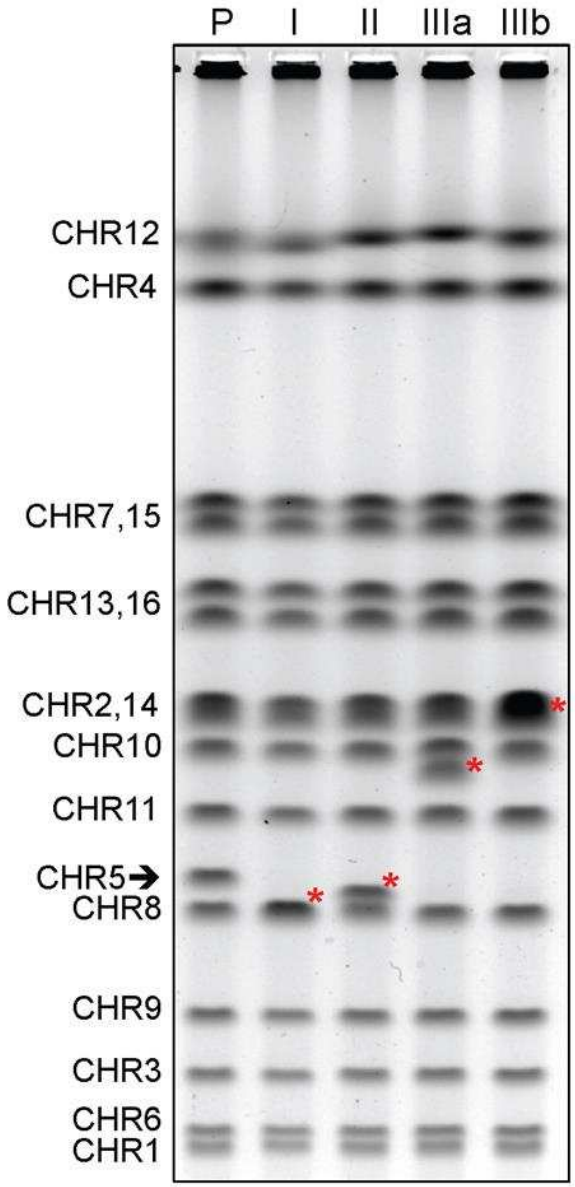


Figure 2.3. Classes of GCRs associated with the highly transcribed G4 DNA motif.
PFGE analysis of different classes of GCR events. The full-length chromosomes of representative 5-FOA^{R} Can^{R} clone derived from the *top1Δ* strain containing the *pTETlys2-GTOP* cassette are separated out by size. Classes I, II, IIIa and IIIb are described in the text. The chromosomes in the parental strain (P) are indicated to the left and the arrow indicates the normal (unrearranged) Chr05. The locations of novel chromosomal bands in each sample are indicated by red asterisks.

Since the *URA3, CAN1* genes and the G4 DNA-forming sequence were all on Chr05 the PFGE location of this chromosome was of particular interest. The red asterisks show the new sizes (larger or smaller) of Chr05 based on its band location.

In addition to PFGE, microarray-based comparative genomic hybridization (array-CGH) was performed in order to analyze copy number differences between the DNA from the GCR-containing clones compared to that of the parental strain. The GCR DNA and reference DNA were each labeled with different fluorophores (red and green) and competitively hybridized to a glass slide coated with approximately 15,000 interrogating oligonucleotide probes spanning the yeast reference genome sequence (S288c; Saccharomyces Genome Database (Engel et al. 2014). A laser scanner was used to detect fluorescent signals and computational analysis provided a plot of the gains and losses of gene copy number. The summarized data plots for all classes of chromosomal rearrangement clones is summarized in Figure 2.4 where red indicates a copy number loss and blue represents a copy number gain. Along the bottom of the diagram is a schematic of the chromosomal map of relevant landmarks on Chr05 including the *GTOP*/G4 DNA insertion site and Ty/delta elements, regions of yeast repetitive DNA sequence where it is relatively common for DNA to recombine. Side-by-side comparisons of PFGE and array-CGH data of Chr05 for each GCR class are shown in Figures 2.5A-D with black dots corresponding to individual probes. The X-axis provides the chromosomal coordinates or loci for each probe and the Y-axis is the Log_2 (red/green signal) for each probe.

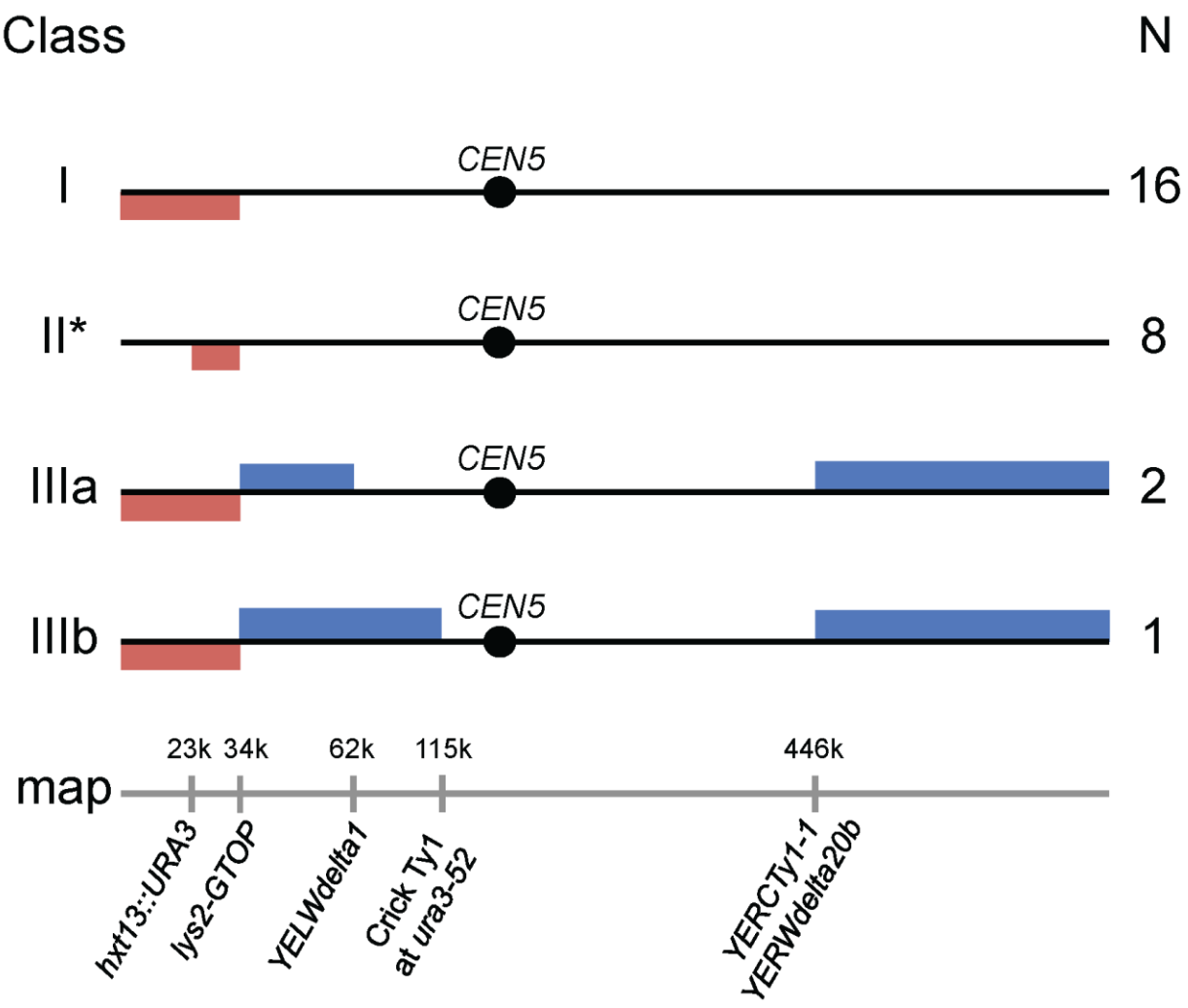


Figure 2.4. Summary of the array-CGH analyses.
The red and blue bars indicate the loss and the gain of the copy number of the corresponding regions of Chr05, respectively. The coordinates are approximate and correspond to the S288c reference genome sequence according to Saccharomyces Genome Database. The relevant genomic features are indicated at the bottom of the figure. The locations of relevant solo LTRs and full length Ty retrotransposons are indicated. At coordinate ~115,000, the *ura3-52* allele with a Crick-oriented insertion of a Ty1 element is present in this strain background.

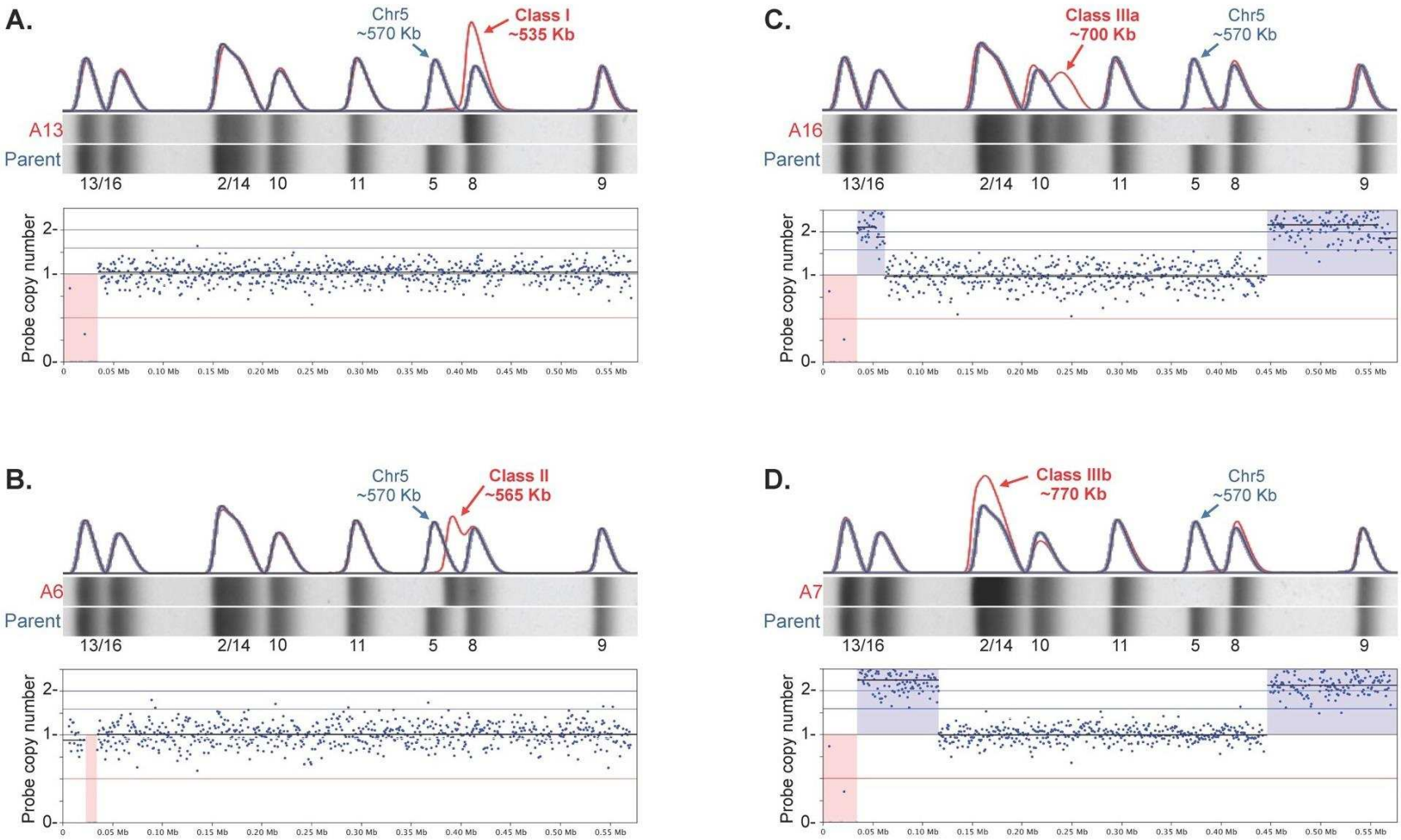


Figure 2.5
Full caption follows.

Figure 2.5. Molecular karyotype analysis of Chr5 GCRs.
A. GCR class I, clone A13; **B.** GCR class II, clone A6; **C.** GCR class IIIa, clone A8 and A16; and **D.** GCR class IIIb, clone A7. Quantitative analysis of the PFGE and the detailed Chr05 gene dosage plots for the GCR classes are shown. In all panels, the top part corresponds to sections of the same PFGE gel shown in Fig. 2.3, cropped between the Chr13/06 and Chr09 regions. Directly above the PFGE images are the superimposed plots of the quantitative trace analysis of the PFGE lanes for the parental strain (dark blue trace) and for the respective GCR clone (red trace). The image pixel intensity traces (vertical axis) for the GCR PFGEs were normalized relative to the trace of the parental strain. The parental and GCR traces closely overlapped for all chromosomes other than Chr05 and the rearranged Chr05 in the various GCR classes. The approximate size of Chr05 and the GCRs are indicated. The lower part of each panel shows the array-CGH copy number plots for the probes (blue dots) from Chr05 in the respective GCR clones. The vertical axis corresponds to the Log2(GCR[Cy5]/Parent[Cy3]) signal and the corresponding positions for 0, 1, and 2 copies of genomic material in the respective GCR clones. The horizontal axis corresponds to the physical position of each probe on Chr05. Regions where full deletions were detected are shaded in red, and regions where duplications were detected are shaded in light blue.

In the most frequent GCR class (16 of 27 clones), Class I, the data from the PFGE in combination with the densitometry measurements (parental in blue and clone in red) showed a smaller (approximately 535 Kb Chr05) than normal parental ~570 Kb Chr05 seen in Figure 2.5A. Along with the array-CGH data, it was apparent there was a terminal deletion occurring distal to the G4 DNA site. Based on this data, the breakpoint was amplified via PCR with degenerate telomere primers and the PCR product was then Sanger sequenced. Analysis indicated a rearrangement consistent with a terminal deletion and *de novo* telomere addition at the switch Mu (Sμ) G4 DNA motif. This result was consistent with the most common class of GCRs observed when a homologous recombination substrate is not present between the *URA3*/*CAN1* and the first essential gene in the chromosome (*PCM1*) (Chen and Kolodner 1999). In the absence of a substrate for homology-directed repair, alternative pathways including *de novo* telomere addition prevail.

In Class II (8 of 27 clones), the PFGE also showed a smaller than normal Chr05 at 565 Kb instead of ~570 Kb, though not as small as that of Class I. In this case, the array-CGH indicated the *URA3* and *CAN1* genes were lost but not the whole end of the chromosome. Further analysis by PCR amplification across the breakpoint and sequencing revealed that it was a recurrent interstitial deletion mediated by 21 bp direct repeats. The repeats were vector sequences that were unintentionally introduced during strain construction. This result was consistent with a single-strand annealing mechanism that is quite frequently observed in yeast when the appropriate sequence substrates are available (Symington et al. 2014).

The Class III clones were more structurally complex, and their characterization was much more challenging. In the 2 of 27 clones classified as Class IIIa, the signature terminal deletion of the left arm of the chromosome occurs distal to the G4 DNA site. There is amplification of the region near the G4 DNA sequence and another seemingly unrelated amplification of the far end of the right arm of Chr05. Due to the copy number gains from these two regions of duplication, the PFGE showed a much larger (~700 Kb) than the normal Chr05. The array-CGH of Class IIIb (1 of 27 clones) was very similar to that of IIIa but the amplification region near the G4 DNA site was larger, resulting in ~770 Kb Chr05. This configuration led us to propose mechanisms to explain the occurrence of these Class III outcomes with two non-exclusive potential pathways ending with the same final structural result (Figure 2.6). The proposed steps in the rearrangements are as follows: A DSB occurs at the G4 site. On the right side of the DSB, 5’ to 3’ DNA resection occurred (not shown) where one strand is trimmed back exposing a single stranded DNA 3’ end. The left side of the break would not be repaired and would be lost from the genome, explaining the deletions seen in array-CGH. This 3’ end of the right side loops back and finds a region of minimal homology and initiates break-induced replication (BIR).

At this point, two alternative pathways may have taken place. Following the pathway on the left of Figure 2.6, BIR continues down the left chromosome arm until it reaches a Ty element/long terminal repeat (LTR), a repetitive region of DNA. Then template switching occurs, moving to another Ty element on the right arm where BIR resumes. The chromosome is now considered "healed" as it regains a second telomere. The final result is an inverted repeat GCR iso-chromosome structure which fits the PFGE and array-CGH data. Based on the lettered segments of the chromosome diagram in Figure 2.6, the copy number outcome results in the absence of segment "A" since it was originally lost when the terminal deletion occurred. There is a duplication of "B" and "C" present on the left arm. There is a single copy of "D" and "E", but also a duplication of "F" which shows as the amplification of the right arm of the chromosome.

In the alternative model on the right half of Figure 2.6, the first two steps are the same. However, in this pathway BIR continues uninterrupted through the centromere all the way to the right telomere resulting in a dicentric chromosome. Since dicentric chromosomes inherently are unstable, the duplicated centromere would have to be looped out by homologous recombination between Ty/LTR repeats (breakage-fusion-bridge) resulting in a final GCR inverted repeat structure, indistinguishable from the one predicted by the pathway shown on the left (template switching). We prefer the former model to the latter. The reasons supporting our inclination are that (i) template switching is commonly associated observed in BIR, (ii) the centromere is generally considered a strong barrier to BIR, and (iii) studies have shown limited synthesis via BIR (Smith et al. 2007; Donnianni and Symington 2013; Stafa et al. 2014). Our own subsequent work (described in Chapter 3) provided further evidence that the replication fork associated with BIR is not capable of efficiently carrying out DNA synthesis for more than about 200 consecutive kilobases.

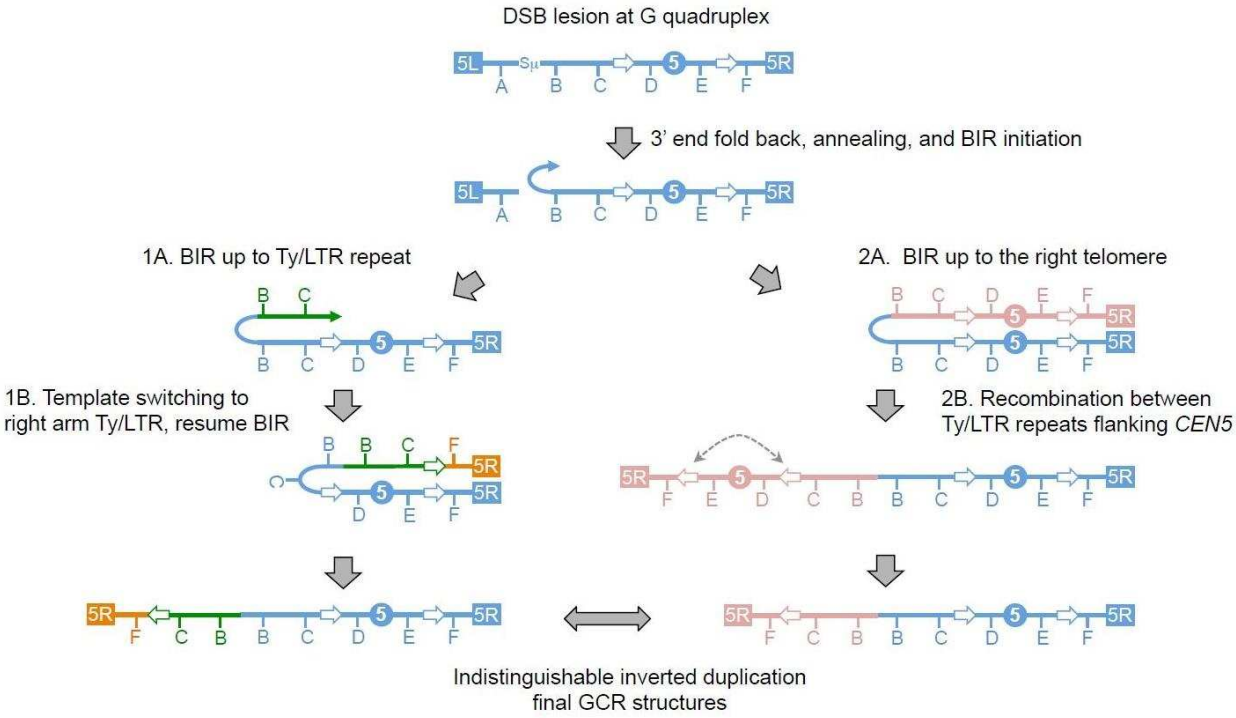


Figure 2.6. Models for the generation of the recovered Class III GCRs.
The inverted duplication rearrangements initiated by a double-strand break lesion at the Sμ sequence. Unique regions in Chr5 are labeled with letters A through F. The Chr5 fragment to the left of Sμ containing *TEL05L*, *URA3*, and *CAN1* is lost (segment A), while the resected 3' end of the Chr5 fragment to the right of Sμ folds back on itself and initiates BIR at the B segment. At this point, two possible models are shown to explain the structure of the recovered clones.

Model 1: Template switching.
1A. BIR proceeds from left to right through the B and C segments (green line) until reaching a dispersed Ty or LTR repeat (open white arrow). 1B. At this point the BIR fork collapses, and re-anneals at a homologous Ty or LTR on the right arm, then resumes extension copying the F segment (orange line) until reaching the right telomere. The resulting GCR product has a deletion of the A segment, an inverted duplication of B and C, single copy of D and E (including the centromere), and a duplication of F.

Model 2: Dicentric formation and secondary rearrangement.
2A. BIR copies an uninterrupted template through segments B, C, and D, then across the centromere, then through segments E and F reaching the right telomere (pink line). 2B. Since the resulting GCR is dicentric, it inevitably accumulates a new DNA break that leads to a secondary rearrangement between directly oriented Ty or LTR repeats flanking one of the two centromeres (dashed double-ended arrow). The resulting secondary GCR becomes monocentric and is recovered as a viable Class III clone.

Since our PFGE and array-CGH data (Figure 2.5) suggested the possible structures of Class III rearrangements (terminal deletion, inverted duplication and terminal duplication), we set out to characterize the specific breakpoint sites to obtain single nucleotide resolution of the junctions, to validate our prediction and to infer repair mechanisms. If the final GCR structure were an inverted duplication, then the break at the G4 site would cause loss of the end of the left chromosome arm, and if it followed the proposed mechanism, then it would fold back and copy the adjacent region causing a duplication. Numerous PCR attempts across the G4-proximal breakpoint either failed or did not provide specific amplification products. This approach for direct amplification of the breakpoint junctions was likely hampered due to formation of a hairpin structure which is refractory to PCR. Alternatively, we used a method called plasmid rescue to physically capture the potential inverted duplications point near a known DNA sequence, thus bypassing PCR amplification. The plasmid rescue technique involved integration, via homologous recombination, of a linearized bacterial plasmid into yeast genomic DNA at a sequence flanking the target site. Following PCR confirmation of integration at the desired locus, the genomic DNA from transformed cells was isolated and restriction digested. Genomic DNA fragments containing plasmid sequences were re-circularized and cloned in *E. coli*. The plasmid clones carrying the breakpoint junctions were then Sanger sequenced, generating the results shown in Figure 2.7A&B, which substantiate the proposed fold-back mechanism from Figure 2.6 with actual data.

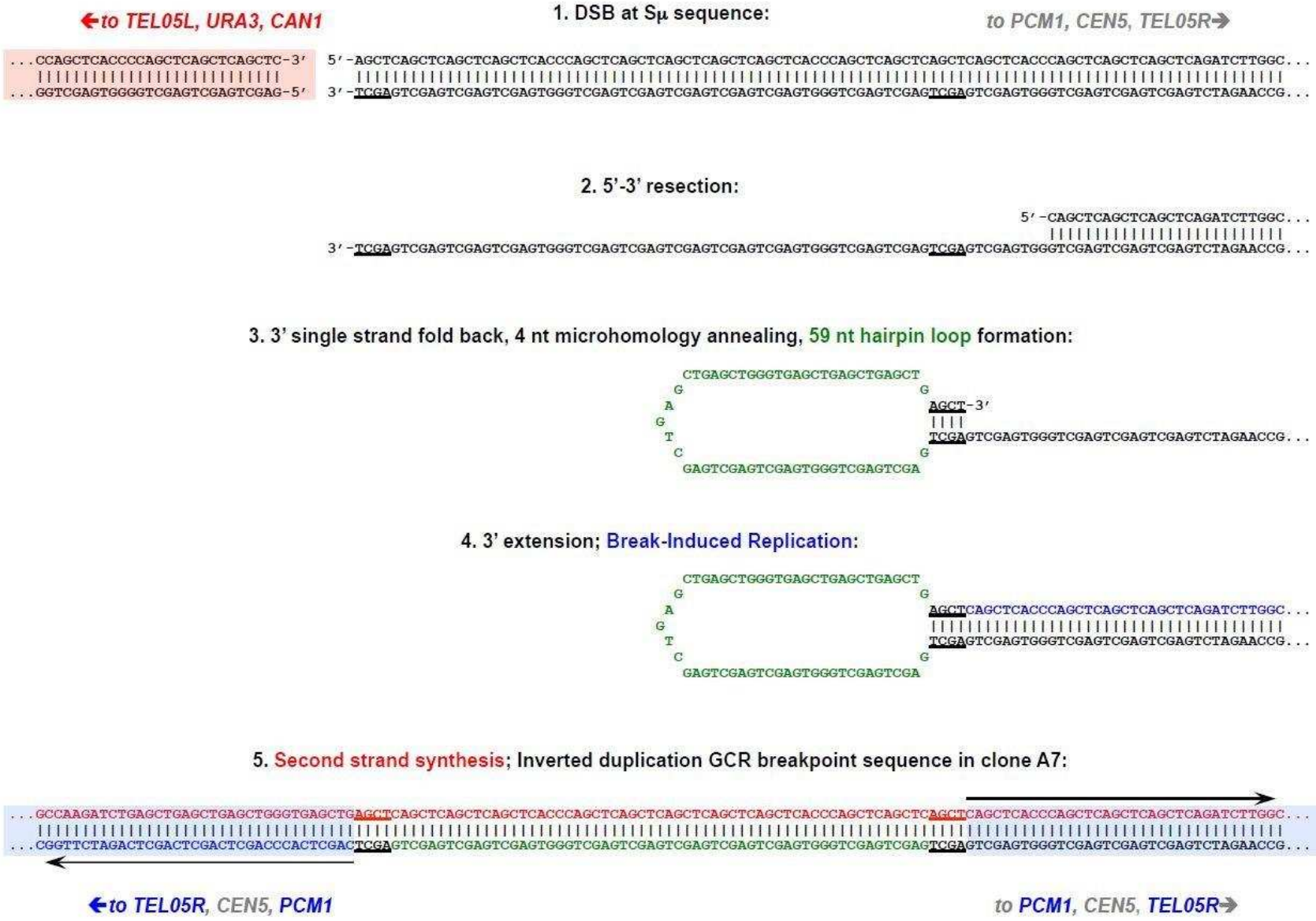


Figure 2.7A
Full caption follows

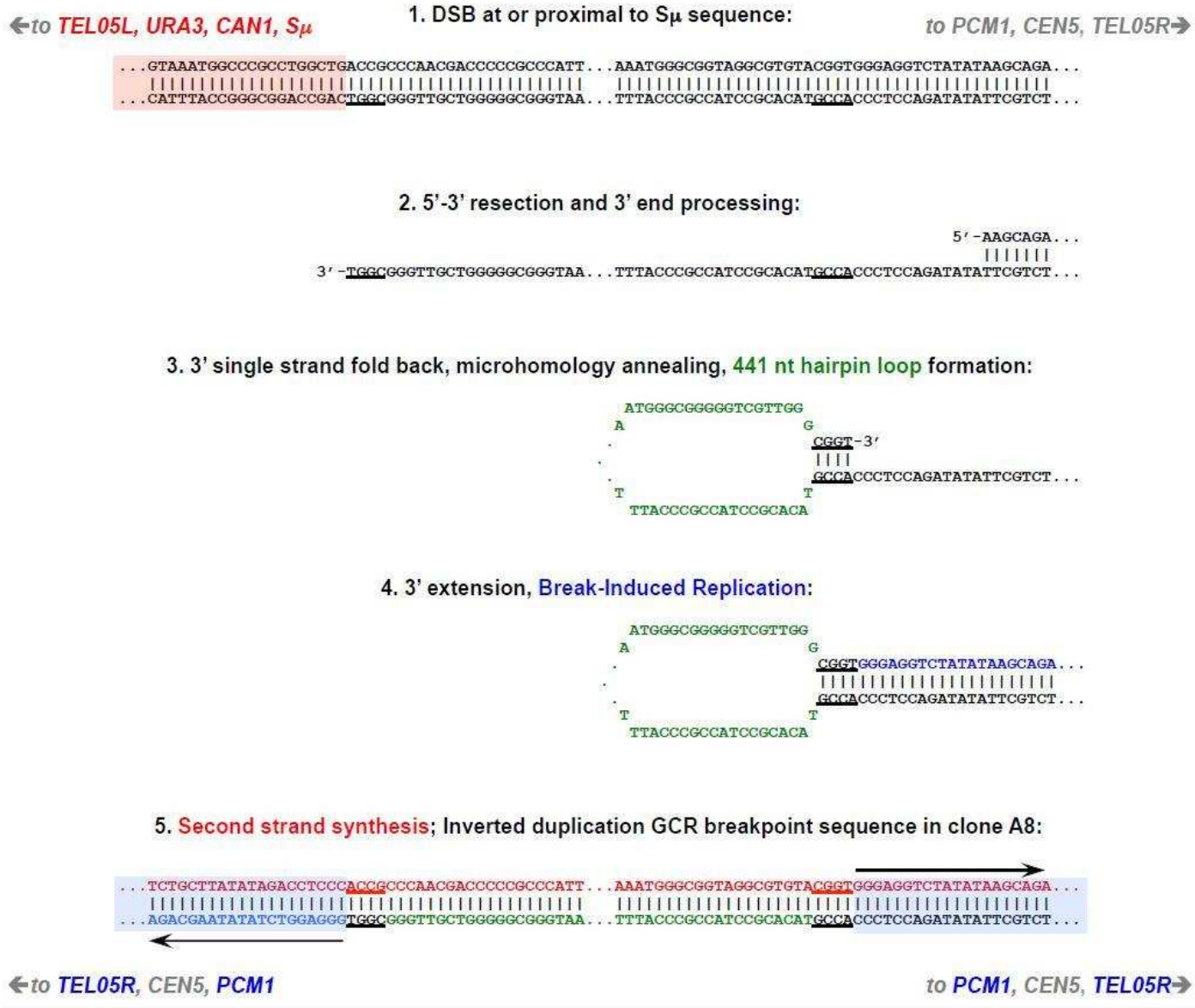


Figure 2.7B
Full caption follows

Figure 2.7A&B. Fold-back inverted duplications originating at the G4 DNA sequence. The DNA sequence determined from the inverted duplication structures in clone A7 (Figure 2.7A) and A8 (Figure 2.7B) are shown at the bottom of each figure (Step 5) and the proposed mechanism of formation is shown above them. Step 1: A DNA double strand break forms within or proximal to the Sμ sequence. Step 2: 5' to 3' resection creates a ssDNA region. Step 3: The 3' end of the ssDNA folds back onto itself forming a loop of 59 nucleotides (clone A7) or 441 nucleotides (clone A8) (nucleotides in green) and anneals to a 4 nt microhomology (underlined). Step 4: The annealed 3' end primes break-induced replication (nucleotides in blue) going toward the right arm. Step 5: After BIR, a complementary DNA strand is synthesized (nucleotides in red) to complete the inverted chromosomal rearrangement event. Regions deleted and duplicated in the resulting clone are shaded in red and blue, respectively. The arrows above and below the DNA sequence show the two sides of the inverted duplication, which flank the microhomologies and the single-copy region corresponding to the original ssDNA loop.

The results support the model where a DSB occurs at the G4 site followed by DNA resection. The exposed 3' end folds back and finds a region of minimal homology (in this case only four nucleotides), it anneals creating a 59 nucleotide loop from which BIR was initiated (Figure 2.7A). When the second strand was completed the result was an inverted duplication since the ends are identical, but inversed, and the middle sequence was unique since it was the spacer between the two repeats due to the temporary hairpin. Sequencing data from clone A8 showed a similar pattern except that a larger (441 nt) hairpin loop was formed (Figure 2.7B).

The secondary chromosomal rearrangements in Class III involving the right arm had breakpoints at Ty/LTR repetitive sequences, and were consistent with a non-allelic homologous recombination mechanism (Argueso et al. 2008). These breakpoint junctions were more amenable to PCR analysis and did not require the plasmid rescue approach. In the schematic diagram in Figure 2.8A, the top portion shows the structure of the normal parental Chr05 with the G4 DNA site. The bottom portion shows the resulting GCR in clone A16 indicating a presumed duplication of the right arm to form and iso-chromosome structure (the duplicated right arm segment is in inverted orientation).

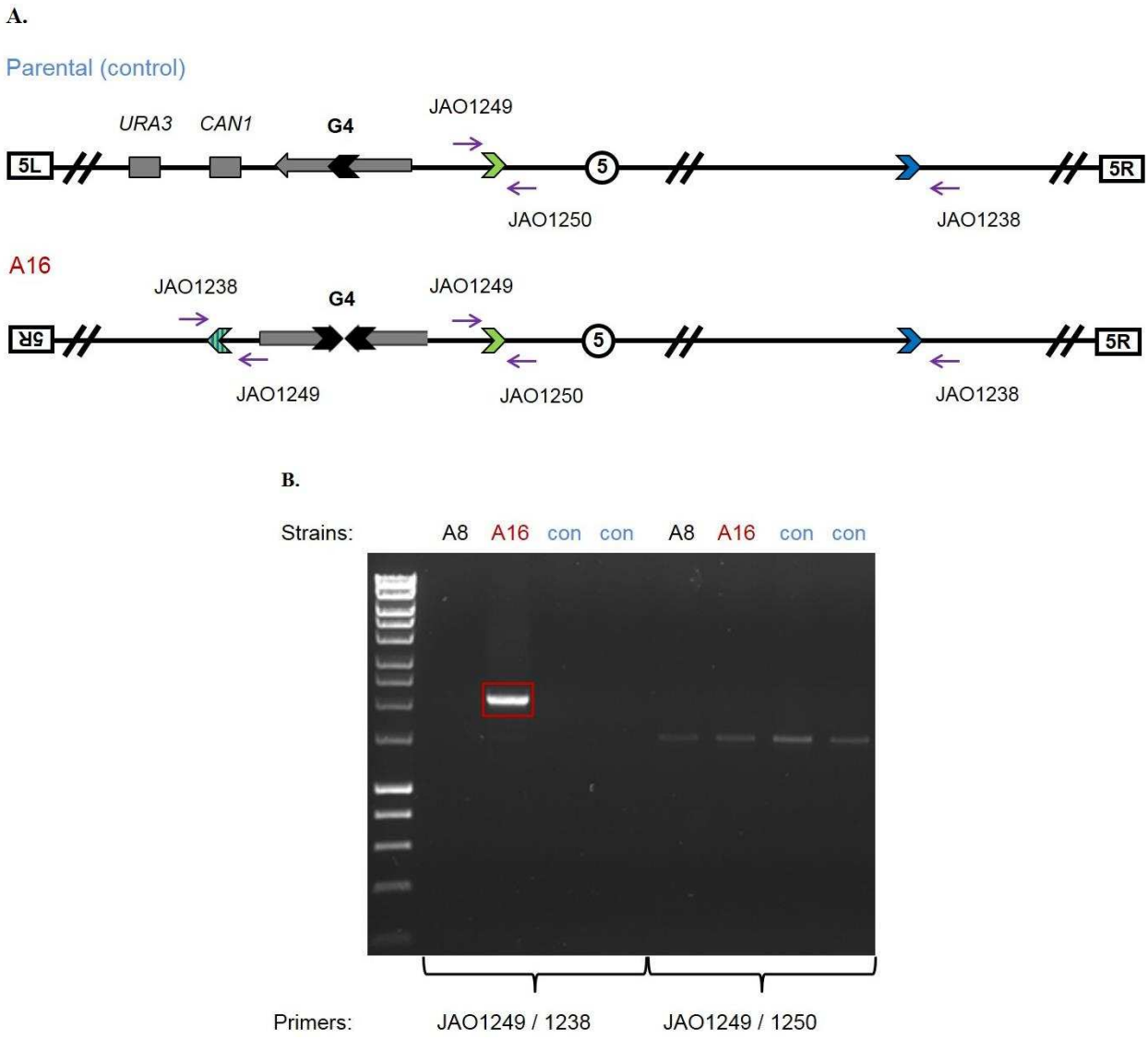


Figure 2.8. Chromosome map and breakpoint analysis for Class IIIa GCR
A. Class IIIa Chromosome 5 map.
Schematic representation of normal Chr05 in the parental strains and the predicted structure from array-CGH following rearrangement in A16. The green chevron symbols represent *YELWdelta1*, the blue chevrons represent *YERWdelta20b* and the hashed chevron depicts a hybrid resulting from a recombination event between these two repetitive DNA elements. The sequence of this recombination junction is shown in Figure 2.9. Telomeres are represented by boxes at ends of the chromosome with left and right indicated within. Primers are shown with labeled purple arrows. Chromosomal sizes not drawn to scale.
B. Results of PCR approach for breakpoint amplification and analysis.
A PCR amplicon was generated (via primers JAO1249/1238) only from Class IIIa clone A16 (red box), not from Class IIIb clone A8 or from parental strains 150-1 and 150-2 (controls). Primers (JAO1249/1250) flanking the native *YELWdelta1* amplified this repeat element in all four strains tested.

PCR was used to amplify the breakpoint region with primers JAO1249 and JAO1250. Only clone A16 showed an amplicon between primers JAO1238 and JAO1249 and since each of these primers were specific (complementary) to opposite ends of the chromosome this indicates a duplication of the right arm of Chr05 now present in place of the original left arm of Chr05 (Figure 2.8A&B). This PCR amplicon was purified and sequenced and the trace data was aligned to the reference yeast genome. The junction where the two sequencing reads overlapped implied that a recombination event occurred between two delta long terminal repeat regions with 198 base pairs of shared homology, 8 mismatches, 96.5% identity (Figure 2.9). This PCR analysis validated the GCR structure predicted by the PFGE and array-CGH data for clone A16.

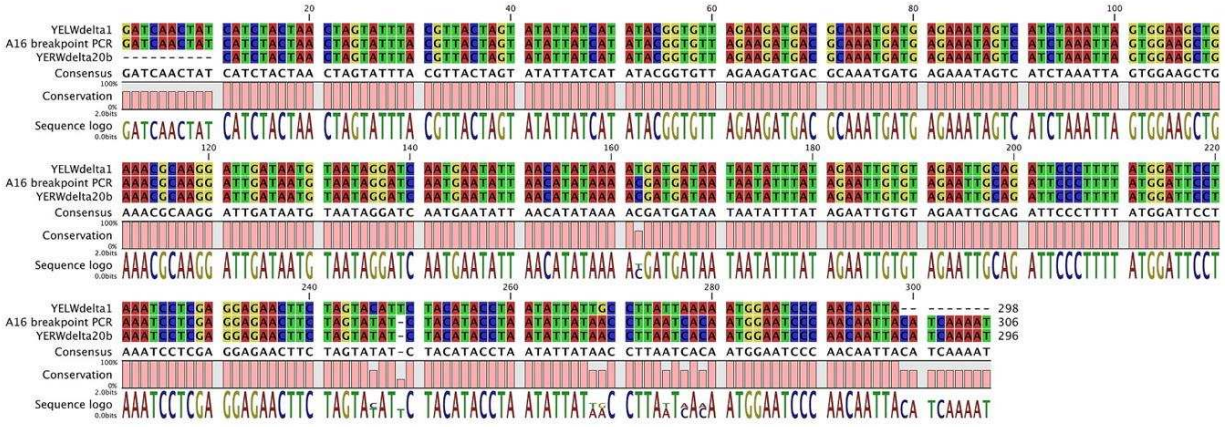


Figure 2.9. Sequencing of LTR – LTR recombination breakpoint in Class IIIa.
The trace data from clone A16 was aligned to the S288c reference genome sequence according to Saccharomyces Genome Database using CLC Genomics Workbench. The detailed breakpoint nucleotide sequencing reads overlapped indicating a recombination event between 2 deltas (repetitive sequence regions) with 198 bp homology, 8 mismatches, 96.5% identity.

In conclusion, the inverted duplications recovered from the GCR assay that we analyzed are indicative of a micro-homology mediated event involving only a few bases to prime the replication process used to repair the G4 chromosome breaks. Similar rearrangement junctions potentially formed by a fold-back mechanism have been noted previously in yeast with a hairpin loop ranging from 25 to 44 bases (Putnam *et al*. 2014). The two fold-back hairpins described in our clones were larger (59 and 441 bases). Analogous structures have also been characterized in the human genome. Hermetz *et al.* analyzed the genomic architecture of a large collection of human inverted duplications. Not surprisingly, their findings supported the existence of a similar loop-back mechanism in humans to that described in our study and earlier studies in yeast. While overall the structure was similar, the sizes of the molecules involved were much larger. They were able to identify spacers of unique sequence fold-back hairpins generated between palindromic inverted duplications (derived from the original) up to 70 Kb in length, three orders of magnitude larger than the yeast examples we characterized (Hermetz *et al*. 2014). This recent finding lends strength to the likelihood that inversion junctions created by these mechanisms are more frequent than originally anticipated by some in the human genome and they do indeed play an important role in the pathogenesis of cancer and other diseases.

Material and Methods

Yeast strains and growth conditions

Rich growth medium (1% yeast extract, 2% peptone, 2% dextrose; YPD) was used to grow clones A7, A8, and A16 containing GCRs, as well as the parental strains 150-1 and 150-2 as controls. Yeast cells were grown at 30C. DNA was isolated by potassium acetate protocol as previously described and quantified using Invitrogen's Qubit fluorometer (Ausubel et al. 1998).

Molecular karyotype analysis

The molecular nature of the chromosomal rearrangements in the clones recovered from the GCR assay was determined by performing Pulsed-Field Gel Electrophoresis and array-CGH as described previously (Argueso et al. 2008).

Plasmid rescue

The chromosomal configuration of the strains used in this experiment included the KanMX4 G418-resistance marker inserted at positions proximal to the -*GTOP* cassette (Figure 2.1C). We took advantage of this sequence as a site for integration of a plasmid sequence in the vicinity of the G4-associated inverted-repeat breakpoint in clones A7 and A8. We linearized plasmid pAG25 (Goldstein and McCusker 1999) carrying the NatMX4 Nourseothricin resistance marker with *Bgl*II to target its integration by homologous recombination to the KanMX4 site. Restriction digest of pAG25 with *Bgl*II and transformed all four yeast strains and plated on YPD. They were then replica plated onto YPD containing 0.2 mg/mL Nat plates. Candidate Nat resistant colonies were struck out to single cell clones and patched for DNA extraction. The genomic DNA from pAG25-transformed yeast was harvested and proper integration was confirmed using one chromosome-specific primer and one plasmid-specific primer was included in each primer pair (JAO979/1258 and JAO581/1005) listed in Table 2.1. Once integration at the intended site in the yeast genome was confirmed by PCR, we prepared genomic DNA, digested it with *Eco*RV, and religated to create circles containing the plasmid and the neighboring breakpoint sequence. The ligation products were transformed into *E. coli* using a heat-shock transformation protocol by Bioline. Six candidate clones were selected and then streaked to single colonies.

The single colonies were then inoculated into 5 mL LB plus 0.1 mg/mL ampicillin broth and incubated overnight at 37C on a 200 rpm shaker. Plasmids were isolated using a miniprep kit from Thermo Fisher Scientific. The rescued plasmids obtained through this procedure were analyzed by restriction digestion and Sanger sequencing to reveal the breakpoint structures shown in Figure 2.7. In some cases, the rescued regions containing the breakpoint were subcloned into pUC18 to reduce the size of the hairpin present in the plasmids and improve the efficiency of the Sanger sequencing reactions. A7 rescue clone#7 was digested with *Sac*I and *Spe*I. A7 rescue clone#3 was restriction digested with *Age*I, *Kpn*I, *Sac*I, and *Sph*I. A7 rescue clone#3, A7 rescue clone#4 and A8 rescue clone#1 were digested with *Eco*RI and *Sac*I. Restriction analyses of the rescued plasmids were consistent with inverted duplication structures.

PCR breakpoint mapping

For the breakpoints involving LTR repeats at the breakpoint, PCR was performed near *YELWdelta1* with primers JAO1249/50/51. PCR of all four strains with JAO1249/1238 (a primer specific to the right arm of the chromosome) showed amplification in only clone A16. This product was PCR purified using Thermo Fisher's purification kit and the Ty rearrangement was sequenced with primers JAO1282/83/84/85 listed in Table 2.1.

Table 2.1. Sequences of synthetic oligonucleotides used in this study

Primer ID#	Primer sequence in 5'-3' orientation
JAO581	CTT GAT GGT CGG AAG AGG CAT
JAO871	CAT CAC AAT TAG TAA TGG AAA GTG TTT GGA AAA AAA AAG AAG AT AA GGC GCG CCA GAT CTG
JAO872	TAA CAA CCA GAA ATA GGC TTT AGT TAA CTC AAT CGG TAA TTA CAT AGG CCA CTA GTG GAT
JAO979	ATCCAGTGCCTCGATGGC
JAO1005	ACA CGG AAA TGT TGA ATA CTC
JAO1009	ATA TTC ATT GAA ACT GAT TAT TCG ATT TTC TTC TTG ACC GCT CGT TTT CGA CAC TGG
JAO1238	ATTTTGGGTGATTTCACAAGG
JAO1249	TGATAAAGACAACTTACAAGTACAG
JAO1250	TGGTATGTCAATGTGAGATATCAC
JAO1251	TGAGTGTTCGTTCAGAAACCACTG
JAO1258	TTCCGAGATACGATTACTCCAG
JAO1282	CTATCAAAGAATTAGGCTCTTCG
JAO1283	CATAGATTAGAACATAACCTTGGA
JAO1284	GTCATCTTCCTAGTTTGTGTAGAACG
JAO1285	TAGGTGTCTGACTTCGATCTTTCG

References

Aplan PD. 2006. Causes of oncogenic chromosomal translocation. *Trends Genet* **22**: 46-55.

Argueso JL, Westmoreland J, Mieczkowski PA, Gawel M, Petes TD, Resnick MA. 2008. Double-strand breaks associated with repetitive DNA can reshape the genome. *P Natl Acad Sci USA* **105**: 11845-11850.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith AJ, Struhl K. 1998. *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.

Bochman ML, Paeschke K, Zakian VA. 2012. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet* **13**: 770-780.

Chen C, Kolodner RD. 1999. Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants. *Nature genetics* **23**: 81-85.

Dunnick W, Hertz GZ, Scappino L, Gritzmacher C. 1993. DNA-Sequences at Immunoglobulin Switch Region Recombination Sites (Vol 21, Pg 365, 1993). *Nucleic Acids Res* **21**: 2285-2285.

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS et al. 2014. The Reference Genome Sequence of Saccharomyces cerevisiae: Then and Now. *G3-Genes Genom Genet* **4**: 389-398.

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

Hermetz KE, Newman S, Conneely KN, Martin CL, Ballif BC, Shaffer LG, Cody JD, Rudd MK. 2014. Large inverted duplications in the human genome form via a fold-back mechanism. *PLoS Genet* **10**: e1004139.

Jackson SP. 2002. Sensing and repairing DNA double-strand breaks - Commentary. *Carcinogenesis* **23**: 687-696.

Katapadi VK, Nambiar M, Raghavan SC. 2012. Potential G-quadruplex formation at breakpoint regions of chromosomal translocations in cancer may explain their fragility. *Genomics* **100**: 72-80.

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

Kim N, Jinks-Robertson S. 2011. Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast. *DNA Repair* **10**: 953-960.

Lemmens B, van Schendel R, Tijsterman M. 2015. Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nat Commun* **6**.

Maizels N. 2006. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol* **13**: 1055-1059.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y et al. 2008. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* **82**: 477-488.

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

Putnam CD, Pallis K, Hayes TK, Kolodner RD. 2014. DNA Repair Pathway Selection Caused by Defects in TEL1, SAE2, and De Novo Telomere Addition Generates Specific Chromosomal Rearrangement Signatures. *Plos Genetics* **10**.

Symington LS, Rothstein R, Lisby M. 2014. Mechanisms and regulation of mitotic recombination in Saccharomyces cerevisiae. *Genetics* **198**: 795-835.

Tornaletti S, Park-Snyder S, Hanawalt PC. 2008. G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. *J Biol Chem* **283**: 12756-12762.

VanHulle K, Lemoine FJ, Narayanan V, Downing B, Hull K, McCullough C, Bellinger M, Lobachev K, Petes TD, Malkova A. 2007. Inverted DNA repeats channel repair of distant double-strand breaks into chromatid fusions and chromosomal rearrangements. *Mol Cell Biol* **27**: 2601-2614.

Willis TG, Dyer MJS. 2000. The role of immunoglobulin translocations in the pathogenesis of B-cell malignancies. *Blood* **96**: 808-822.

Yadav P, Harcy V, Argueso JL, Dominska M, Jinks-Robertson S, Kim N. 2014. Topoisomerase I Plays a Critical Role in Suppressing Genome Instability at a Highly Transcribed G-Quadruplex-Forming Sequence. *Plos Genetics* **10**.

Yu VPCC, Koehler M, Steinlein C, Schmid M, Hanakahi LA, van Gool AJ, West SC, Venkitaraman AR. 2000. Gross chromosomal rearrangements and genetic exchange between nonhomologous chromosomes following BRCA2 inactivation. *Gene Dev* **14**: 1400-1406.

Chapter Three

Relative Contributions of Break-Induced Replication and Canonical Reciprocal Homologous Recombination to the Formation of Spontaneous Translocations in Yeast

Summary

Our laboratory has previously developed an assay system to investigate the formation of spontaneous/un-induced chromosomal rearrangements in yeast diploids through the use of a gene amplification selectable marker (*SFA1-CUP1*; (Zhang et al. 2013)). The majority of rearrangements recovered from this selection are non-reciprocal translocations (NRTs) involving the deletion of a terminal segment of a chromosome and amplification of a terminal segment of the chromosome containing the *SFA1-CUP1* marker. The breakpoints of these rearrangements typically occur at dispersed repeats, suggesting the involvement of a non-allelic homologous recombination (NAHR) double-strand break repair (DSBR) pathway. However, the specific mechanism of homologous recombination responsible cannot be unambiguously assigned. This is because NRTs formed through NAHR are indistinguishable at the karyotype level whether they originated from canonical reciprocal homologous recombination (CRHR) or from non-reciprocal break-induced replication (BIR). One feature that does distinguish the two is that the CRHR pathway gives rise to balanced and unbalanced translocations at a 1:1 ratio, whereas BIR only produces unbalanced translocations. Thus, the ratio of balanced to unbalanced translocations can give an indication of the relative contribution of each pathway.

I designed an experimental system to investigate the relative contributions of CRHR and BIR to the formation of spontaneous non-reciprocal translocations by selecting for yeast clones carrying a non-reciprocal translocation through the generation of a functional copy of the *LYS2* gene as a result of homologous recombination between two non-allelic *LYS2* truncated alleles. One truncated allele is present on chromosome 10 (Chr10) and it is missing the 5' end of the gene (-*YS2*), while the second allele is present on Chr04 and is missing the 3' end of the gene (*LYS*-). Furthermore, the *LYS*- recombination substrate was integrated at five positions on Chr04 increasingly distant from the right telomere in separate clones. Neither allele alone is sufficient to render the cells Lys+, able to grow on media lacking lysine, but cells carrying a translocation mediated by recombination between *LYS*- and -*YS2* have a complete and functional *LYS2* gene at the rearrangement junction. The *LYS*- allele on Chr04 was designed and synthesized such that it contains silent single nucleotide polymorphisms (SNPs) at regular 206 bp intervals along the *YS* portion of the gene still coding for the same amino acids as the wild type sequence. I grew cultures of this strain and plated them on selective medium to identify Lys+ clones. Incorporation of a *SFA1-CUP1* reporter cassette conferring gene dosage-dependent tolerance to formaldehyde and copper served as a means to screen the Lys+ clones for unbalanced translocations. Additionally, since yeast Pol32 protein is an accessory subunit of DNA polymerase delta complex that is required for BIR but not for CRHR, I constructed isogenic *pol32Δ* strains and characterized their rate of Lys+ recombination and their balanced to unbalanced translocations ratio in *pol32*Δ mutants that are compromised for BIR. The analysis of the variation in recombination rates and balanced to unbalanced translocation ratios in the various experimental strains suggests that BIR primarily contributes to translocations involving recombination near the telomere, and not at inner regions of Chr04.

Introduction

The Copy Number Variation Reporter system

The importance of understanding the mechanisms and consequences of Copy Number Variations (CNVs) is due to the fact that they underlie natural genetic diversity and susceptibility and pathogenesis for many diseases. CNVs involve an alteration in gene dosage (amplifications and/or deletions of chromosomal segments) (Hastings et al. 2009; Mishra and Whetstine 2016). Additionally, translocations can occur in which duplicated segments end up on a different chromosome other than where that segment originated from, replacing a deleted segment on another chromosome.

Cumulative research evidence within the past few years has brought to the attention of scientific and medical communities that CNVs have been under-recognized in comparison to nucleotide mutations. CNVs play a previously underestimated role in the development of a myriad of medical disorders and are also a hallmark of cancer genomes (Gu and Lupski 2008; Ionita-Laza et al. 2009; Zack et al. 2013; Cooper et al. 2014). In 2013, a dataset analysis of approximately 3000 tumor samples from The Cancer Genome Atlas Project identified CNVs to some degree in almost all tumor types (Weinstein et al. 2013). Furthermore, the tumor types were separated into two major categories based on whether those tumor displayed a majority of point mutations (nucleotide-level alterations) or structural variations (large-scale mutations) (Ciriello et al. 2013). Results showed that CNVs are particularly prevalent in breast and ovarian tumor types (Krepischi et al. 2012; Park et al. 2015).

A diploid CNV assay previously developed by our laboratory (Zhang et al. 2013) plays a central role by helping us to detect chromosomal rearrangements associated with gene amplifications within the context of our research studies. It involves a reporter cassette that can be

integrated anywhere in the yeast genome which consists of three genes, (*SFA1*, *CUP1* and *KAN*). The *SFA1* gene encodes formaldehyde dehydrogenase, an enzyme that functions in formaldehyde detoxification. The *CUP1* gene encodes metallothionein which binds to and mediates resistance to high concentrations of copper. The genes *SFA1* and *CUP1* confer gene dosage dependent resistance to formaldehyde and copper. Therefore, the more copies of the reporter present, the more resistant the yeast strain becomes.

One such example of a Formaldehyde and Copper Resistant (FCR112) clone obtained in our laboratory from this assay is shown in Figure 3.1. This schematic representation illustrates the relevant features of Chromosome 4 (Chr04) and Chr07 with telomeres represented by squares, the centromere as a circle, the reporter integrated on the right arm of the chromosome and the black arrows to indicate Ty elements or dispersed repeats. There are analogous regions in human genomic DNA called HERV-1 (Human Endogenous Retrovirus 1) elements and also other repetitive DNA elements that can serve as recombination substrates called LINEs (Long Interspersed Nuclear Elements) (Bose et al. 2014; Startek et al. 2015b; Trela et al. 2016). When the parent strain of FCR112 is plated in the absence of formaldehyde and copper it shows normal growth, but at high concentrations of these two inhibitory chemicals it fails to grow. However, in cells where a genome rearrangement has caused duplication of the *SFA1-CUP1* reporter, the amplification of the resistance genes allows the yeast to grow, allowing us to isolate FCR clones. In this example, the reporter was integrated on the right arm of Chr04 in the parent diploid strain and the FCR112 clone was isolated. By using a molecular cytogenetic technique called array comparative genomic hybridization (array CGH) to gauge relative ploidy level, data showed that there was an amplification of the right arm of Chr04 and a terminal deletion of Chr07, and that the rearrangement occurred at Ty repetitive elements.

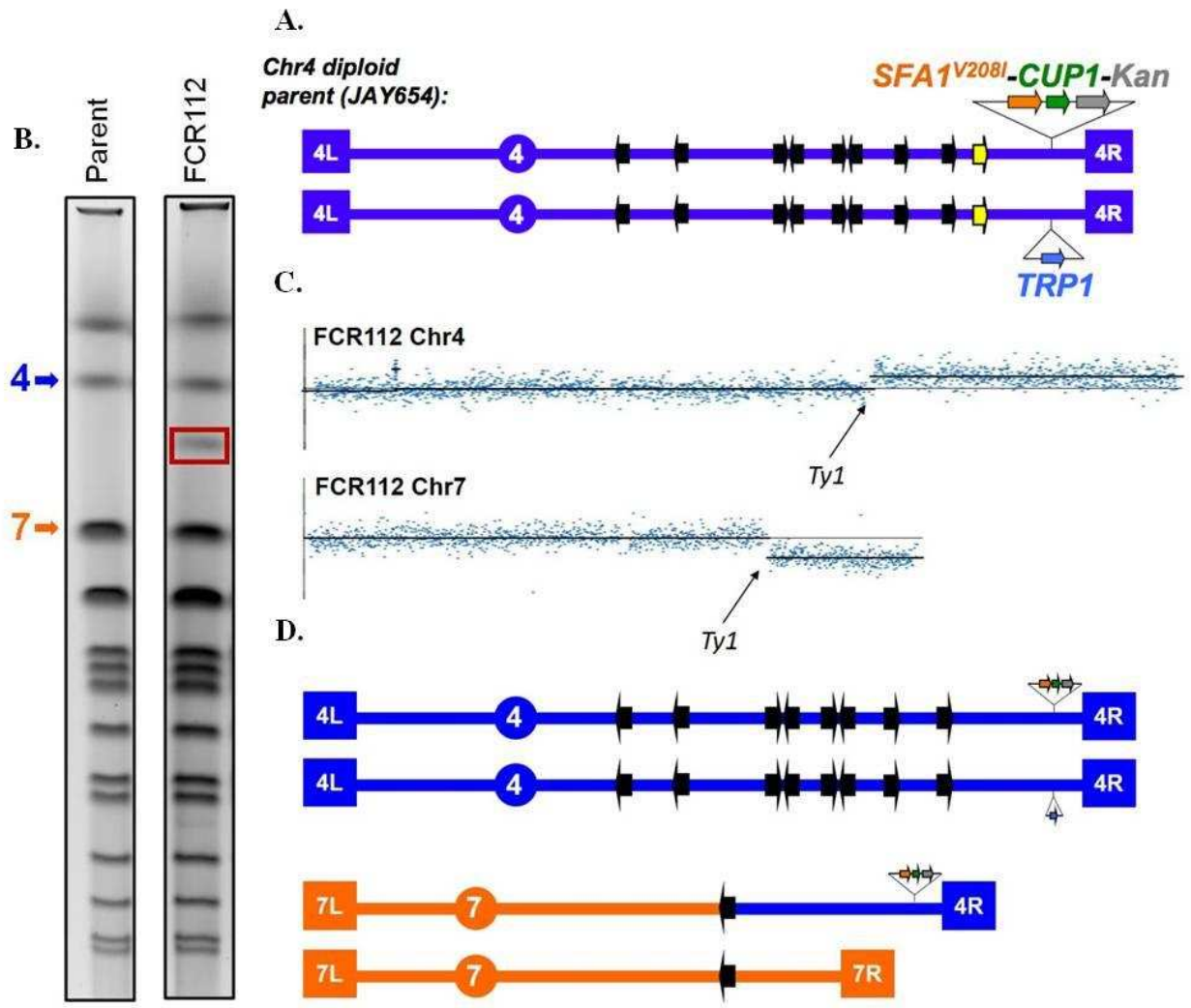


Figure 3.1. Qualitative analysis of a CNV-associated rearrangement in a diploid.
A. Chromosome 4 of JAY654 (parental diploid strain) showing the CNV reporter cassette represented by three block arrows on the right arm, labeled *SFA1-CUP1-KAN*. Black arrows represent Ty elements.
B. Electrophoretic results from the parental strain are on the left. On the right is FCR112, a copper-formaldehyde resistant clone obtained from the CNV assay with a new chromosome band highlighted in red.
C. Array-CGH results from FCR112 showing amplification of the right arm of Chr04 and a deletion of the right arm of Chr04. Breakpoints are identified at Ty1 elements.
D. FCR112 karyotype: Chr07/Chr04 non-reciprocal translocation products inferred by combination of PFGE, array-CGH results and summation of chromosome segments lost and/or gained.

Pulsed-field gel electrophoresis (PFGE), which allowed us to separate out whole, intact yeast chromosomes showed a new chromosomal band in the FCR clone that was previously absent in the karyotype of the parental strain. The new chromosome is ~1.3 Mb in this clone and the combined size of the preserved segment of Chr07 and the amplified segment of Chr04 is 1.3 Mb. Therefore, we can infer the molecular structure of this chromosomal rearrangement to be as such: A terminal deletion on Chr07 where the terminal amplification of Chr04 is now attached (Zeidler and Argueso, unpublished data). Different projects in our laboratory have isolated several FCRs that display non-reciprocal translocations such as this one.

Project Rationale

Typically, a double-ended DNA double-strand break can be repaired relatively error-free by a conservative allelic homologous recombination mechanism. However, other events such as replication fork collapse and telomere uncapping typically lead to a DSB in which there is only one free DNA end available to initiate repair which occurs by a process known as break-induced replication (BIR)(Llorente et al. 2008a).

Non-reciprocal translocation (NRTs) outcomes, the most abundant type of translocation identified using the *SFA1-CUP1* CNV experimental system, involve a terminal deletion on one chromosome with an amplification of the end of the chromosome containing the reporter. The breakpoints are often at repetitive regions of DNA suggesting non-allelic homologous recombination, of which the mechanism remains ambiguous. We were interested in investigating which double-strand DNA break repair process brings about these outcomes. Therefore, resolving the ambiguity of the mechanism responsible is the motivation behind this project. Homologous recombination requires a homologous sequence to guide repair, and break-induced replication, a

variant of homologous recombination, involves repair of only one side of the break, but both of these pathways can lead to identical NRT karyotypes. Therefore, we seek to understand what the contribution of BIR is to the formation of NRTs.

The schematic diagram Figure 3.2 depicts the chromosomes relevant to our study, Chr04 and Chr10. In the canonical reciprocal homologous recombination (CRHR) pathway, if a double-ended break occurs within a repeat there could be a reciprocal non-allelic crossover event between the regions in grey boxes representing regions of homology. If a reciprocal crossover event occurred, following chromosomal segregation during cell division, the karyotypes of the resultant cells would be CNVs (Chr10 amplified and attached to Chr4 and vice versa) which would occur in a 1:1 ratio. Otherwise, the resulting karyotypes would include outcomes either resembling that of the parental chromosome and ones with a reciprocal (balanced) translocation, where ends of each chromosome were exchanged but copy number remained unchanged.

In break-induced replication (BIR), after the double-strand DNA break, the centromere distal end of the chromosome is lost and only the centromeric side of the break undergoes repair by strand invasion into a non-allelic region of homology which, in this case, is on Chr04 and proceeds by replication to the chromosome end, healing the broken Chr10. It should be noted that these two identical outcomes (highlighted in red) can arise by two very different mechanisms. Furthermore, even if amplification tracts are longer than generally expected for BIR, is still not possible to rule it out since the resultant karyotypes are the same. Therefore, two of the major questions to investigate were (i) what is the relative contribution of BIR to the formation of these products? (ii) is BIR more likely to occur or contribute to these translocations depending on how far it would need to proceed? In other words, “does the location of the donor sequence (region of homology) affect the relative abundance of BIR repair?”

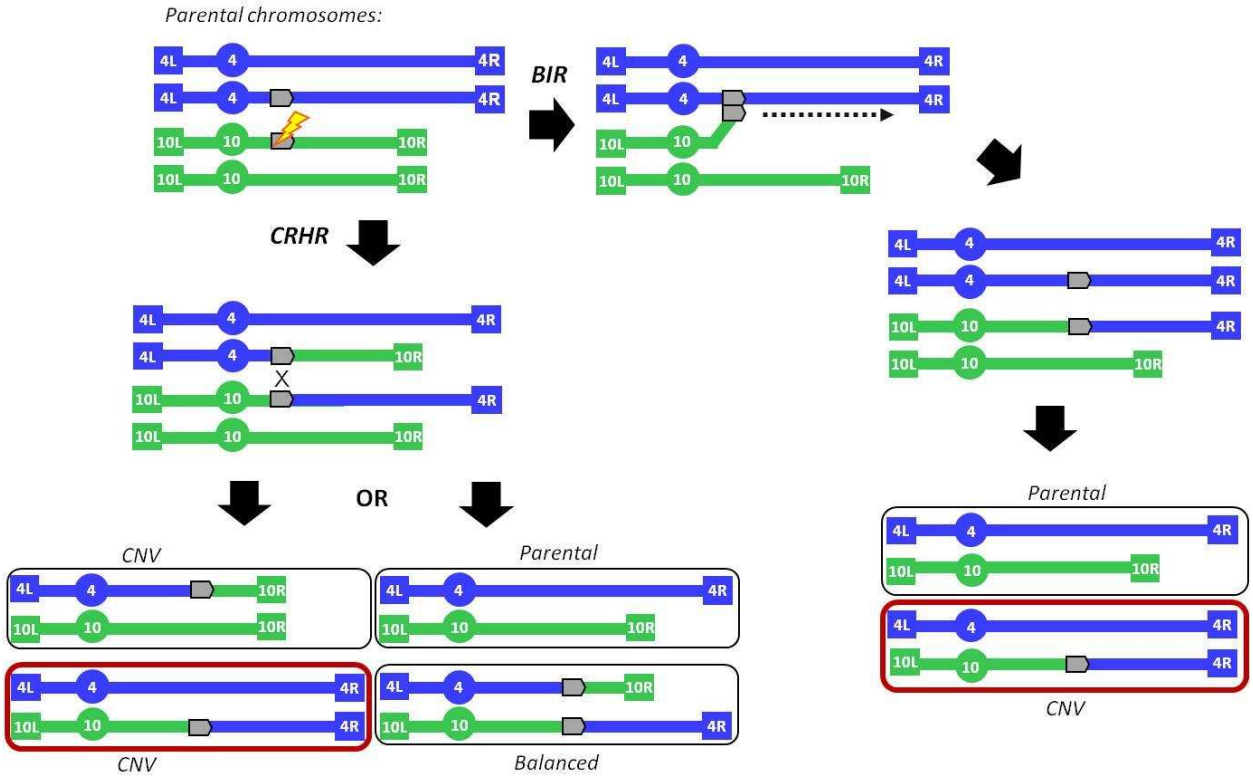


Figure 3.2. Mechanisms of Non-Allelic Homologous Recombination (NAHR). Schematic representation of the Canonical Reciprocal Homologous Recombination (CRHR) pathway vs. the Break-Induced Replication (BIR) pathways depicted in diploid cells at S-phase with duplicated chromosomes 4 and 10 relevant to our experimental design shown at top. Inserted regions of homology are shaded grey. The formation of a double-strand break (shown by lightning bolt) initiates the DNA break repair process. In CRHR (on left), after a crossover event the progenies that would be produced after chromosome segregation and cell division are depicted at the bottom. This would lead to copy number variants or to karyotypes that either resemble the parental chromosomes or involve a balanced/reciprocal translocation. In BIR (upper and right side), the end of Chr10 could be lost, necessitating the repair by using a non-allelic template. The exposed ssDNA homology could invade Chr04 at the homologous region and proceed to copy to the end of the chromosome. The final karyotypes would be similar to that of the parent or be a non-reciprocal translocation. Distinguishing between the identical non-reciprocal translocations highlighted in red is the primary focus of this project.

Description of the experimental system

To provide a substrate of homology for recombination to occur, a portion of the *LYS2* gene was integrated on Chr10, missing the 5' end of the gene (-*YS2*). The other substrate was integrated on Chr04 and was missing the 3' end of the gene (*LYS*-) but both alleles have the *YS* portion in common. These recombination substrates share ~2 Kb of homology (*YS* segment). Neither truncated allele encodes a functional version of *LYS2*, but if recombination between their shared regions occurs it generates a translocation resulting in a functional full-length gene that restores lysine prototrophy. Figure 3.3 shows the constructs, in addition to the CNV reporter, and the possible resultant recombined products. Thus, lysine dropout medium was used to select for recombination products producing functional *LYS2* genes. The *SFA1-CUP1* marker is present at a distal position on Chr04 to phenotypically distinguish if a given Lys+ clone carries an amplification of the right arm of Chr04 or not, thus allowing us to determine if the associated *LYS*-/-*YS2* translocation is non-reciprocal/unbalanced or reciprocal/balanced (Formaldehyde-Copper resistant or sensitive, respectively).

I created a diploid experimental strain set in which the Chr10 -*YS2* is present at a fixed position, while the Chr04 *LYS*- was been inserted at five different loci at increasing distances from the right telomere (Figure 3.4) to determine if a more distal donor substrate position facilitates the formation of BIR-mediated non-reciprocal translocations. Furthermore, we constructed *pol32*Δ mutant strains since yeast Pol32 protein, an accessory subunit of DNA polymerase delta complex, is required for BIR. Thus, the BIR pathway is severely deficient in the mutant. Using this strain construction, we should be able to indirectly measure the relative contribution of BIR to the formation of chromosomal rearrangements by monitoring the formation of balanced and unbalanced translocations.

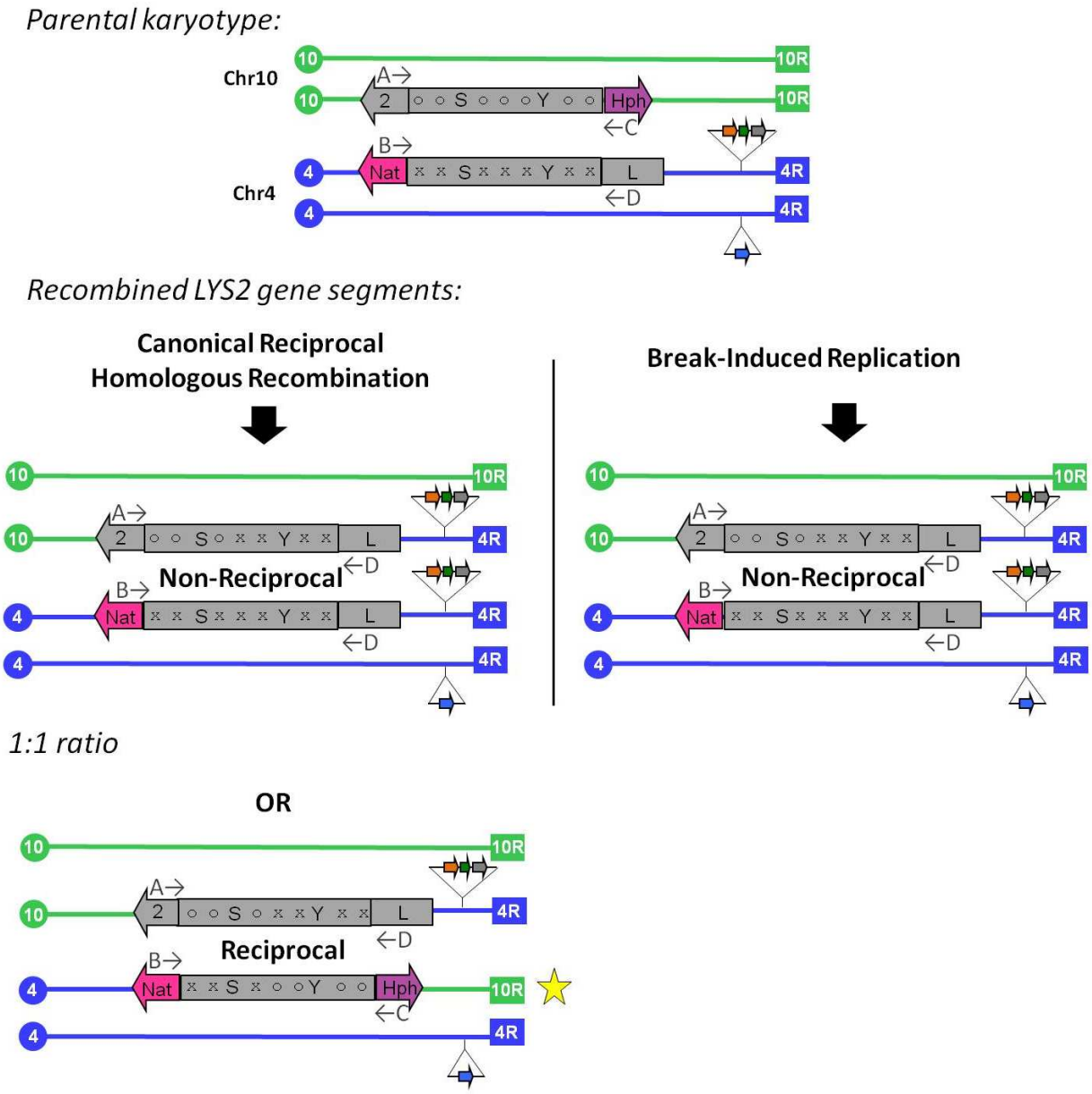


Figure 3.3. Maps of the right arms of Chr04/Chr10 and alternate outcomes.
Engineered *LYS2* homologous recombination substrates present in the parental strains (top). Outcomes from canonical reciprocal homologous recombination (right) showing both non-reciprocal and reciprocal outcomes which would occur at a ratio of 1:1. The star indicates the Chr04/Chr10 translocation that can only be produced through the canonical reciprocal homologous recombination pathway. It can be identified molecularly by PFGE and PCR (by using primers labeled as A, B, C and D.

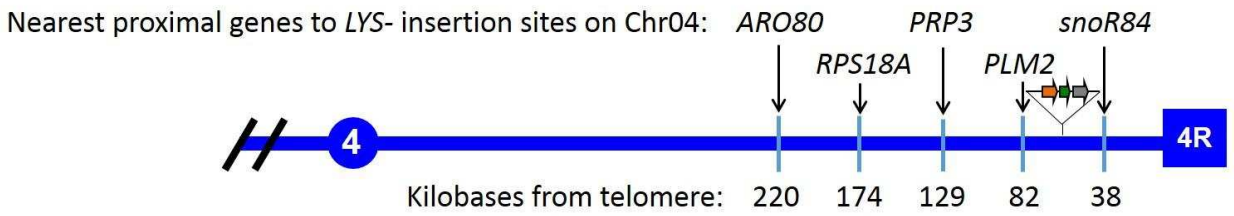


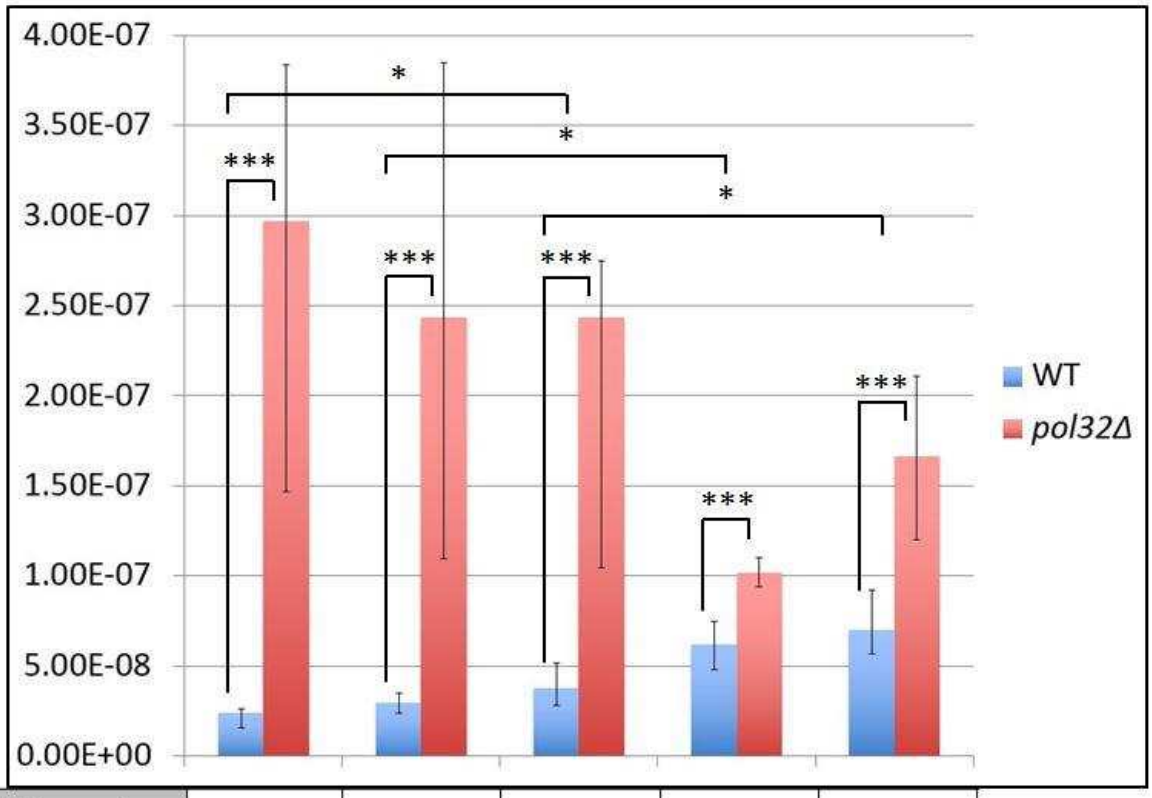
Figure 3.4. Chromosome 4 *LYS*- insertion sites
The right arm of Chr04 is shown with the insertion sites of the *LYS*- construct identified by the name of the most proximal gene name at each locus. The CNV reporter cassette is represented between the *PLM2* (82 Kb) and *snoR84* (38 Kb) sites. Note: Distances not drawn to scale. The distance from the centromere to the right telomere is approximately 10.9 Kb.

Results & Discussion

Quantitative analysis of LYS-/-YS2 translocations

The first step in our analysis was to determine the overall rate at which recombination occurred between the *LYS*- and -*YS2* substrates. Figure 3.5 shows the rates of translocation formation per genome per cell generation and 95% confidence intervals for wild type and *pol32*Δ mutants calculated using the Lea-Coulson method of the median for ≥ 20 independent cultures. In wild type strains, we noted a pattern in which there was approximately a 3-fold increase in *LYS2* recombination as the donor sequence was moved closer to the telomere in wild type strains, these comparisons are noted with a single asterisk in Figure 3.5 where $p = <0.05$. The overall rate of recombination is dependent on the frequency of the initiating DSB lesion, and on the efficiency of pathway used to repair such lesions. Assuming that the frequency of DSBs remains comparable for all *LYS*- insertions, the efficiency of repair would be the main factor causing the progressive elevation in recombination that we observed. While the efficiency of the CRHR pathway is not predicted to change between the sites, the efficiency of BIR is. It has been previously shown that BIR does not operate efficiently when the replication tracts are larger than ~100 Kb (Donnianni and Symington 2013). Therefore, our results are consistent with a model in which the relative

contribution of BIR is higher for telomere proximal *LYS*- substrates (*PLM2* and *snoR84*) than it is for centromere proximal ones (*ARO80* and *RPS18A*).



Chr04 insertion site	*ARO80*	*RPS18A*	*PRP3*	*PLM2*	*snoR84*
Distance from telomere	220 Kb	174 Kb	129 Kb	82 Kb	38 Kb
Ratio of *pol32Δ*/WT	12.4	8.2	6.4	1.6	2.4

Figure 3.5. *LYS- / -YS2* Non-allelic Homologous Recombination translocation rates.
The comparison of the number of Lys+ colonies to those appearing on YPD are represented by the recombination rate data. The mutation rates per cell generation and 95% confidence intervals for wild type and *pol32*Δ calculated using the Lea-Coulson method of the median for ≥ 20 independent cultures are shown with computed ratios identifying the fold-difference between the mutant and wild type strain at five different locations.
*** = $p < 0.0001$; * = $p < 0.05$

If the variation in the *LYS2* recombination rate between *LYS*- insertion sites was caused primarily by variation in the contribution of the BIR pathway, then we would expect that *pol32*Δ mutant cells unable to carry out BIR would display comparable rates of recombination regardless of the position of the *LYS*- substrates. In each of the five *pol32*Δ derivative strains, there was an overall elevation in the recombination rate compared to the corresponding wild type. The significantly elevated recombination rates relative to wild type levels (p = <0.0001 for each; represented by three asterisks in Figure 3.5) suggest a general genome stability defect in *pol32*Δ. The moderate increases observed are likely because, even though Pol32 protein is not essential, its absence likely affects the functionality of the replicative DNA polymerase delta. Less than optimal replication processivity could lead to more replication fork stalling and collapse, thus lending to higher initiation of recombination genome wide. The spontaneous recombination rates determined from our experiment are in the 10^{-7} range. Compared to other BIR studies in inducible systems, there are six orders of magnitude difference in the overall occurrence of these events meaning that the events we investigated are actually quite rare, while conventional HO-induced systems may convey the idea that BIR is very frequent. The primary driver of this difference is that induction causes more breaks.

Interestingly, there was no significant difference in the *LYS2* recombination rates between four of the five *pol32*Δ donor site locations, including the two most extreme sites (ARO80 vs. snoR84; p = 0.12). This eliminated the progressive elevation in the recombination rate observed in wild type, and possibly even reversed it. The only exception was for *PLM2*, where *LYS*- substrate insertion site is immediately adjacent to where the CNV reporter cassette was integrated, which could be contributing to this confounding result through an unknown mechanism.

It is of note that larger genomic contexts may contribute to the differences observed at the *PLM2* site as well, such as altered DNA architecture due to nucleosome position, etc.

Together these quantitative recombination rate observations indicated that there may be a higher contribution of BIR to the formation of Lys+ clones as the donor sequence is moved closer to the telomere. In order to definitively show such an effect, we proceeded to characterize the qualitative nature of the translocations formed in both wild type and *pol32Δ* strains across all five *LYS*- insertion sites, specifically determining the relative abundance of non-reciprocal and reciprocal translocations.

Phenotypic descriptions and spectra of structural variants

CRHR is predicted to yield Lys+ clones carrying balanced and unbalanced translocations at a ~1:1 ratio, while BIR is predicted to yield only unbalanced translocations. Our intention was to infer the relative contribution of each recombination mechanism by the comparison of these outcomes. We examined approximately 50 independent Lys+ clones from each of the ten diploid strains constructed and classified them into categories based on their biochemical or physiological properties dictated by their genotype. Phenotypes were scored based on growth on selective media. The phenotype expected for clones carrying a non-reciprocal translocation would be: $CuFA^R$, Hyg^S, Lys+, Gen^R, Trp+, and Nat^R (R and S indicate resistance and sensitivity, respectively). The phenotype for clones carrying a reciprocal translocation would be: $CuFA^S$, Hyg^R, Lys+, Gen^R, Trp+, Nat^R.

Unexpectedly, upon phenotypic analysis (Table 3.1), we observed that many Lys+ clones did not fit neatly into the two ascribed categories predicted by the model in Figure 3.3. A substantial fraction of clones were: Gen^S or Trp- or NatS or ($CuFA^R$ and Hyg^R) or ($CuFA^S$ and Hyg^S). We did not fully anticipate the abundance of so many clones exhibiting these other

characteristics. These findings were indicative that a variety of alternative and more complicated mechanisms of recombination occurred in order to generate Lys+ clones. Disassociation and reinvasion of ssDNA during replication, referred to as template switching, is a common occurrence in yeast and may partially explain these alternate genetic outcomes (Smith et al. 2007). The abundance of clones in the "other" category drastically reduced the overall number of clones in the non-reciprocal and reciprocal categories, and thus impaired our ability to calculate robust non-reciprocal to reciprocal translocation ratios from phenotypes alone. Therefore, the numbers we did get in these two categories were relatively small and no conclusive trends were readily apparent.

Table 3.1. Phenotype classes for Lys+ clones.
Non-Reciprocal Phenotype: $CuFA^R$, Hyg^S, Lys+, Gen^R, Trp+, Nat^R
Reciprocal Phenotype: $CuFA^S$, Hyg^R, Lys+, Gen^R, Trp+, Nat^R
Other Phenotype: Gen^S or Trp- or Nat^S or ($CuFA^R$ and Hyg^R) or ($CuFA^S$ and Hyg^S)
N/A: The CuFA resistance phenotype was not applicable because the *LYS*- recombination substrate was integrated at a locus distal to the CNV reporter cassette.

Integration Site	Distance to TEL04R	Strain	Non-reciprocal	Reciprocal	Other	Total
ARO80	220 Kb	WT	16	20	15	51
		pol32Δ	11	16	21	48
RPS18A	174 Kb	WT	11	11	28	50
		pol32Δ	13	8	29	50
PRP3	129 Kb	WT	14	9	27	50
		pol32Δ	12	19	15	46
PLM2	82 Kb	WT	7	22	22	51
		pol32Δ	12	4	34	50
snoR84	38 Kb	WT	N/A	N/A	15	N/A
		pol32Δ	N/A	N/A	23	N/A

In order to better gauge the relative abundance of non-reciprocal and reciprocal outcomes present among all Lys+ clones, we would need to directly examine the karyotypes of each clone. These studies were ongoing at the time this thesis was submitted. Specifically, we are using PCR and PFGE to identify the presence of the reciprocal Chr04/Chr10 translocation with a *YS* structure

at the breakpoint (star in Figure 3.3). This approach directly interrogates the occurrence of the CRHR pathway and should allow us to obtain a conclusive assignment for every Lys+ clone. The confirmation of the presence of a *YS* PCR product would be a direct indication of CRHR (Figure 3.3), although it is still possible that the sample size of our data set precludes the identification of a specific pattern. In this case, additional Lys+ clones would be isolated and analyzed by PCR.

Future Directions and Studies

BIR impairment alternative

In a haploid yeast CNV experiment conducted by Payen *et. al.* they reported that for a specific segmental duplication in a chromosome, *POL32* was required. So, the duplication of that region occurred at 10^{-7} in wild type and in the *pol32*Δ mutant it was basically undetected, indicating it was strictly required (Payen et al. 2008b). However, we see something different in our experiment using diploids and looking for translocations associated with long range BIR. Somehow, the *POL32* defect has an overall effect of destabilizing the genome by a three to 12-fold difference (depending on where in the chromosome the *LYS*- insertion was placed). We believe that the necessity of *POL32* for BIR is likely completely independent of the increased recombination rate. Pol32 protein is a non-essential subunit of the replicative DNA polymerase delta. By missing this subunit in the mutant strains, the DNA replication defect is revealing itself. There are probably more breaks occurring spontaneously from additional collapsed replication forks, thus driving more recombination. The repair of the increased amount of spontaneous breaks will be different since they do not have the BIR option anymore. Interestingly, there is repair of those additional breaks, but it is not changing on account of whether the recombination substrate is close to or far from the end of the chromosome. This is logical since BIR is not playing a role.

Pif1, an evolutionarily conserved helicase has recently been identified to play a crucial role in BIR. It is purported to actively minimize topological constraint during the extensive DNA synthesis activity that creates a migrating bubble structure, also known as a D-loop (Saini et al. 2013; Wilson et al. 2013). Since *PIF1* is also required for BIR it could be used as an alternative way to investigate the BIR contribution to NRTs.

Identifying reciprocal translocations and gene conversion patterns

In our experimental system, a balanced Chr04/Chr10 translocation would produce a recombined *LYS2* gene segment of only the -*YS*- portion, along with selectable markers. This reciprocal product can only be produced through the canonical reciprocal homologous recombination pathway, indicated by the star in Figure 3.3. It would be possible to take advantage of this unique characteristic by identifying it molecularly by pulsed-field gel electrophoresis and by detecting it by PCR. These additional methods would aid in resolving the ambiguity presented by the current phenotypic analysis of the Lys+ clones recovered. Particularly, it would enable us to further differentiate between the clones currently classified in the "other" phenotype group (Table 3.1).

By design, our recombination substrates share ~2 Kb of homology (the *YS* region). The *YS* segment integrated on Chr10 maintains a wild type nucleotide sequence. On the other hand, the *YS* segment integrated at different positions along Chr04 was synthesized such that it contains synonymous single nucleotide polymorphisms (SNPs) at regular 206 bp intervals. This allelic difference would allow us to map the site of genetic exchange by following the distribution of SNPs at the recombination breakpoints. SNP analysis by Sanger sequencing would reveal if there were any patterns to the gene conversion tracts to investigate whether the CRHR and BIR might

produce different SNP signatures at their respective recombination sites. These investigative steps would provide more evidence as to whether there is indeed a correlation between the locations of the donor sequence and how it may affect the relative contribution of BIR to double-strand DNA break repair.

Mutation clusters aka hypermutation

Recently, it was been shown that exposure to methyl methanesulfonate (MMS) causes DNA base damage to the long, single-stranded DNA tracts transiently formed during break-induced replication. The base damage ultimately results in a signature pattern of non-random, co-localized hypermutation ("kataegis"), closely resembling mutation clusters found in cancer (Sakofsky et al. 2014). In contrast, reciprocal canonical homologous recombination is not associated with extensive tracts of ssDNA or of DNA synthesis, thus, should not lead to kataegis. One could take advantage of this new finding to assign a mechanism of formation to the non-reciprocal translocations identified through genome stability assays employed in our laboratory. The BIR pathway is absent in the mutant and the relative contribution of BIR to formation of chromosomal rearrangements is typically *indirectly* determined by subtracting the *pol32*Δ rate from the wild type rate. In this experimental approach, one would potentially be able *directly* make this same estimate by whole genome sequencing of the Lys^+ clones and determining the presence or absence of mutation clusters on the region of Chr04 that is amplified. Specifically, if the translocation was originally formed by CRHR, no mutation clusters would be seen within the tract. In contrast, BIR generated translocations should have such clusters.

Investigation of half-crossover cascades

In the present study, a single translocation event that led to the conjunction of two truncated alleles to make a fully functional gene was investigated. However, the processes that chromosomes regularly undergo to repair broken chromosomes can be much more complex and involve the interaction and exchange of multiple chromosome pieces. Genome instability results in cases of more complicated rearrangements, where more than one non-allelic chromosome may be used as the template for repair of other chromosomes (Aguilera 2009). This ultimately results in a type of genome rearrangement known as half-crossover initiated cascades (HCC)(Vasan et al. 2014). The "cascade" portion of the nomenclature denotes the chain reaction of events that transpire within a single cell cycle to result in the accumulation of several segmental translocations originating from just one DSB lesion. Specifically, when single-ended DSBs are used in CRHR, the repair of the initial lesion leads to the breakage of the template region. The broken template then initiates a secondary repair reaction that can generate another chromosomal rearrangement. Evidence of HCC events have been identified in many forms of cancer. Thus, we would be able to implement our established yeast-based bioassay by integration of an additional recombination substrate on another chromosome (Figure 3.6). In this proposed HCC system, each of the *YS* segments (on Chr04, Chr7and Chr10) of the *LYS2* gene would contain SNP markers. Overall, this strain would be unable to produce its own lysine (Lys-) by lacking a continuous *LYS2* gene. A HCC event would begin with cascading genome rearrangements at a DSB on Chr07. 5′–3′ degradation would generate stretches of ssDNA at the *YS* sequence promoting homology driven invasion into the Chr10 *YS*, followed by a half-crossover resolution. A Chr07/Chr10 non-reciprocal translocation would be formed but also generate a broken Chr10 containing remnant segments of the Chr07 *YS* (orange). Chr10 repair would initiate by invading the Chr04 *YS* and begin BIR and proceed to the

telomere such that the *SFA 1-CUP1* reporter would be amplified. The end result is a Chr10/Chr04 non-reciprocal translocation containing a complete *LYS2* gene, enabling own lysine biosynthesis. This newly formed *LYS2* gene now contains SNP markers from all three chromosomes involved in the cascade (green, orange, blue). Additionally, amplification of the *SFA 1-CUP1* reporter on the right arm of Chr04 would have occurred as well as a deletion of the *2xURA3* reporter previously on the right arm of Chr07.

This system would be useful for determining basal rates of HCC mutation rates quantitatively in wild type yeast cells. Furthermore, in light of the abundant number of clones recovered from our experiment that fit into the "other" phenotypic category, it is likely that further analysis of the molecular nature of those structural variations would reveal HCC events.

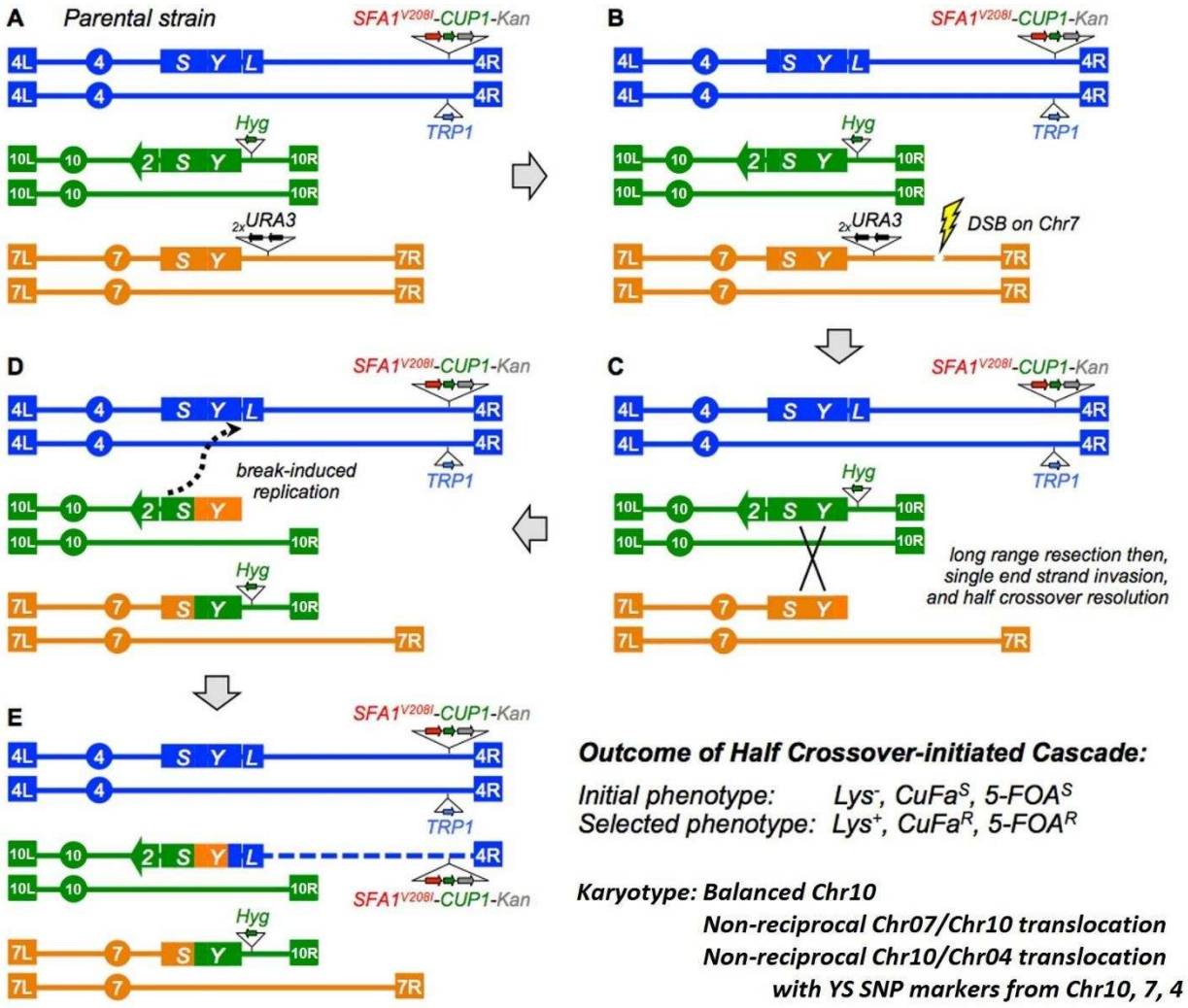


Figure 3.6. Half-crossover mechanism and detection assay. A. Schematic map of the diploid test strain. The *YS* segment of the *LYS2* gene contains unique synthetic SNP markers in each of the three genomic positions. There are no complete/continuous *LYS2* sequences in this strain so it is unable to produce its own lysine. **B.** Cascading genome rearrangements start with a double-strand break (DSB) lesion on Chr07 followed by long-range resection. **C.** Once resection reaches the *YS* sequence, homologous recombination promotes a single-end invasion into the Chr10 *YS*, followed by a half-crossover resolution. This creates the Chr07/Chr10 non-reciprocal translocation and leaves a broken Chr10 containing remnant segments of the Chr07 *YS* (orange). **D.** The broken Chr10 end invades the Chr04 *YS* and initiates break-induced replication that extends through the right arm of Chr04, through the *SFA 1-CUP1* amplification reporter, until it reaches the telomere. **E.** This creates the Chr10/Chr04 non-reciprocal translocation, which contains a complete *LYS2* gene. This *LYS2* gene now contains SNP markers from all three chromosomes involved in the cascade (green, orange, blue). It also amplifies the right arm of Chr04 including *SFA 1-CUP1* reporter and deletes the right arm of Chr07 including the *2xURA3* reporter.

Materials and Methods

Growth conditions and media for propagating strains

E. coli culture medium consisted of Luria Bertani broth (bacteriological agar not added to mixture) or agar prepared by adding 0.1 mg/mL ampicillin for drug resistance selection. *E.coli* plates were placed in a 37C incubator (with shaking rack at 250RPM when necessary).

Yeast culture medium (10 g yeast extract, 20 g glucose, 20 g peptone, and 10 g bacteriological agar in 1 L of distilled water) was used for propagating strains with supplement drugs to the media (e.g. Hygromycin B, Geneticin, etc. where noted) Synthetic drop-out plates were prepared by mixing 1.7g yeast nitrogen base (without amino acids or ammonium sulfate), 1.4g drop-out mix (lysine, tryptophan or complete drop-out), 5 g ammonium sulfate, 20 g glucose, 20 g bacteriological agar, and 1L deionized H20 and then autoclaving. *Petite* clones (carrying genetic mutations causing mitochondrial dysfunction) were screened out on YPGE medium containing 2% ethanol and 2% glycerol as non-fermentable carbon sources.

For the CNV assay, $CuSO_4$ and formaldehyde (FA) were added to autoclaved molten agar medium once it had cooled to 70°C (0 µM $CuSO_4$ and 0 mM FA, 75 µM $CuSO_4$ and 0.75 mM FA, 150 µM $CuSO_4$ and 1.5 mM FA, 225 µM $CuSO_4$ and 1.6 mM FA, 300 µM $CuSO_4$ and 1.7 mM FA). Since aqueous formaldehyde solutions are considered unstable, a fresh 1 M solution of formaldehyde was prepared from a methanol-stabilized stock solution immediately before it was mixed with the autoclaved medium. After all components were mixed, media was poured into Petri dishes and used within 48 hours.

All yeast plates were incubated at 30C (with rotating test tube rack or flask shaker at 200 rpm when necessary). Plates were either loosely plastic wrapped or placed in a plastic box with a damp paper towel if incubated for extended periods of time to prevent desiccation of the agar.

Transformations and DNA extractions

All *E.coli* transformations were performed using either α-Select Silver or Bronze-efficiency competent cells purchased from Bioline following their recommended heat-shock protocol. Yeast transformations were performed using a LiAc/ssDNA/PEG protocol as previously described (Gietz and Woods 1998). Plasmid were recovered from *E.coli* using Qiagen Miniprep kit. Yeast genomic DNA was harvested using zymolyase and potassium acetate. Total yields of dsDNA were determined using Invitrogen Qubit Fluorometer. Sanger sequencing data was obtained by GENEWIZ, Inc. and Colorado State University's Proteomics and Metabolomics Facility. DNA sequence analysis, primer and plasmid construction was performed with Qiagen's CLC Genomics Workbench v4.9

Synthetic LYS2 reporter - gBlock assembly and subcloning

We designed gBlock Gene Fragments (custom synthesized by Integrated DNA Technologies) such that the coding sequence for the *YS* portion (corresponding to SGD coordinates 472974 to 470674 of Chr02) of the *LYS2* gene would include synonymous point mutations spaced every 206 bases, taking in to consideration preferential *S. cerevisiae* codon usage.

gBlocks 1 and 2, the "*S*" and "*Y*" regions respectively, of the *LYS2* gene were diluted 1:1000 and each amplified by PCR with primers JAO1356/1359 and JAO1358/1357. Amplification products were restriction digested with *Sac*I and *Hin*dIII. Each gene segment was amplified and joined by overlapping PCR using primers JAO1356/1360 then purified with a gel band-extraction kit by Fermentas. The purified product was digested with *Nae*I and *Sac*I and cloned into the plasmid carrying the "*L*" segment of *LYS2* (described below).

Strain construction: LYS- and -YS2 constructs

The oligonucleotide sequences and complete genotypes for yeast strains used in this study are listed on the following page in supporting information, Table 3.2 and Table 3.3, respectively. Additional primer sequences used for DNA sequence verification and intermediate construction steps are not listed but available upon request.

The *LYS2*- construct was created by PCR amplification of the wild type "*L*" segment of the *LYS2* gene (corresponding to SGD coordinates 469748 to 468797 of Chr02) from JAY357. The PCR product was ligated into pUC19 plasmid vector following digestion with *Hin*dIII and *Kpn*I and transformed into *E.coli*. The recovered plasmids were then digested with *Nae*I and *Sac*I and the joined *YS* gBlock gene fragments (described above) were ligated into this vector which was transformed into *E.coli* for propagation and then recovered.

The -*YS2* construct was created via knockout deletion of the "*L*" segment from a wild type *LYS2* sequence in strain JAY648. The deletion cassette was PCR-amplified with primers JAO1009/1389 using pAG32 (HphMX4; hygromycin resistance) as template. Transformations were plated on YPD, replica plated to YPD plus Hygromycin and colony singles tested on Hygromycin, Geneticin, YPGE and lysine drop-out plates to confirm phenotypes.

All plasmid constructions were verified by Sanger sequencing to ensure that the PCR amplification process had not introduced any mutations. Selection for *E.coli* transformants was achieved by ampicillin drug resistance selection.

The CNV reporter cassette was previously integrated into the yeast genome on the right arm of Chr04 between the *PML2* and *SAM2* at a distal location 78558 Kb from the right telomere along with an auxiliary marker at the allelic position in the other homolog.

Each of the genetic changes that composed the final genotypes of the diploid experimental strains were generated individually in isogenic haploid strain. These haploids were then successively intercrossed and taken through meiotic tetrad analysis to obtain combined genotypes.

Table 3.2. Sequences of synthetic oligonucleotides used in this study

Primer ID#	Primer sequence in 5'-3' orientation
JAO871	CAT CAC AAT TAG TAA TGG AAA GTG TTT GGA AAA AAA AAG AAG AT AA GGC GCG CCA GAT CTG
JAO872	TAA CAA CCA GAA ATA GGC TTT AGT TAA CTC AAT CGG TAA TTA CAT AGG CCA CTA GTG GAT
JAO1009	ATA TTC ATT GAA ACT GAT TAT TCG ATT TTC TTC TTG ACC GCT CGT TTT CGA CAC TGG
JAO1356	atcgatcggagCTCTCAATCCGCCGGTGTTAC
JAO1357	TAGGGGTGTGGATTTGATGGTATG
JAO1358	atcgatcggagCTC AAATGTGCGTATCTATTTCTCC
JAO1359	atcgatcg AAGCTT GGGTTCAGAATCGAATTAGGAG
JAO1360	atcgatcg AAGCTT TAGGGGTGTGGATTTGATGGTATG
JAO1389	GCAACACCTAAGTAAATGGTTTGTCTGGCTGGGGGGATATGCAGGGTCGATAACTGAAAAGGTTGCGCC AGGTCGACGGATCCCCGG

Table 3.3. Yeast strains used in this study

Strain ID #	Ploidy	Mating type	Genotype
JAY357	1n	alpha	ade5-1 his7-2 leu2-3,112 LEU+ ura3-52 trp1-289 cup1D RSC30 sfa1D::hisG
JAY648	1n	alpha	ade5-1 his7-2 leu2-3,112 LEU+ ura3-52 trp1-289 cup1D RSC30 sfa1D::hisG PLM2::SFA1-V208I-CUP1-Kan
JAY654	2n	a/alpha	ade5-1 his7-2 leu2-3,112 LEU+ ura3-52 trp1-289 cup1D RSC30 sfa1D::hisG homozygous heterozygous PLM2::SFA1-V208I-CUP1-Kan/PLM2::TRP1

Integration of LYS- and -YS2 constructs into yeast chromosomal loci

Sequence integration was guided, in all instances, to the chromosomal locus of choice by homology-directed homologous recombination utilizing extended oligonucleotide "tails" with at least 40 bp of homology to allow for reassembly into the yeast genome following a single transformation. Chr04 loci of the five *LYS*- insertion were named according to the nearest proximal gene: *ARO80, RPS18A, PRP3, PLM2, snoR84*. The corresponding SGD coordinates for these insertion sites were: 1315070, 1361005, 1406774, 1453105, 1493622, respectively.

All integrations of constructs into yeast genomes were verified by PCR, either by amplification across the entire inserted region using primers complementary to the native flanking chromosome sequence or using primer pairs at the "borders" of the integrations sites (i.e. one primer complementary to the linearized plasmid and one primer complementary to a nearby sequence on the chromosome near the integration site.

POL32 knockout

Isogenic *pol32Δ*::Nat homozygous diploid strains were constructed by PCR-mediated gene knockout using JAO871/872 to amplify the Nourseothricin resistance gene from pAG25 (natMX4) to replace *POL32* gene in haploid cells (Goldstein and McCusker 1999). Strains were crossed to ones previously described and yeast tetrad analysis was performed to generate the final diploid strains with desired genetic markers.

Quantitative LYS-/-YS2 non-allelic recombination assay

Lys- parental strains were struck out to single colonies on YPD medium and grown for 2 days at 30C. Independent 5mL YPD cultures were inoculated with individual colonies and grown for 24 hours. 2mL from each culture were centrifuged at 3500rpm, and the supernatant discarded. The cell pellet was resuspended in 2ml deionized (DI) water and the centrifugation and discarding process was repeated. The washed pellet was resuspended in a final volume of 1mL DI water. 500uL of this undiluted cell suspension was plated on lysine drop-out (selective) medium and spread on the agar using autoclaved glass beads. A serial dilution of the cell suspension was performed in 96-well plates for each sample and 100uL of a 10^{-5} final dilution was plated on YPD (permissive) medium plates. Colonies were counted on YPD plates following two days of growth

and at five days for ones grown on lysine drop-out medium. A single Lys+ clone from each culture was selected to ensure mutation independence before it was further characterized. Approximately 50 independent cultures were inoculated for each strain (overall total of five wild type strains and five *pol32*Δ strains), to generate a final total of approximately 500 Lys+ clones.

Mutation rate determination

The total number of colonies formed on lysine drop-out plates were counted and compared to the number of colonies formed on YPD plates. The colony numbers were used to determine the median mutation rate and 95% confidence intervals using FALCOR fluctuation calculator (Lea-Coulson method of the median) as has been previously described (Lea and Coulson 1949; Foster 2006; Schmidt et al. 2006). All statistical analyses were performed in Prism version 6.0h. Differences in mutation rates were evaluated by using the non-parametric Mann–Whitney *U* test to determine statistical significance. A *p*- value of less than 0.05 was considered significant.

Phenotypic characterization

Cells from all Lys+ clones were plated on the following media types for growth analysis: YPD, YPD plus Hygromycin (0.4 mg/ml), YPD plus Geneticin (G418) Sulfate (0.4 mg/mL), YPD plus Nourseothricin (Nat) (0.2 mg/mL), YPGE and drop-out plates lacking Tryptophan or Lysine or Uracil.

Copy Number Variation assay

To qualitatively verify amplification of the copy number *SFA1-CUP1* cassette, cells from each of the aforementioned Lys+ clones, in addition to the Lys- parent (carrying 1 copy of the

cassette) as a negative control to verify viability, were lightly patched to a numbered grid on YPD plates. The cells were allowed to grow for 24 hours. The cell patches were then replica plated to separate plates of complete drop-out medium containing the following concentrations of copper and formaldehyde: 0 µM CuSO4 and 0 mM FA (control),

75 µM CuSO4 and 0.75 mM FA,

150 µM CuSO4 and 1.5 mM FA,

225 µM CuSO4 and 1.6 mM FA,

300 µM CuSO4 and 1.7 mM FA

Cultures were grown for 2 to 4 days. Viability was checked each day visually by naked eye and under a microscope. Resistant clones carrying two or more copies of the *SFA1-CUP1* reporter grew prolifically even in high concentrations of Cu and FA and were easily identifiable. Cell patches from Cu-FA sensitive clones showed weak, filamentous colony morphology and individual cells in the periphery of the colony showed minimal to absent cell division over time.

References

Aggarwal M, Brosh RM. 2012. Functional analyses of human DNA repair proteins important for aging and genomic stability using yeast genetics. *DNA Repair* **11**: 335-348.

Aguilera A. 2009. Chromosomal translocations caused by either pol32-dependent or pol32-independent triparental break-induced replication. *MolecCell Biol* **29**: 5441-5454.

Anand RP, Tsaponina O, Greenwell PW, Lee CS, Du W, Petes TD, Haber JE. 2014. Chromosome rearrangements via template switching between diverged repeated sequences. *Gene Dev* **28**: 2394-2406.

Aplan PD. 2006. Causes of oncogenic chromosomal translocation. *Trends Genet* **22**: 46-55.

Argueso JL, Westmoreland J, Mieczkowski PA, Gawel M, Petes TD, Resnick MA. 2008. Double-strand breaks associated with repetitive DNA can reshape the genome. *P Natl Acad Sci USA* **105**: 11845-11850.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith AJ, Struhl K. 1998. *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.

Bell S, Putnam CD, Kolodner RD. 2014. Template homology determines the genetics and mechanisms of gross chromosomal rearrangements in S. cerevisiae *The FASEB Journal* **28**.

Bochman ML, Paeschke K, Zakian VA. 2012. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet* **13**: 770-780.

Bose P, Hermetz KE, Conneely KN, Rudd MK. 2014. Tandem Repeats and G-Rich Sequences Are Enriched at Human CNV Breakpoints. *Plos One* **9**.

Capra JA, Paeschke K, Singh M, Zakian VA. 2010. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS computational biology* **6**: e1000861.

Carvalho CM, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nature reviews Genetics* **17**: 224-238.

Chen C, Kolodner RD. 1999. Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants. *Nature genetics* **23**: 81-85.

Chin L, Hahn WC, Getz G, Meyerson M. 2012. Making sense of cancer genomic data (vol 25, pg 534, 2011). *Gene Dev* **26**: 1003-1003.

Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. 2013. Emerging landscape of oncogenic signatures across human cancers. *Nature Genetics* **45**: 1127-U1247.

Conover H, Arugeso J, Kunkel T. 2015. Stimulation of Chromosomal Rearrangements by Ribonucleotides. *Genetics*: 951-961.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V et al. 2014. A copy number variation morbidity map of developmental delay (vol 43, pg 838, 2011). *Nature Genetics* **46**: 1040-1040.

Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, Mathur R, Chabes A, Malkova A. 2011a. Break-Induced Replication Is Highly Inaccurate. *PLoS biology* **9**.

Donnianni RA, Symington LS. 2013. Break-induced replication occurs by conservative DNA synthesis. *P Natl Acad Sci USA* **110**: 13475-13480.

Donnianni RA, Symington LS. 2016. *The initiation of DNA replication in eukaryotes*. Springer Science+Business Media, New York, NY.

Dunnick W, Hertz GZ, Scappino L, Gritzmacher C. 1993. DNA-Sequences at Immunoglobulin Switch Region Recombination Sites (Vol 21, Pg 365, 1993). *Nucleic Acids Res* **21**: 2285-2285.

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS et al. 2014. The Reference Genome Sequence of Saccharomyces cerevisiae: Then and Now. *G3-Genes Genom Genet* **4**: 389-398.

Foster PL. 2006. Methods for determining spontaneous mutation rates. *Method Enzymol* **409**: 195-213.

Gietz D, Woods RA. 1998. Transformation of yeast by the lithium acetate single-stranded carrier DNA/PEG method. *Method Microbiol* **26**: 53-66.

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

Gu W, Lupski JR. 2008. CNV and nervous system diseases - what's new? *Cytogenet Genome Res* **123**: 54-64.

Gu W, Zhang F, Lupski JR. 2008. Mechanisms for human genomic rearrangements. *Pathogenetics* **1**: 4.

Hastings PJ, Lupski JR, Rosenberg SM, Ira G. 2009. Mechanisms of change in gene copy number. *Nature Reviews Genetics* **10**: 551-564.

Hermetz KE, Newman S, Conneely KN, Martin CL, Ballif BC, Shaffer LG, Cody JD, Rudd MK. 2014. Large inverted duplications in the human genome form via a fold-back mechanism. *PLoS Genet* **10**: e1004139.

Ionita-Laza I, Rogers AJ, Lange C, Raby BA, Lee C. 2009. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics* **93**: 22-26.

Jackson SP. 2002. Sensing and repairing DNA double-strand breaks - Commentary. *Carcinogenesis* **23**: 687-696.

Kachroo AH, Laurent JM, Yellman CM, Meyer AG, Wilke CO, Marcotte EM. 2015. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* **348**: 921-925.

Katapadi VK, Nambiar M, Raghavan SC. 2012. Potential G-quadruplex formation at breakpoint regions of chromosomal translocations in cancer may explain their fragility. *Genomics* **100**: 72-80.

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

Kim N, Jinks-Robertson S. 2011. Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast. *DNA Repair* **10**: 953-960.

Kolodner RD. 2000. Links between replication, recombination and genome instability in eukaryotes. *Trends in biochemical sciences* **25**: 196-200.

Kraus E, Leung WY, Haber JE. 2001. Break-induced replication: A review and an example in budding yeast. *P Natl Acad Sci USA* **98**: 8255-8262.

Krepischi ACV, Achatz MIW, Santos EMM, Costa SS, Lisboa BCG, Brentani H, Santos TM, Goncalves A, Nobrega AF, Pearson PL et al. 2012. Germline DNA copy number variation in familial and early-onset breast cancer. *Breast Cancer Res* **14**.

Lea DE, Coulson CA. 1949. The Distribution of the Numbers of Mutants in Bacterial Populations. *Journal of Genetics* **49**: 264-285.

Lemmens B, van Schendel R, Tijsterman M. 2015a. Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nat Commun* **6**.

Llorente B, Smith CE, Symington LS. 2008a. Break-induced replication - What is it and what is it for? *Cell Cycle* **7**: 859-864.

Maizels N. 2006. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol* **13**: 1055-1059.

Malkova A, Haber JE. 2012. Mutations Arising During Repair of Chromosome Breaks. *Annual Review of Genetics, Vol 46* **46**: 455-473.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y et al. 2008. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* **82**: 477-488.

Mishra S, Whetstine JR. 2016. Different Facets of Copy Number Changes: Permanent, Transient, and Adaptive. *Mol Cell Biol* **36**: 1050-1063.

Morales ME, White TB, Streva VA, DeFreece CB, Hedges DJ, Deininger PL. 2015. The Contribution of Alu Elements to Mutagenic DNA Double-Strand Break Repair. *Plos Genetics* **11**.

Murat P, Balasubramanian S. 2014. Existence and consequences of G-quadruplex structures in DNA. *Curr Opin Genet Dev* **25**: 22-29.

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

O'Driscoll M, Jeggo PA. 2006. The role of double-strand break repair - insights from human genetics. *Nature reviews Genetics* **7**: 45-54.

Onishi-Seebacher M, Korbel J. 2011. Challenges in studying genomic structural variant formation mechanisms: The short-read dilemma and beyond. *BioEssays* **33**: 840–850.

Paeschke K, Bochman ML, Garcia PD, Cejka P, Friedman KL, Kowalczykowski SC, Zakian VA. 2013. Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497**: 458-462.

Park RW, Kim TM, Kasif S, Park PJ. 2015. Identification of rare germline copy number variations over-represented in five human cancer types. *Mol Cancer* **14**.

Payen C, Koszul R, Dujon B, Fischer G. 2008a. Segmental Duplications Arise from Pol32-Dependent Repair of Broken Forks through Two Alternative Replication-Based Mechanisms. *Plos Genetics* **4**.

Preston BD, Albertson TM, Herr AJ. 2010. DNA replication fidelity and cancer. *Semin Cancer Biol* **20**: 281-293.

Putnam CD, Hayes TK, Kolodner RD. 2009. Specific pathways prevent duplication-mediated genome rearrangements. *Nature* **460**: 984-989.

Putnam CD, Pallis K, Hayes TK, Kolodner RD. 2014. DNA Repair Pathway Selection Caused by Defects in TEL1, SAE2, and De Novo Telomere Addition Generates Specific Chromosomal Rearrangement Signatures. *Plos Genetics* **10**.

Putnam CD, Srivatsan A, Nene RV, Martinez SL, Clotfelter SP, Bell SN, Somach SB, J ESdS, Fonseca AF, de Souza SJ et al. 2016. A genetic network that suppresses genome rearrangements in Saccharomyces cerevisiae and contains defects in cancers. *Nature communications* **7**: 11256.

Saini N, Ramakrishnan S, Elango R, Ayyar S, Zhang Y, Deem A, Ira G, Haber JE, Lobachev KS, Malkova A. 2013. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502**: 389-392.

Sakofsky CJ, Ayyar S, Malkova A. 2012. Break-induced replication and genome stability. *Biomolecules* **2**: 483-504.

Sakofsky CJ, Roberts SA, Malc E, Mieczkowski PA, Resnick MA, Gordenin DA, Malkova A. 2014. Break-Induced Replication Is a Source of Mutation Clusters Underlying Kataegis. *Cell Rep* **7**: 1640-1648.

Schmidt KH, Pennaneach V, Putnam CD, Kolodner RD. 2006. Analysis of gross-chromosomal rearrangements in Saccharomyces cerevisiae. *Method Enzymol* **409**: 462-476.

Smith CE, Llorente B, Symington LS. 2007. Template switching during break-induced replication. *Nature* **447**: 102-105.

Stafa A, Donnianni RA, Timashev LA, Lam AF, Symington LS. 2014. Template Switching During Break-Induced Replication Is Promoted by the Mph1 Helicase in Saccharomyces cerevisiae. *Genetics* **196**: 1017-+.

Startek M, Szafranski P, Gambin T, Campbell IM, Hixson P, Shaw CA, Stankiewicz P, Gambin A. 2015a. Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination. *Nucleic Acids Res* **43**: 2188-2198.

Symington LS, Rothstein R, Lisby M. 2014. Mechanisms and Regulation of Mitotic Recombination in Saccharomyces cerevisiae. *Genetics* **198**: 795-835.

Tornaletti S, Park-Snyder S, Hanawalt PC. 2008. G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. *J Biol Chem* **283**: 12756-12762.

Trela M, Nelson PN, Rylance PB. 2016. The role of molecular mimicry and other factors in the association of Human Endogenous Retroviruses and autoimmunity. *Apmis* **124**: 88-104.

VanHulle K, Lemoine FJ, Narayanan V, Downing B, Hull K, McCullough C, Bellinger M, Lobachev K, Petes TD, Malkova A. 2007. Inverted DNA repeats channel repair of distant double-strand breaks into chromatid fusions and chromosomal rearrangements. *Mol Cell Biol* **27**: 2601-2614.

Vasan S, Deem A, Ramakrishnan S, Argueso JL, Malkova A. 2014. Cascades of Genetic Instability Resulting from Compromised Break-Induced Replication. *Plos Genetics* **10**.

Verma P, Greenberg RA. 2016. Noncanonical views of homology-directed DNA repair. *Genes & development* **30**: 1138-1154.

Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM, Network CGAR. 2013. The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics* **45**: 1113-1120.

Willis TG, Dyer MJS. 2000. The role of immunoglobulin translocations in the pathogenesis of B-cell malignancies. *Blood* **96**: 808-822.

Wilson MA, Kwon Y, Xu Y, Chung WH, Chi P, Niu H, Mayle R, Chen X, Malkova A, Sung P et al. 2013. Pif1 helicase and Poldelta promote recombination-coupled DNA synthesis via bubble migration. *Nature* **502**: 393-396.

Yadav P, Harcy V, Argueso JL, Dominska M, Jinks-Robertson S, Kim N. 2014. Topoisomerase I Plays a Critical Role in Suppressing Genome Instability at a Highly Transcribed G-Quadruplex-Forming Sequence. *Plos Genetics* **10**.

Yu VPCC, Koehler M, Steinlein C, Schmid M, Hanakahi LA, van Gool AJ, West SC, Venkitaraman AR. 2000. Gross chromosomal rearrangements and genetic exchange between nonhomologous chromosomes following BRCA2 inactivation. *Gene Dev* **14**.

Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang CZ, Wala J, Mermel CH et al. 2013. Pan-cancer patterns of somatic copy number alteration. *Nature Genetics* **45**: 1134-U1257.

Zarrei M, MacDonald JR, Merico D, Scherer SW. 2015. A copy number variation map of the human genome. *Nature Reviews Genetics* **16**: 172-183.

Zhang HS, Zeidler AFB, Song W, Puccia CM, Malc E, Greenwell PW, Mieczkowski PA, Petes TD, Argueso JL. 2013. Gene Copy-Number Variation in Haploid and Diploid Strains of the Yeast Saccharomyces cerevisiae. *Genetics* **193**: 785-801.