DISSERTATION

GENOME INSTABILITY: A PRE-EXISTING CONDITION

Submitted by

Hailey Nicole Conover Sedam

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 2018

Doctoral Committee:

Advisor: Lucas Argueso

Jennifer DeLuca
Jac Nickoloff
Claudia Wiese

ABSTRACT

GENOME INSTABILITY: A PRE-EXISTING CONDITION

Copy number variations (CNV), or large amplifications or deletions in the genome, account for about 50% of human genetic diversity. CNVs across genomic regions essential for development and function can lead to disease. The underlying mechanisms of CNV formation are typically traced to a combination of endogenous or environmental sources of DNA damage coupled with defects in DNA repair, replication, and recombination. This dissertation describes two endogenous sources of genome instability involved in both mitotic and meiotic CNVs. Each chapter of this dissertation focuses on one endogenous contribution to genome instability, using the budding yeast *Saccharomyces cerevisiae* as a model system to investigate the conserved cellular processes that, when gone awry, can lead to CNVs.

In the first phase of my research, I focused on the mitotic mutagenic effects of ribonucleotide incorporation into DNA. In the absence of RNase H2, RNA-DNA hybrids (R-loops) accumulate in the genome and ribonucleotides that are misincorporated into the DNA are not efficiently excised. Instead, the latter function is taken over by topoisomerase 1 which inappropriately removes ribonucleotides in a way that leads to accidental/unscheduled DNA double strand breaks (DSBs). My work showed that the accumulation of these lesions in RNase H deficient mutants was sufficient to increase the rate of genome rearrangements through both Loss of Heterozygosity (LOH) and Non Allelic Homologous Recombination (NAHR). Modulating the number of ribonucleotides incorporated into the leading DNA strand during replication through the use of DNA polymerase ε mutants affected the rate of LOH and NAHR. Additionally, the RNase H2-Ribonucleotide Excision Repair Deficient (RNase H2-RED) separation of function allele allowed further investigation of genomic instability when R-loops are properly processed but misincorporated ribonucleotides are not. The RNaseH2-RED study revealed that both ribonucleotide excision repair and R-loop removal contribute roughly equally to chromosomal stability under normal

conditions. Together, the results of these studies indicated that the effects of ribonucleotides and R-loops on chromosomal instability may vary under different genomic contexts of variable R-loop formation and ribonucleotide density.

Next, I designed, constructed, optimized and validated a new yeast assay system to study meiotic NAHR leading to *de novo* recurrent CNVs. The chromosomal rearrangements analyzed through this system are directly analogous to human pathogenic CNVs that are formed in germ cells through recombination between Low Copy Repeat elements (LCRs). While there are assays available to investigate factors involved in mitotic CNV formation, few assays have been developed to experimentally test factors involved in meiotic recurrent CNVs. Previous studies of human patient cohorts have shown that the size and distance between LCRs is strongly correlated with the frequency of recurrent CNV formation. We used this basic observation to validate our experimental system and ask whether it could faithfully recapitulate the phenomenon in our yeast model system. I constructed four diploid strains containing LCRs engineered to range in size from 5-35 Kb and determined the meiotic NAHR frequency in each construct. We detected a very clear linear correlation between LCR size and CNV frequency, and thus established our system as a pertinent assay for interrogation of factors involved in meiotic recurrent CNV formation.

The results described within this dissertation have deepened our understanding of the endogenous causes of genome instability leading to CNVs, and provide perspective into the ability of normal cellular processes to trigger both mitotic and meiotic CNV formation. Additionally, I describe a unique method for future screens of both endogenous and exogenous stimulants of meiotic CNV.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Lucas Argueso. I had just started in Dr. Argueso's lab when a special model organism issue of *Trends in Genetics* published an editorial which concluded with the statement, “If physicists stand on the shoulders of giants, geneticists stand on the shoulders of yeast, and flies, and worms, and plants.” (MACRAE 2014). As a genetics graduate student I have certainly stood on the shoulders of yeast, but more importantly I am grateful for the times I've stood on the shoulders of Dr. Argueso. He has been my dedicated mentor since the day we met, and his wisdom and guidance will remain an inspiration for the rest of my scientific career.

My thanks also go to the members of my graduate advisory committee, Dr. Claudia Wiese, Dr. Jennifer DeLuca, and Dr. Jac Nickoloff, for providing such positive encouragement and direction. I am grateful for your insight and suggestions throughout my projects. Thank you for your support, your guidance, and your time.

The members of the Argueso laboratory filled my PhD with laughter, joy, and baked goods. They exemplify what it means to work as a team, and it was a wonderful team to be a part of. I specifically want to thank Nadia Sampaio who over the years has played the roles of best friend, therapist, and mentor. Her constant support and guidance has kept me encouraged and motivated.

I am grateful for the education provided to me by the Cell and Molecular Biology (CMB) graduate program. When I joined the CMB program, I did not expect to gain so many life-long friendships. I have met some of the world’s brightest, funniest, and kindest people through this graduate program. Thank you all for challenging me, for supporting me, and for helping me see the world through your eyes. To my cohort; Adam, Alli, Josh, Kate, Laird, Laylaa, and Tom — I have cherished our time together. To the friends I’ve made because of CMB; Aliena, Annie, Hannah, Katie, Talia, PJ — I am so lucky that everything in our lives led us to each other. To the CMB Directors, Dr. Howard Liber and Dr. Carol Wilusz, I have thoroughly enjoyed our talks about science, and about life — thank you for your counsel.

My life in Fort Collins was surrounded by a community of scientists. It was a truly amazing and humbling experience to know that every event I attended, whether social or scientific in nature, would provide opportunities for learning. The passion displayed by my PhD colleagues, the veterinary students, and the rest of the scientific community will forever remind me why I joined this field. My friends here have become more like family, thank you all for making Fort Collins my home.

Finally, I would like to whole-heartedly express my gratitude to my family. My parents, Kris and Kurt, for a lifetime of encouragement, education, and love. It's said that we tend to become like those we admire, and I strive to be more like the both of you. My sister, Shannon, who paved the path towards graduate school, and always put extra effort into travelling to visit me, reducing my stress, and ensuring I was well fed and warm through the Colorado winters. And to my husband, Tyler, who inspired me to apply to graduate school in the first place, and continues to inspire me every day. Each of you motivates me to be better by being more like you, thank you for everything.

The work in this dissertation was made possible by funding from the NIH - 1R35GM119788-01, the Boettcher Foundation, and the Escher Fund for Autism Research. I am especially grateful for scientific training opportunities through the Environmental Mutagenesis and Genomics Society's travel award and the Colorado State University Vice President for Research Graduate Fellows Program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One

Chapter Two

Chapter Three

Chapter Four

Supplemental

LIST OF ACRONYMS

aCGH- Array comparative genomic hybridization

Alt-NHEJ- Alternative non-homologous end joining

CI- Confidence Interval

CNV- Copy Number Variation

CN- Copy Number

ddPCR- Digital Droplet Polymerase Chain Reaction

DP LCR- Directly Paralogous Low Copy Repeat

DSB- Double Strand Break

EMGS- Environmental and Molecular Mutagenesis Society

EMM- Environmental and Molecular Mutagenesis

GCR- Gross Chromosomal Rearrangement

HU- Hydroxyurea

IR- Ionizing Radiation

LOH- Loss of Heterozygosity

LCR- Low Copy Repeat

MMR- Mismatch Repair

MMBIR- Microhomology-mediated break-induced replication

NAHR- Non Allelic Homologous Recombination

PCR- Polymerase chain reaction

PTLS- Potocki-Lupski Syndrome

RED- Ribonucleotide Excision Repair Deficient, Separation of function RNAseH2 allele

RER- Ribonucleotide Excision Repair

SMS- Smith-Magenis Syndrome

YAC- Yeast artificial chromosome

CHAPTER ONE: CONTRASTING MECHANISMS OF DE NOVO COPY NUMBER MUTAGENESIS SUGGEST THE EXISTENCE OF DIFFERENT CLASSES OF ENVIRONMENTAL COPY NUMBER MUTAGENS[1]

Summary

While gene copy number variations (CNVs) are abundant in the human genome, and often are associated with disease consequences, the mutagenic pathways and environmental exposures that cause these large structural mutations are understudied relative to conventional nucleotide substitutions in DNA. The members of the environmental mutagenesis community are currently seeking to remedy this deficiency, and there is a renewed interest in the development of mutagenicity assays to identify and characterize compounds that may induce *de novo* CNVs in humans. To achieve this goal, it is critically important to acknowledge that CNVs exist in two very distinct classes: Non-recurrent and recurrent CNVs. The goal of this introduction is to emphasize the deep contrasts that exist between the proposed pathways that lead to these two mutation classes. Non-recurrent *de novo* CNVs originate primarily in mitotic cells through replication-dependent DNA repair pathways that involve micro-homologies (<10 bp), and are detected at higher frequency in children of older fathers. In contrast, recurrent *de novo* CNVs are most often formed in meiotic cells through homologous recombination between non-allelic large low copy repeats (>10,000 bp), without an associated paternal age effect. Given the biological differences between the two CNV classes, it is our belief that non-recurrent and recurrent CN mutagens will probably

[1] *This chapter is an adaptation of previously published material, the figures have been renumbered to indicate both chapter and figure number.*

Reference for the full article:

Conover, H. N., Argueso, J. L. 2016. Contrasting mechanisms of de novo copy number mutagenesis suggest the existence of different classes of environmental copy number mutagens. Environmental and Molecular Mutagenesis. DOI: 10.1002/em.21967.

differ substantially in their modes of action. Therefore, each CNV class may require their own uniquely designed assays so that we as a field may succeed in uncovering the broadest possible spectrum of environmental CN mutagens.

Introduction

Beginning in the early 1980s, evidence started accumulating to establish a link between genome rearrangements leading to gene copy number variation (CNV) and a wide variety of human disorders. By the late 1990s, Dr. James R. Lupski coined the term "genomic disorders" to describe inherited conditions caused by structural changes in chromosomes, rather than conventional nucleotide substitutions in the linear sequence of DNA (LUPSKI 1998). However, it wasn't until the discovery that *de novo* CNVs are strongly associated with autism (SEBAT *et al.* 2007), and subsequently, the understanding that CNVs account for up to 25% of all severe intellectual disability cases (GIRIRAJAN and EICHLER 2010; SCHAAF *et al.* 2011; BATTAGLIA *et al.* 2013; BEAUDET 2014), that the urgency of studying the mechanisms of CNV formation decisively grabbed the attention of the environmental mutagenesis field. The members of the Environmental Mutagenesis and Genomics Society (EMGS) have become increasingly interested in the mutational mechanisms that lead to CNVs. The disease consequences of CNVs and the pathways involved in copy number mutagenesis have become prominent topics at our annual meetings, most recently in 2014 when Dr. Lupski received the EMGS Award. Both the Environmentally Induced Germline Mutation Analysis Workshop (ENIGMA - Montreal 2011; (YAUK *et al.* 2013)) and the International Workshop on Genotoxicity Testing (IWGT - Foz do Iguaçu 2013; (YAUK *et al.* 2015b)) specifically discussed the potential relationship between environmental exposures and CNV formation. Finally, Environmental and Molecular Mutagenesis (EMM) recently published two comprehensive review articles that described copy number mutagenesis mechanisms in rich detail (LUPSKI 2015a; RUSSO *et al.* 2015).

In his EMM review article, Dr. Lupski posed three key questions that encapsulate the challenges that the environmental mutagenesis community should strive to overcome in the coming years: *"Do*

current mutagenesis assays measure CNV formation?; Can we design such an assay?; Are we introducing compounds into our environment that induce CNV mutagenesis?". In our opinion, the answers to these questions are, respectively: Partly; Yes; and Probably. The goal of this commentary is to highlight the biological differences that exist between the two major classes of *de novo* CNVs associated with human genomic disorders: Non-recurrent and recurrent CNVs. As discussed below, we believe that, as a consequence of these fundamental differences, assays designed to measure non-recurrent CNVs will not be suitable for the identification of recurrent CN mutagens, and vice-versa. Understanding and acknowledging these differences will be the first step to design effective approaches for the future identification of the majority of environmental mutagens that can induce the formation of both classes of CNVs.

Two CNV classes: Differences in breakpoints, mutagenesis mechanisms, and cell types

The two CNV classes are defined by their differences in genome rearrangement breakpoint positions. When analyzing the genomes of a group of unrelated individuals who share a disorder caused by gene duplication or deletion, the specific CNV breakpoints may occur at different (non-recurrent) or similar (recurrent) positions. All affected individuals in this group have altered copy number spanning the same relevant dose-sensitive gene, however, the boundaries and size of their CNV regions may vary. This concept is illustrated in Figure 1.1, and other key associated differences between the two CNV classes are summarized in Table 1.1.

Non-recurrent CNVs are thought to originate primarily when micro-homology mediated pathways are inappropriately used to repair DNA double-strand breaks (DSBs). These include microhomology-mediated break-induced replication (MMBIR) and alternative non-homologous end joining (Alt-NHEJ) (HASTINGS *et al.* 2009; BETERMIER *et al.* 2014). Breakpoints resembling a blunt-end ligation between sequences that do not share any homology are also observed, but their formation mechanism is not yet clear since an essential component of the canonical-NHEJ pathway (Xrcc4) was shown to be dispensable for zero-homology breakpoint formation in mouse embryonic stem cells (ARLT

et al. 2012a). While micro- and non-homologous breakpoints have been well documented, recent findings also show that a substantial fraction of human non-recurrent CNVs result from non-allelic homologous recombination (NAHR) between short dispersed DNA repeats (*e.g.*, LINE, Alu, HERV mobile element insertions [MEIs]) (GIRIRAJAN *et al.* 2013; ROBBERECHT *et al.* 2013; BOONE *et al.* 2014; BOSE *et al.* 2014; CAMPBELL *et al.* 2014; CARVALHO *et al.* 2014; HERMETZ *et al.* 2014; NEWMAN *et al.* 2015; STARTEK *et al.* 2015; WECKSELBLATT *et al.* 2015).

Non-recurrent *de novo* CNVs are believed to form in mitotic pre-meiotic germ cells, or through post-zygotic somatic mosaicism (CAMPBELL *et al.* 2015) that may contribute to the germline. Regardless of the amount of sequence homology involved in their formation, non-recurrent CNVs arise in proliferating cells primarily as a consequence of direct DNA damage or DNA replication stress that leads to the formation of accidental DSBs (ARLT *et al.* 2012b). The replication stress connection, in particular, has been conclusively demonstrated by joint work from the Glover and Wilson laboratories, both following environmental exposure (ARLT *et al.* 2009; ARLT *et al.* 2011a; ARLT *et al.* 2011b; ARLT *et al.* 2014) and at genomic regions such as common fragile sites and active large transcription units where replication forks are prone to slow down and collapse (WILSON *et al.* 2015). Corroborating their conclusions, a significant correlation was observed between computationally predicted regions of slow DNA replication and the incidence of non-recurrent CNVs in the *Drosophila*, mouse and human genomes (CHEN *et al.* 2015).

In contrast, recurrent *de novo* CNVs are thought to form primarily during the meiotic cell cycle, specifically the MI prophase, through the NAHR mechanism. The breakpoints in recurrent CNVs occur within directly-oriented paralogous low copy repeats (DP-LCRs) which are thought to be the remnants of recent segmental duplications in the human lineage (BAILEY *et al.* 2002). They are at least 10 Kb in size (often larger) and share more than 95% nucleotide sequence identity. Although DP-LCR sequences are present in low copy numbers (often just two), because each individual repeat unit is long, altogether they represent about 5% of the human genome (MEFFORD and EICHLER 2009; LIU *et al.* 2012). The presence of these long homologous sequences in relative close proximity to each other poses a challenge to the

recombination machinery, which must accurately discern between the correct (allelic) and incorrect (non-allelic) template substrates to repair the programmed meiotic DSBs produced by the SPO11 enzyme (SASAKI *et al.* 2010). This vulnerability is further exacerbated when the DP-LCRs contain binding sequences for the PRDM9 histone methyltransferase, which modifies the local chromatin structure, recruiting SPO11 and thus creating meiotic recombination hotspots (DITTWALD *et al.* 2013; PRATTO *et al.* 2014). In addition to the local genomic architecture (*i.e.*, number, distance, and size of DP-LCRs), evidence from classic studies in yeast and recent analyses of large human CNV databases strongly suggest that inaccurate meiotic chromosome synapsis and/or defective meiotic checkpoints facilitate NAHR between DP-LCRs, thus leading to recurrent CNV formation (PETES and HILL 1988; HABER *et al.* 1991; GOLDMAN and LICHTEN 1996; GRUSHCOW *et al.* 1999; THOMPSON and STAHL 1999; GOLDMAN and LICHTEN 2000; LIU *et al.* 2011b; DITTWALD *et al.* 2013; SHINOHARA and SHINOHARA 2013; PENG *et al.* 2015).

The classification system described above, along with the corresponding CNV properties summarized in Table 1.1, are adequate to describe the majority of *de novo* CNVs observed in humans. However, exceptions to these general rules do exist. (1) Short repeat (*i.e.*, MEIs) -mediated mitotic NAHR events have the potential to be recurrent, and in fact sometimes are (HERMETZ *et al.* 2012; SHUVARIKOV *et al.* 2013); and (2), MEI-NAHR is not necessarily restricted to mitotic cells, as recombination between short repeats may also occasionally occur during meiosis in humans (ROBBERECHT *et al.* 2013). However, because the MEI sequences in question are highly abundant (STEWART *et al.* 2011), and some of them are still actively moving in the human genome (BURNS and BOEKE 2012), it is unusual for CNV breakpoints to involve the exact same pair of MEI repeat units in unrelated individuals. For these reasons, we believe that it is more appropriate to place them in the non-recurrent CNV class, rather than in the recurrent class. Another noteworthy exception (3) to the properties listed in Table 1.1 regards cases in which DP-LCR mediated CNVs are detected as somatic mosaics (SHINAWI *et al.* 2010; DITTWALD *et al.* 2013), which implies that they must also form at some

appreciable frequency during the mitotic cell cycle. This subclass of mitotic recurrent CNVs might be triggered by perturbation of the same biological processes (*e.g.*, DNA replication) that typically induce non-recurrent CNVs during mitosis.

Differences in paternal age effect

Genome-wide analyses of family trios and quartets have established a strong correlation between advanced paternal age and increased frequency of *de novo* mutations in their offspring (KONG *et al.* 2012; CAMPBELL and EICHLER 2013; SEGUREL *et al.* 2014). Importantly, while this correlation holds true for nucleotide mutations and non-recurrent CNVs, no association with the father's age has been found for recurrent CNVs (HEHIR-KWA *et al.* 2011; SIBBONS *et al.* 2012; MACARTHUR *et al.* 2014). This is likely due to the distinct cell types where to the two different classes of CNVs arise. Mitotic mutations (including non-recurrent CNVs) accumulate in male germ cells as the number of successive stem cell divisions increases with age. In contrast, recurrent CNVs do not have the opportunity to build up because only one meiotic cell cycle occurs between generations, regardless of the parent's age.

This key difference was evident in a recent elegant study which directly measured rates of NAHR at the *CMT1A-REP* locus in sperm from donors that included monozygotic twins (MACARTHUR *et al.* 2014). The authors found that the rate of recurrent *CMT1A-REP* deletions was variable between unrelated men, and was independent of the sperm donor's age. Interestingly, the NAHR rate was strongly correlated between identical twin brothers. Therefore, a combination of genes and environment likely influences the rate of recurrent CNV formation in humans, although the limited size of the twin sperm donor cohort available for sampling precluded the quantification of the specific contribution from each. Despite this, the results reported by MacArthur et al. strongly suggest that the rate of meiotic NAHR at a specific locus is a movable quantity, which would have profound implications for future mutagenesis studies. If genes and environmental exposures do indeed play a role in making some parents transmit more recurrent CNVs to their children than other parents, then those specific factors must be identified and their biological activity characterized.

Likely differences in environmental copy number mutagens

Environmental mutagens, including agents that cause DNA damage and/or DNA replication stress, are known to play a role in the genome instability processes associated with mitotic non-recurrent CNVs in cancer (BOFFETTA 2006; RELIENE *et al.* 2007; ARLT *et al.* 2012b; CIRIELLO *et al.* 2013). It is likely that the same carcinogens that trigger chromosomal rearrangements may also cause non-recurrent CNVs to accumulate in mitotic germ cells. The existence of environmental mutagens that can induce mitotic non-recurrent CNVs leads us to expect that mutagens that cause meiotic recurrent CNVs may also exist. Even though no agent with such specific activity has been identified to date, given the major mechanistic differences between the two CNV classes, we expect that there will be hardly any overlap between the eventual lists of non-recurrent and recurrent CN mutagens.

The common feature of mitotic non-recurrent CN mutagens is that they directly or indirectly cause accidental/unscheduled DSBs leading to genome rearrangements. In contrast, during meiosis DSBs are carefully programed to occur at a specific time and their repair is highly coordinated to produce "healthy" allelic crossovers (KEENEY *et al.* 2014; ZICKLER and KLECKNER 2015). SPO11 produces an excess of DSBs, many more than those that eventually are resolved as allelic or even as "unhealthy" non-allelic crossovers (*i.e.*, recurrent CNVs). Thus, an environmental mutagen that causes a relatively small number of additional DSBs would not likely lead to an increase in the overall NAHR rate, especially since these extra DSBs would have to form precisely during the narrow window of time when meiotic recombination is taking place. Instead, an agent that somehow perturbs the structure of the synaptonemal complex, homologous chromosome engagement, meiotic sister chromatid cohesion, or meiotic checkpoints could potentially alter the normal ratio of allelic to non-allelic recombination, and thus become an effective recurrent CN mutagen. This yet unidentified activity is very different from the known mode of action of mitotic non-recurrent CN mutagens.

Conclusion: New and diverse copy number mutagenesis assays are needed

It has been the consensus in the environmental mutagenesis field that mutagens that cause nucleotide mutations or chromosomal rearrangements in mitotic cells will also be mutagenic in germ cells. This is likely the case for DNA replication-based mutagenic processes in pre-meiotic germ stem cells or even during the round of DNA replication that occurs in the first meiotic division, immediately before recombination. The expert panel that met in the IWGT 2013 generally agreed that somatic mutagenicity predicts germ cell effects, although they warned that caution should be used, particularly in light of the full known spectrum of mutations that cannot be detected by current laboratory tests (YAUK *et al.* 2015b). Our field should heed this warning when searching for agents that can induce the formation of both classes of *de novo* CNVs.

A variety of assays have been developed to identify mitotic CN mutagens in different model systems. These include selection-based assays that can identify CNVs at targeted genomic regions (PETES and HILL 1988; SCHIESTL 1989; ZIMMERMANN 1992; HOFFMANN 1994; MYUNG and KOLODNER 2003; RELIENE and SCHIESTL 2003; PAYEN *et al.* 2008; LYNCH *et al.* 2011; ZHANG *et al.* 2013b) and selection-free genome-wide assays in yeast and mammalian cell culture (ARGUESO *et al.* 2008; ARLT *et al.* 2009; ARLT *et al.* 2011a; ARLT *et al.* 2011b; ARLT *et al.* 2014). To the best of our knowledge, there are currently no validated assays available to specifically detect meiotic recurrent CNVs, although at least one promising prototype is under development (ARGUESO *et al.* 2014). Affordable unbiased whole genome survey approaches are ushering a new era in this field, exemplified by an exciting new report that investigated *de novo* germline mutations induced by ionizing radiation (IR) (ADEWOYE *et al.* 2015). These authors observed a convincing eight-fold increase (relative to control) in *de novo* CNVs in the offspring of male mice following a 3 Gy X-ray exposure, although the breakpoint features of the CNVs were not specified, and thus it was not clear which recurrence class they might have belonged to. Over the next few years, similar approaches will be used to continue expanding the list of known somatic and germ cell CN mutagens. Model organism-based assays that are amenable to high throughput chemical screens will be particularly useful in this regard (FERREIRA and ALLARD 2015).

Through the efforts of our field, evidence has progressively been gathered to support officially declaring IR, air pollution, and cigarette smoke as human germ cell mutagens in the near future (DEMARINI 2012). Much more work, using new methods, will be needed to identify and validate additional agents that have similar activity. However, as we continue to expand the catalog mutagens affecting germ cells, it is critically important for our field to be mindful that the mutagens identified through mitotic assays might only identify part of the risk factors. To reach a comprehensive understanding of our vulnerability to environmental mutagens, we will be required to also develop and apply methods that can interrogate the ability of toxicants to specifically perturb meiotic functions.

Figures

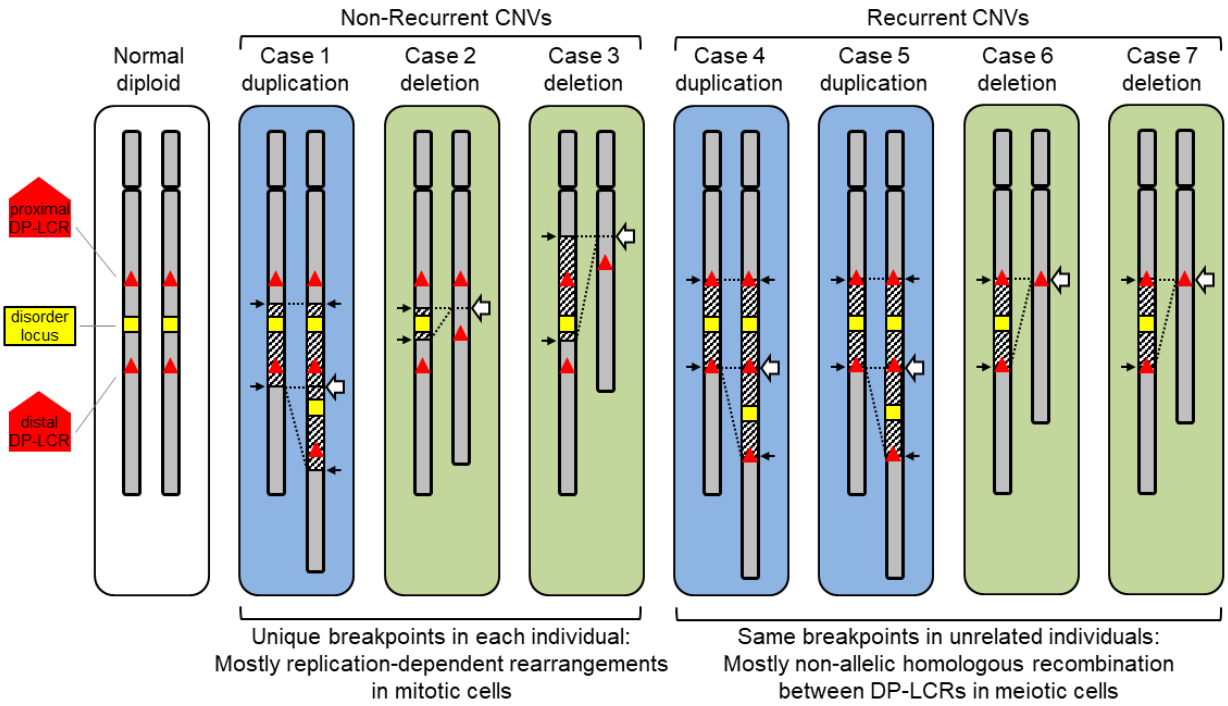


Figure 1.1 *de novo* CNV classes: Non-Recurrent and Recurrent

This figure depicts the two major classes of *de novo* CNVs observed in humans. The white cell (left) displays a karyotype with two normal copies of a chromosome, each containing one copy of a locus (yellow squares) associated with a hypothetical genomic disorder. This locus is flanked by a pair of proximal and distal direct paralogous low copy repeats (DP-LCRs; red triangles). An individual with this normal karyotype would not be affected by the disorder, while individuals with an up or down deviation in copy number of the locus would likely show symptoms. **Cases 1, 4, and 5** represent duplications (blue cells) of a genomic segment that includes the disorder locus resulting in three total copies of the dosage-sensitive gene in the region; **Cases 2, 3, 6, and 7** represent deletions (green cells) resulting in a single copy of that same gene. Arrows indicate the CNV breakpoints, *i.e.* the DNA sequences involved in the chromosomal rearrangements, where a new DNA connection is formed (white block arrow). The breakpoints are used to classify CNVs as non-recurrent or recurrent, and the presence or absence of DNA repeats at the breakpoints is used to infer their molecular mechanism of formation. **Cases 1, 2, and 3** each carry CNVs with unique breakpoints not seen in any other cases, thus are classified as non-recurrent. Such mutations arise most often in mitotic cells through DNA replication-dependent mechanisms. **Cases 4, 5, 6, and 7** have the same CNV boundaries, even though these individuals are not members of the same family. This suggests that their CNVs occurred *de novo*, independently, multiple times, and therefore are recurrent. Recurrent CNVs are typically associated with DP-LCRs at the breakpoints, and are believed to arise through non-allelic homologous recombination in meiotic cells.

Tables

Table 1.1 General properties of *de novo* CNV classes*.

CNV Class:	NON-RECURRENT	RECURRENT
DNA sequence at the rearrangement breakpoints:	Unique to each individual	Same in unrelated individuals
Mutational mechanisms:	Microhomology-mediated BIR, Alt-NHEJ, C-NHEJ, NAHR between *short* dispersed high copy repeats (*e.g.* LINE, Alu, HERV)	NAHR between *long* direct paralogous - low copy repeats (DP-LCRs)
Cell type:	Pre-meiotic mitotic germ cells, Post-zygotic mitotic cells (somatic mosaicism)	Meiotic germ cells (MI prophase)
Paternal age effect:	YES	NO
Known or predicted genetic predispositions:	Mutations in genes involved in DNA damage response and repair; initiation and progression of DNA replication	Presence, architecture, and sequence of DP-LCR recombination substrates; mutations in genes involved in control of the meiotic program
Known or predicted mode of action of environmental mutagens:	Agents that cause DNA damage, DNA replication stress, others?	Agents that may perturb meiotic chromosome synapsis; none identified to date

*Table format modeled after and modified from [Arlt et al., 2012].

CHAPTER TWO: STIMULATION OF CHROMOSOMAL REARRANGEMENTS BY RIBONUCLEOTIDES[2]

Summary

We show by whole genome sequence analysis that loss of RNase H2 activity increases loss-of-heterozygosity (LOH) in *Saccharomyces cerevisiae* diploid strains harboring the *pol2-M644G* allele encoding a mutant version of DNA polymerase ε that increases ribonucleotide incorporation. This led us to analyze the effects of loss of RNase H2 on LOH and on non-allelic homologous recombination (NAHR) in mutant diploid strains with deletions of genes encoding RNase H2 subunits (*rnh201Δ*, *rnh202Δ*, and *rnh203Δ*), topoisomerase 1 (*top1Δ*), and/or carrying mutant alleles of DNA polymerases ε, α, and δ. We observed a ~7-fold elevation of the LOH rate in RNase H2 mutants encoding wild-type DNA polymerases. Strains carrying the *pol2-M644G* allele displayed a 7-fold elevation in the LOH rate, and synergistic 23-fold elevation in combination with *rnh201Δ*. In comparison, strains carrying the *pol2-M644L* mutation that decreases ribonucleotide incorporation displayed lower LOH rates. The LOH rate was not elevated in strains carrying the *pol1-L868M* or *pol3-L612M* alleles that result in increased incorporation of ribonucleotides during DNA synthesis by polymerases α and δ, respectively. A similar trend was observed in an NAHR assay, albeit with smaller phenotypic differentials. The ribonucleotide-

[2] *This chapter is an adaptation of previously published material, the figures have been renumbered to indicate both chapter and figure number.*

Reference for the full article:

Conover, H. N., Lujan, S. A., Chapman, M. J., Cornelio, D. A., Sharif, R., Williams, J. S., Clark, A. B., Camilo, F., Kunkel, T. A., Argueso, J. L. 2015 Stimulation of chromosomal rearrangements by ribonucleotides. Genetics. DOI: 10.1534/genetics.115.181149.

Contributions to this research are as follows:
Whole Genome Sequencing: SAL, JSW, ABC, FC, TAK, LOH Assay: **HNC**, NAHR Assay: DAC, **HNC**, Karyotype Analysis: **HNC**, DAC, JLA, Data Analysis: **HNC**, SAL, DAC, JLA, TAK, Strain Construction: **HNC**, SAL, MJC, RS, JSW, ABC, FC, Manuscript Preparation: JLA, TAK, **HNC**, SAL

mediated increases in the LOH and NAHR rates were strongly dependent on *TOP1*. These data add to recent reports on the asymmetric mutagenicity of ribonucleotides caused by topoisomerase 1 processing of ribonucleotides incorporated during DNA replication.

Introduction

The replicative DNA polymerases of *Saccharomyces cerevisiae*, DNA polymerases α (Pol α), δ (Pol δ) and ε (Pol ε), frequently incorporate ribonucleotides into DNA both *in vitro* and during nuclear DNA replication *in vivo* (NICK MCELHINNY *et al.* 2010a; NICK MCELHINNY *et al.* 2010b; WILLIAMS and KUNKEL 2014; WILLIAMS *et al.* 2015). These ribonucleotides are efficiently removed when RNase H2 incises the DNA backbone containing a ribonucleotide to initiate Ribonucleotide Excision Repair (RER) (NICK MCELHINNY *et al.* 2010a; SPARKS *et al.* 2012). When the *RNH201* gene that encodes the catalytic subunit of RNase H2 (CERRITELLI and CROUCH 2009) is deleted, RER is defective and many unrepaired ribonucleotides remain in the genome. A subset of these unrepaired ribonucleotides can be removed when topoisomerase 1 (Top1) incises a DNA backbone containing a ribonucleotide (WILLIAMS *et al.* 2013). However, Top1 incision creates nicks with unligatable ends and elicits several RNA-DNA damage phenotypes, including slow growth, activation of the genome integrity checkpoint and altered progression through the cell cycle, sensitivity to the replication inhibitor hydroxyurea (HU), and strongly elevated rates for deletion of 2-5 base pairs from low complexity DNA sequences (NICK MCELHINNY *et al.* 2010a; CLARK *et al.* 2011; KIM *et al.* 2011). These effects are elicited primarily by ribonucleotides incorporated by Pol ε, but not by ribonucleotides incorporated by Pol α or Pol δ (WILLIAMS *et al.* 2015). Loss of RNase H2 is also associated with reduced efficiency of mismatch repair (MMR), thereby elevating the rate of single-base mutations (GHODGAONKAR *et al.* 2013; LUJAN *et al.* 2013). This mutator phenotype is consistent with the hypothesis (NICK MCELHINNY *et al.* 2010a) that nicks resulting from RNase H2 incision at ribonucleotides can signal for strand discrimination during removal of DNA replication errors.

In addition to the point mutations mentioned above, larger types of genome instability have also been observed in RNase H2-defective cells. For example, in a study of Gross Chromosomal Rearrangements (GCRs) in haploid yeast cells, RNase H2 defects alone had little effect, but GCR rates were elevated in double mutant strains lacking the non-catalytic Rnh203 subunit in combination with deletions of any of eight other genes affecting DNA metabolism (ALLEN-SOLTERO *et al.* 2014). An earlier GCR study reported that *rnh201Δ* single mutants displayed a four-fold increase in instability of a non-essential Yeast Artificial Chromosome (YAC loss and terminal deletions) (WAHBA *et al.* 2011). This instability may also be related to the fact that defects in the yeast Rnh202 subunit of RNase H2 increase the rate of gene conversion (also in haploids), an effect that is partially suppressed by deleting *TOP1* (AGUILERA and KLEIN 1988; II *et al.* 2011; POTENSKI *et al.* 2014). Similarly, mouse embryonic fibroblasts lacking the non-catalytic RNASEH2B subunit of RNase H2 have increased levels of micronuclei and chromosomal rearrangements (REIJNS *et al.* 2012). The mechanisms responsible for these types of large-scale genome instability are not yet fully understood, but could involve DNA strand breaks arising during processing of unrepaired ribonucleotides incorporated during replication, processing of unresolved R-loops formed during transcription, or both.

The present study was designed to answer three questions. First, do ribonucleotides incorporated during nuclear DNA replication in RER-defective yeast strains elevate the rate of two types of large-scale genome instability in diploid cells: mitotic interhomolog allelic homologous recombination leading to loss-of-heterozygosity (LOH), and non-allelic homologous recombination (NAHR) leading to chromosomal translocations and copy number alterations? If so, do elevated LOH or NAHR rates depend on ribonucleotides incorporated by Pol ε, Pol α, or Pol δ? Thirdly, do elevated LOH or NAHR rates depend on Top1? To answer these questions, we studied a series of homozygous diploid yeast strains that vary in RNase H2 and Top1 status and that also vary in the propensity to incorporate ribonucleotides by Pol ε (*pol2-M644G*, increased; *pol2-M644L*, decreased) or by Pols α and δ (increased in both *pol1-L868M* and *pol3-L612M*, respectively). These strains, and their parents encoding wild-type replicases

were modified to assay for LOH and NAHR. The results indicate that ribonucleotides incorporated by Pol ε cause *TOP1*-dependent chromosome instability, while the role of those incorporated by Pols α and δ is not as strong. The results support a model in which asymmetric processing of ribonucleotides in DNA by Top1 causes both local and large-scale genome destabilization.

Methods and Materials

Yeast strains

The *Saccharomyces cerevisiae* strains used for the whole genome sequencing mutation accumulation experiment were diploids descended from Δ|(-2)|-7B-YUNI300 (PAVLOV *et al.* 2001). They were homozygous for *his7-2, leu2Δ::kanMX, ura3Δ, trp1-289, ade2-1, lys2ΔGG2899-2900* and *agp1::URA3*. Sanger sequencing confirmed that the *pol2-M644G* mutation was homozygous, and *rnh201Δ::hphMX* was confirmed to be homozygous by PCR fragment size analysis. The strains used in the main set of LOH assays and all NAHR assays (presented in Table S2.1) were isogenic to the CG379 strain background (MORRISON *et al.* 1991; ARGUESO *et al.* 2008). We also conducted a limited number of LOH assays using hybrid diploids resulting from the mating of *MATa* haploids isogenic to CG379 and *MATα* haploids isogenic with YJM789 (WEI *et al.* 2007). CG379 is a laboratory strain background closely related to the S288c reference yeast strain, while YJM789 is a diverged background derived from a clinical isolate. There are roughly 60,000 single nucleotide polymorphisms (SNPs) between two strains' genomes, some of which are associated with restriction fragment length polymorphisms (RFLPs) that can be used to monitor recombination between homologous chromosomes. For the isogenic and hybrid LOH assays, the CORE2 cassette containing the *Kluyveromyces lactis URA3* gene, the *S. cerevisiae URA3* gene, and the *kanMX* geneticin resistance gene was amplified from plasmid pJA40 (ZHANG *et al.* 2013a) and integrated at chromosome 7 (Chr7) downstream of the *MAL13* gene (distal side), approximately 20 Kb from the right telomere. For the NAHR assay, we used a PCR-based approach to delete a 180 bp segment spanning the 3' end of the open reading frame (ORF) and the immediate downstream sequence of the *URA3* gene at

its native position on Chr5. This *ura3Δ3'* allele corresponds to a clean (marker-less) deletion between Chr5 coordinates 116,823 and 117,003 from the reference *S. cerevisiae* S288c genome (CHERRY *et al.* 2012). The specific coordinates given are from the release version 64-2-1 of the S288c reference genome. In addition, a cassette containing the *kanMX* marker and a segment of the *URA3* gene missing the promoter sequences and 34 bp from 5' end of the ORF was integrated on Chr14, downstream of the *PEX17* gene (proximal side). This non-allelic *PEX17::kanMX-ura3Δ5'* (*-RA3*) insertion shares 622 bp of perfect homology (*RA* region) with the *ura3Δ3'* allele described above. Both *RA* sequences are oriented 5' to 3' in the top strand of the reference genome (Watson strand), on the left arms of their respective chromosomes, and recombination between them can regenerate a fully functional *URA3* gene that can be selected for by cell growth on uracil drop-out medium.

The *RNH201*, *RNH202*, *RNH203*, and *TOP1* genes were deleted with the *hphMX* or *natMX* drug resistance cassettes using a PCR based approach (GOLDSTEIN and MCCUSKER 1999). The *pol2-M644G* and *pol2-M644L* alleles were integrated through a two-step allele replacement procedure (*URA3* pop-in, pop-out) using plasmids described earlier (KIRCHNER *et al.* 2000). We built custom plasmids to integrate the *pol3-L612M* and *pol1-L868M* alleles, pHC1 and pHC2, respectively. We used overlapping PCR approaches to create mutant *pol3-L612M* and *pol1-L868M* fragments containing synonymous substitutions that created diagnostic *Taq*I restriction enzyme polymorphisms immediately adjacent to the codons encoding the Leu-Met mutations of interest in each gene. These PCR fragments were cloned into the pRS306 *URA3* integrative vector (SIKORSKI and HIETER 1989) and Sanger sequenced to validate the constructions. pHC1 was linearized with *Bam*HI and pHC2 was linearized with *Sal*I prior to integration into *POL3* and *POL1* homozygous diploid strains, respectively. Ura^+ clones containing the respective hemizygous plasmid integrations were patched onto 5-FOA medium to select for plasmid pop-outs. PCR and *Taq*I restriction analysis were used to screen candidate 5-FOA resistant derivatives and identify diploids heterozygous for *POL3*/*pol3-L612M* or *POL1*/*pol1-L868M*, from which haploids carrying the respective DNA polymerase δ and α mutations were obtained by sporulation and tetrad dissection. All

combinations of LOH and NAHR experimental markers and mutant backgrounds were obtained by intercrossing isogenic haploid strains followed by sporulation, tetrad dissection, and genotyping of haploid spores. The DNA sequences of all plasmids and oligonucleotide primers used in the constructions above are available upon request.

Detection of LOH by whole genome sequencing

RNH201-dependent LOH events were monitored during mutation accumulation experiments performed as previously described (LUJAN *et al.* 2014). LOH rates are lower bound estimates calculated from the number of homozygous mutations accumulated in a particular isolate, divided by both the total number of mutations and by the number of accumulated yeast generations for that isolate. *p*-values comparing LOH rates between strains were calculated with a one-tailed Welch's t test (WELCH 1947). The whole genome sequencing data used to identify genome wide LOH events (findings in Table 2.1) have been deposited at the Sequence Read Archive (SRA) database under study number SRP062900; experiments SRX1165947, SRX1165952, SRX1165955, and SRX1165953.

Quantitative recombination assays

Yeast cells were streaked to single colonies on solid YPD medium and incubated on plates at 30° C for three days. For the LOH assays, fresh whole colonies were picked, resuspended in 1 ml of sterile distilled water, serially diluted, and plated on synthetic complete medium containing 1 g/l of 5-FOA (selective) and YPD (permissive). For the NAHR assays, fresh single colonies were picked, inoculated in tubes containing 4 ml of liquid YPD, and incubated for 24 hours at 30° C in a rotating drum. One ml aliquots from these cultures were transferred to microcentrifuge tubes, pelleted, washed and resuspended in 1 ml of sterile distilled water. Serial dilutions of cells were plated on uracil drop-out (selective) and YPD (permissive). For both assays, plates were incubated at 30° C, and colonies were counted from the permissive plates after 2 days of growth and from the selective plates after 4 days. The colony counts were used to calculate recombination rates and 95% confidence intervals (Table S2.2) using the Lea &

Coulson method of the median within the FALCOR web application [http://www.keshavsingh.org/protocols/FALCOR.html] (LEA and COULSON 1949; HALL *et al.* 2009). Statistical analyses of comparisons between recombination rates were performed using a two-sided nonparametric Mann Whitney test in GraphPad Prism software. The calculated *p*-values for several of the most relevant pairwise recombination rate comparisons are shown in Table S2.3.

Molecular karyotype analysis

We analyzed the spectra of chromosomal rearrangements present in recombinant Ura^+ clones derived from WT and *rnh201Δ* NAHR diploid strains. Independent Ura^+ clones were picked from uracil drop-out plates, purified to single colonies in YPD plates, and then patched on to diagnostic plates containing uracil drop-out, 5-FOA, YPD Geneticin, and YPD media. Full length chromosomal DNA embedded in agarose was prepared and fractionated by pulse field gel electrophoresis (PFGE), and in some cases the DNA was used for comparative genome hybridization microarrays (array-CGH) using the methods and microarray design described previously (ZHANG *et al.* 2013a).

Results

Genome-wide ribonucleotide-dependent LOH in a pol2-M644G rnh201Δ strain.

We constructed diploid yeast strains carrying the *pol2-M644G* allele and either *RNH201* or *rnh201Δ*. Multiple clonal isolates were passaged on solid, complete medium, their genomes were sequenced, and mutations were identified by comparison to "zero passage" genomes for each strain (LUJAN *et al.* 2014). In strains proficient in RNase H2, the majority of the mutations that accumulated were single base events whose allelic fractions were between 40 and 60%, as expected for a heterozygous state for new mutations. The identities, distributions in the genome and rates of formation of these mutations in RNase H2-proficient strains have recently been described in detail (LUJAN *et al.* 2014). In comparison, the *pol2-M644G rnh201Δ* strain accumulated single base mutations and 2-5 bp deletions at higher rates (to be

described in detail elsewhere). However, what was not expected was that the allelic fraction of a substantial number of point mutations approached 100%. These loss-of-heterozygosity (LOH) mutations suggested that a subset of ribonucleotide-induced mutations quickly became homozygous through interhomolog recombination. The number of these LOH events was larger in the *pol2-M644G rnh201Δ* strain as compared to the *pol2-M644G RNH201* strain (14 *vs.* 111; Table 2.1). Given the number of generations over which these events accumulated, we conservatively estimate that a defect in RNase H2 increased the rate of LOH in the *pol2-M644G* strain by 3.7-fold. This result motivated the more specific genetic analyses described next.

Chromosome instability as measured by loss-of-heterozygosity (LOH)

We investigated the role that ribonucleotides play in stimulating structural genomic alterations using assays specifically designed to measure allelic and non-allelic homologous recombination in proliferating yeast diploid cells. The first assay was designed to characterize the loss-of-heterozygosity (LOH) effect observed in the mutation accumulation experiment described above. We used isogenic diploids (Fig. 2.1; Table S2.1) that were homozygous at all positions on Chr7, except that they contained a single hemizygous copy of the CORE2 counter-selectable cassette with two diverged orthologous copies of the *URA3* gene and a marker for geneticin resistance (ZHANG *et al.* 2013a) inserted near the right end of Chr7. An allelic interhomolog recombination event occurring anywhere within the 575 Kb region between *CEN7* and the CORE2 insertion can result in LOH that may cause a diploid to become homozygous for the distal regions of the right arm of Chr7. This strain setup allows selection for homozygosity of the homolog lacking the CORE2 insertion, therefore the recombination rates measured by this approach correspond to only half of the LOH events that occur in the region.

We also conducted a limited number of experiments using hybrid diploids that resulted from mating between two diverged haploids (Methods and Materials; Fig. S2.1). In this case, the genotype for two heterozygous positions along Chr7 were monitored in clones selected for the loss of the hemizygous CORE2 insertion. All examined 5-FOA^{R} clones derived from wild type and RNase H2 deficient hybrid

diploids remained heterozygous at a position on the left arm of Chr7, and most became homozygous for a marker located on the right arm approximately 11 Kb proximal to the CORE2 insertion. This result indicated that the majority of 5-FOA^R clones selected through this approach formed by interhomolog recombination leading to LOH, and that chromosome loss was not a frequent occurrence. In both the isogenic (Fig. 2.1) and hybrid (Fig. S2.1) diploid backgrounds, the LOH rates measured from diploids carrying the hemizygous CORE2 insertion were in the 10^{-5} to 10^{-4} 5-FOA^R/cell/division range, and all 5-FOA^R clones examined became concomitantly sensitive to geneticin. When similar measurements were made from haploid strains, the rates were in the 10^{-8} range (data not shown). Thus, this assay provides a specific measurement for allelic interhomolog mitotic recombination that happens frequently in diploid genomes, without interference from chromosome loss, or from rare and more complex mutational mechanisms such as gross chromosomal rearrangements or clustered point mutations that simultaneously inactivate the two *URA3* genes.

Ribonucleotide-dependent LOH

We analyzed isogenic diploid strains (Fig. 2.1 and Fig. 2.2A) lacking each of the three genes encoding subunits of the RNase H2 enzyme: *rnh201Δ*, *rnh202Δ*, and *rnh203Δ*. All three single mutants showed a similarly significant 6 to 8-fold elevation of the LOH rate relative to wild type (Tables S2.2 and S2.3; *RNH201* vs. *rnh201Δ* $p<0.0001$). A similar phenotypic differential was observed between the LOH rate measurements made using the hybrid diploid strain background (rate in CG379xYJM789 *RNH201* $1.8\text{x}10^{-5}$, and $8.5\text{x}10^{-5}$ in CG379xYJM789 *rnh201Δ*). These direct measurements were consistent the genome-wide result presented in Table 2.1, confirming that a defect in RER can cause chromosomal instability in the form of LOH.

The mutagenic effects of unrepaired ribonucleotides have previously been determined to be dependent on inappropriate cleavage by Top1, including induction of 2-5 bp deletions, gene conversions and chromosomal rearrangements. We found this to also be true for LOH stimulation, as the *rnh201Δ*

top1Δ double mutant displayed a rate that was significantly lower than that of the *rnh201Δ* single ($p<0.0001$). The LOH rate in the *top1Δ* single mutant was slightly lower than that of wild type, but not significantly so ($p<0.0853$).

A recent study used haploid strains carrying mutant alleles of three different DNA polymerase genes that differentially alter the rate of incorporation of ribonucleotides into DNA (WILLIAMS *et al.* 2015). We pursued a similar approach to investigate whether the asymmetric mutagenic effect of ribonucleotides also applies to LOH stimulation. Diploids homozygous for the *pol2-M644G* allele that encodes a mutant version of Pol ε that increases ribonucleotide incorporation showed an 8-fold elevation in LOH ($p<0.0001$). This rate was further elevated to 23-fold above wild type ($p<0.0001$) when the ability to remove these extra ribonucleotides was eliminated in a *pol2-M644G rnh201Δ* double mutant. We also examined strains carrying the *pol2-M644L* allele that encodes a mutant version of Pol ε that incorporates fewer ribonucleotides. By itself this mutation did not significantly alter the LOH rate, but in combination with *rnh201Δ*, the LOH rate was lower than that measured for the wild type version of Pol ε (*POL2 rnh201Δ* vs. *pol2-M644L rnh201Δ*; $p=0.0349$). For both alleles of *POL2*, deletion of *TOP1* caused significant decreases in the LOH rate. Together, these results support the existence of a strong direct relationship between the frequency of ribonucleotide incorporation by Pol ε and the rate of Top1-dependent LOH.

Diploids homozygous for the *pol1-L868M* or the *pol3-L612M* alleles which respectively encode mutant versions of Pol α and Pol δ that increase ribonucleotide incorporation did not display altered rates of LOH relative to diploids with wild-type polymerases. Neither mutation increased the rate of LOH when combined with loss of RNase H2 activity. The rate of LOH in *pol1-L868M rnh201Δ* was similar to the rate in *POL1 rnh201Δ* ($p=0.1821$), and surprisingly, the rate of LOH in *pol3-L612M rnh201Δ* was ~50% lower than that in *POL3 rnh201Δ* strains ($p=0.0026$). Finally, deletion of *TOP1* from strains

containing the *pol1-L868M* or the *pol3-L612M* alleles in combination with *rnh201Δ* significantly decreased the LOH rate. We conclude from these results that ribonucleotides incorporated by Pol α or Pol δ do not induce Top1-dependent LOH as potently as those incorporated by Pol ε.

Chromosome instability as measured by non-allelic homologous recombination (NAHR)

In the second assay, we investigated the formation of chromosomal translocations that result from non-allelic homologous recombination (NAHR) in diploid cells. We used a classic heteroallele recombination approach in which two incomplete overlapping sequences can recombine to regenerate a functional selectable marker. In this case, a 3' truncation of the *URA3* gene designated *URA*- was present on Chr5 while a 5' truncation (-*RA3*) was located on Chr14 (Fig. 2.3A). Homologous recombination involving the shared central regions (*RA*; 622 bp, 100% sequence identity) can create a functional copy of the *URA3* gene at the breakpoint of a Chr14/Chr5 translocation. This event may occur through a reciprocal mitotic crossover, or non-reciprocal mechanisms such as break-induced replication (BIR) or half-crossover (HC) (MALKOVA and IRA 2013; SYMINGTON *et al.* 2014; VASAN *et al.* 2014). In the simple recombination scenarios presented in Fig. 2.3, if a reciprocal crossover occurs between the *RA* substrates, depending on how the recombinant chromatids segregate in the following cell division, the karyotype of the resulting Ura^+ cells may contain either two balanced reciprocal translocations (class 1; Fig. 2.3B), or only the non-reciprocal Chr14/Chr5 translocation associated with a terminal deletion on the left arm of Chr14 and a terminal amplification on the left arm of Chr5 (class 2; Fig. 2.3C). If BIR or HC were the mechanisms of recombination between the *RA* sequences, the same class 2 outcome would be observed. Additionally, since the *kanMX* marker is distal to the -*RA3* sequence on Chr14, the expectation is that Ura^+ clones carrying a terminal deletion of Chr14 should lose *kanMX*, thus becoming sensitive to geneticin. The class 1 and class 2 karyotype configurations predicted from this model were validated using PFGE and array-CGH (Fig. 2.3B and 2.3C).

To characterize this NAHR assay system, we analyzed the karyotypes and the presence of the *kanMX* marker in 67 independent Ura$^+$ clones derived from *RNH201* and *rnh201Δ* diploids (Table 2.2). We analyzed the PFGE profiles of these clones to determine the number of copies of the parental-sized Chr5 and Chr14, the presence of the 650 Kb Chr14/Chr5 translocation with *URA3* at the breakpoint, and the reciprocal 700 Kb Chr5/Chr14 translocation with *kanMX-RA* at the breakpoint. In addition, we inspected the PFGE karyotypes for the presence of any other chromosomal rearrangements not predicted by the model in Fig. 2.3. The results showed that the reciprocal crossover outcome (class 1) was infrequent, having only one example detected from each genotype. This was expected, since the reciprocal crossover mechanism requires the spontaneous initiating DSB lesion to occur within the relatively small *RA* region of the recipient chromosome. In contrast, non-reciprocal outcomes can be initiated by DSBs within *RA* or anywhere in the 245 Kb distal region of Chr14. This higher probability for the formation of spontaneous initiating lesions was reflected in the higher abundance of non-reciprocal recombination outcomes (classes 2 to 7). Interestingly, while class 2 clones were the most abundant category recovered, other configurations were also detected at substantial frequency, particularly classes 3 and 4 (Fig. S2.2). Class 3 clones had karyotypes and copy number profiles that were indistinguishable from those of class 2, with the exception that they were resistant to geneticin. Because class 3 clones contained only one parental-sized copy of Chr14 and retained the *kanMX* marker, they must somehow have lost the copy of Chr14 that did not contain the *RA* recombination substrate. Class 4 was also resistant to geneticin, but contained two parental-sized copies of Chr14. Finally, the relatively rare non-reciprocal classes 5, 6 and 7 involved loss of one of the parental copies of Chr5. Although these alternative events were quite interesting, the detailed characterization of the recombination mechanisms leading to their formation was beyond the scope of this project. Regardless of the final karyotype configurations of the Ura$^+$ clones, all of them resulted from homologous recombination between non-allelic *RA* repeats leading to gross chromosomal rearrangements.

The overall distribution of Ura$^+$ clones in the various karyotype classes was very similar between *RNH201* and *rnh201Δ*. However, we detected six cases in which *rnh201Δ*-derived clones each carried one additional chromosomal rearrangement band with size different from those predicted for the *URA-* / *-RA3* translocations between Chr14 and Chr5 (Table 2.2). No such other chromosomes were detected among the wild type-derived Ura$^+$ clones. While we did not characterize the structures of these other rearrangements, and the numbers of clones analyzed were relatively small, their presence exclusively in the *rnh201Δ* background suggested a higher occurrence of complex genome rearrangements in diploids defective for RNase H2.

Ribonucleotide-dependent NAHR

The experiments described above showed that the recombinogenic effect of RNA-DNA damage, first demonstrated for intramolecular gene conversion (AGUILERA and KLEIN 1988; II *et al.* 2011; POTENSKI *et al.* 2014), also extend to allelic interhomolog recombination leading to LOH. In the absence of RNase H2, these recombination events are presumably initiated by DNA breaks that accumulate following processing of ribonucleotides or R-loops by pathways other than RER. Therefore, it is to be expected that the same lesions may also increase the formation of more complex outcomes such as gross chromosomal rearrangements. This hypothesis was tested previously using the original version of the haploid *URA3-CAN1* GCR assay (ALLEN-SOLTERO *et al.* 2014), and a YAC stability assay (WAHBA *et al.* 2011). Importantly, the majority of the GCR events detected with these two specific strain setups were formed through mechanisms other than homologous recombination, as there were no significant proximal homologous sequence substrates in the regions assayed (CHEN and KOLODNER 1999). Allen-Soltero *et al.* showed that single mutants lacking RNase H2 did not alter the GCR rate, but synergistic stimulation was observed in double mutant combinations with specific suppressors of chromosomal instability. Interestingly, deletion of *RAD51* partly rescued this phenotype, suggesting that a substantial fraction of the GCR events that are increased in RNase H2 mutants form through the homologous recombination

pathway, possibly involving non-allelic repeats. We specifically investigated this possibility using the diploid NAHR assay described above, which was designed to detect chromosomal translocations formed by recombination between homologous substrates present on Chr5 and Chr14 (Fig. 1.3A).

Our quantitative analyses of chromosomal instability (Fig. 2.1, Tables S2.2 and S2.3) showed that Chr14/Chr5 NAHR events were quite rare compared to Chr7 LOH events (the baseline rate of NAHR was two to three orders of magnitude lower than LOH). Only minor, not significant, stimulation of the NAHR rate was detected in single mutants lacking the RNase H2 catalytic subunit (1.3-fold up in *rnh201Δ* vs. *RNH201*, p=0.2817). A similarly insignificant increase was observed in *rnh202Δ*, and no alteration at all in *rnh203Δ*. These small (or no) rate differentials in the single mutants, within a context of rare mutational events, indicated that the NAHR rate measurements were more susceptible than LOH to interference from other confounding pleiotropic effects of the various DNA polymerase genotypes tested, such as differences in cell growth kinetics, mutator phenotypes, roles in replicative repair, or others. Specifically, single mutants carrying either the *pol1-L868M* or *pol2-M644G*, which result in higher incorporation of ribonucleotides by Pol α and Pol ε, respectively, showed mild, yet significant decreases in the NAHR rate (each ~40% reduction relative to wild type, each with p<0.0001). Despite this complication, the results obtained when we combined RNase H2 deficiency with mutant replicases and *top1Δ* showed a trend analogous to that observed in the LOH experiments, and suggested that ribonucleotides incorporated into DNA also contribute to NAHR.

Within each of the four mutant replicase genetic backgrounds tested, the largest NAHR rate increase between *RNH201* and *rnh201Δ* was observed in strains carrying the *pol2-M644G* allele encoding a mutant version of Pol ε that increases the incorporation of ribonucleotides (5.4-fold NAHR rate elevation in *pol2-M644G rnh201Δ* vs. *pol2-M644G RNH201*, p<0.0001). *pol2-M644G* strains also displayed the largest Top1-dependent NAHR rate reduction (9.4-fold decrease in *pol2-M644G rnh201Δ top1Δ* vs. *pol2-M644G rnh201Δ TOP1*, p<0.0001). *pol2-M644L* strains that have lower Pol ε ribonucleotide incorporation did not significantly increase the NAHR rate in combination with *rnh201Δ* (only 1.6-fold

elevation in *pol2-M644L rnh201Δ* vs. *pol2-M644L RNH201*, p=0.0871). These results were consistent with our observations for LOH, and showed that the frequency of ribonucleotide incorporation by Pol ε is an important contributor to the formation of chromosomal rearrangements by NAHR.

The NAHR results obtained within the other two mutant replicase backgrounds that increase ribonucleotide incorporation were not as straightforward as in the LOH experiments. Deficiency of RNase H2 in combination with either *pol1-L868M* or *pol3-L612M* resulted in ~2.3-fold increases in the NAHR rate. These increases were not as robust as the 5.4-fold effect observed for *pol2-M644G*, but they were both significant (*pol1-L868M rnh201Δ* vs. *pol1-L868M RNH201* and *pol3-L612M rnh201Δ* vs. *pol3-L612M RNH201*, each had p<0.0001). In both cases, the increases in the NAHR rate were dependent on *TOP1*. A notable result was the significant ~60% increase in the NAHR rate in *pol3-L612M rnh201Δ* compared to *POL3 rnh201Δ* (p<0.0001), a recombination rate change of approximately same magnitude but opposite direction as observed in the LOH assay (~50% decrease; Fig 2.2A). One possible reason for these contrasting results might be related to the fact that Pol δ is known to participate in the BIR mechanism (SYMINGTON *et al.* 2014), which we showed accounts for the majority of the NAHR events detected in our assay (Table 2.2). In this scenario, the *pol3-L612M* mutation might somehow make Pol δ more efficient at initiating or sustaining BIR. This would promote the non-reciprocal recombination mechanism associated with the NAHR assay, and conversely, might disfavor the reciprocal interhomolog mitotic crossover pathway most often associated with LOH.

The results obtained for NAHR with the mutant alleles of *POL1* and *POL3* can be interpreted as sign that ribonucleotides incorporated by these polymerases have a larger contribution to structural chromosomal rearrangements than they do to allelic interhomolog recombination (*i.e.* LOH). However, even in this scenario, the ribonucleotides incorporated by Pol ε remain as the ones with the most destabilizing effect in both recombination assays used in this study. Alternatively, the inconsistencies in

pol1-L868M and *pol3-L612M* behavior between our two assays might be primarily the result of interference in NAHR rate measurements caused by confounding phenotypes associated with these mutations, and not by their higher inherent rate of ribonucleotide incorporation.

Discussion

A number of recent studies have shown that ribonucleotides are incorporated throughout the genome during DNA replication (reviewed in (JINKS-ROBERTSON and KLEIN 2015)). Incorrect removal of such abundant ribonucleotides can lead to genomic instability in the form of point mutations, short deletions within low complexity regions, and chromosomal rearrangements. Interestingly, the mutagenic effect of ribonucleotides appears to be asymmetric, according to which DNA polymerase was responsible for their incorporation. Specifically, *TOP1*-dependent 2-5 bp deletions in an RNase H2 mutant background were shown to be the product of ribonucleotides incorporated by Pol ε, but not by Pol α and Pol δ (WILLIAMS *et al.* 2015).

Earlier reports described increased chromosome-scale instability in RNase H2 mutant haploid yeast cells, including Top1-dependent gene conversion (POTENSKI *et al.* 2014) and gross chromosomal rearrangements (ALLEN-SOLTERO *et al.* 2014). Another study measured the rate of LOH at Chr3 in RNase H homozygous mutant diploids (WAHBA *et al.* 2011). Although Wahba *et al.* did not observe significant change in LOH in *rnh1Δ* or *rnh201Δ* single mutants, they did report an 8-fold elevation in *rnh1Δ rnh201Δ* double mutants lacking both RNase H1 and H2 activities. Finally, a study complementary to our own has been performed recently in hybrid yeast diploids to quantify LOH stimulation genome-wide and characterize the associated recombination tracts in RNase H defective cells. These experiments showed elevated LOH in *rnh201Δ* and *rnh1Δ rnh201Δ*, but not in *rnh1Δ* single mutants (K. O'Connell, S. Jinks-Robertson, T. Petes; personal communication). Neither of the diploid studies described above investigated the *TOP1* dependency of the reported increases in LOH.

In this study, we show that genome-wide LOH occurs more frequently in diploids lacking RNase H2 activity, and that the rate of LOH on Chr7 is elevated in mutants carrying deletion of RNase H2 subunits. The increased LOH phenotype in *rnh201Δ* was almost entirely suppressed by deletion of the *TOP1* gene. Further, we showed that the rate of LOH is modulated by increasing or decreasing the number of ribonucleotides incorporated by mutant alleles of Pol ε, and this effect was also dependent on *TOP1*. In contrast, alleles of Pol α or Pol δ that increase ribonucleotide incorporation did not alter the rate of allelic interhomolog recombination.

We also observed a role for ribonucleotides incorporated by Pol ε in stimulating chromosomal translocations formed by NAHR. Although the phenotypic differential provided by this assay system is small, the highest rate of NAHR was measured in the *rnh201Δ pol2-M644G* strain, and chromosomal size polymorphisms other than the selected translocations were only detected in the *rnh201Δ* background. Together, these observations were consistent with earlier reports for a role of ribonucleotides in the generation of gross chromosomal rearrangements in yeast (WAHBA *et al.* 2011; ALLEN-SOLTERO *et al.* 2014), and increased cytogenetic abnormalities in mammalian cells (REIJNS *et al.* 2012).

The recombinogenic effects associated with RNase H2 mutants may result from misprocessing of scattered ribonucleotides incorporated into DNA, misprocessing of R-loops, or a combination of these two defects. The observation that increased and decreased ribonucleotide incorporation by Pol ε correlates with the LOH rate leads us to propose that much of the LOH observed here in RNase H2 mutants is triggered by ribonucleotides incorporated by Pol ε during leading strand replication. This idea is consistent with our earlier proposal that ribonucleotides incorporated by Pol ε, but not (or less so) by Pols α or δ, result in 2-5 bp deletions that result from incisions by Top1 (Williams *et al.* 2015). Thus, it appears that after Pol ε incorporates a ribonucleotide, topoisomerase 1 cleavage of the DNA backbone at a ribonucleotide provides opportunities for multiple types of genome instability, including loss of a short repeat, LOH and NAHR. Possible mechanistic reasons for this observed asymmetry have been proposed (Figure 5 in (WILLIAMS *et al.* 2015)) and are currently being tested. The first centers on the fact that the

RER-defective *pol2-M644G* strain has a higher density of ribonucleotides in the nascent leading strand than is observed in the nascent lagging strand for the *pol1-L868M* and *pol3-L612M* mutator strains lacking *RNH201*, suggesting that ribonucleotide density must exceed a certain threshold in order to observe Top1-dependent recombination events. Alternatively, there may be additional pathways involved in removal of nascent lagging strand ribonucleotides that are either specific for this DNA strand or less available for ribonucleotide removal from the nascent leading strand. Finally, we propose that negative supercoils may accumulate in the continuous nascent leading strand in the wake of the replisome and Top1, through its interaction with the CMG helicase (GAMBUS *et al.* 2006), would be in an ideal physical position to relieve this torsional stress. Such helical tension may not build up in the discontinuous nascent lagging strand because of the presence of DNA ends that could allow rotation. As with more localized forms of RNA-DNA damage, the asymmetry of Top1-dependent recombination events support the concept that failure to remove ribonucleotides incorporated into DNA by Pol ε during leading strand synthesis puts genome stability at risk.

Figures

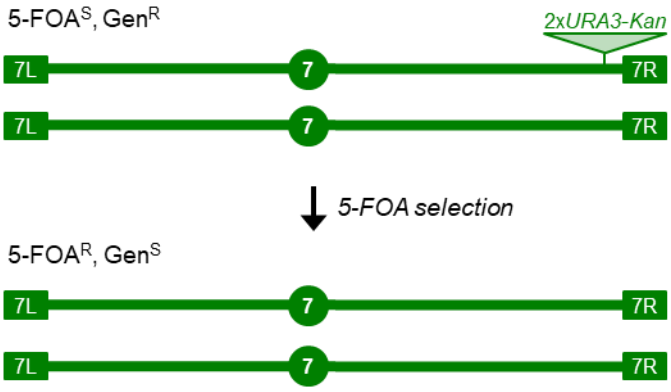


Figure 2. 1 Loss-of-heterozygosity (LOH) assay system in isogenic diploids.

Schematic representation of the karyotype of the isogenic diploid strains used in the LOH assays in Fig. 2.2A and Table S2.2. A hemizygous counter-selectable CORE2 cassette was inserted near the right end of one of the Chr7 homologs. Allelic homologous recombination between the two homologs may result in homozygosity for the region lacking the CORE2 insertion. Derived clones carrying this LOH event were resistant to 5-FOA and sensitive to geneticin. Terminal boxes labeled L and R correspond to the left and right telomeres, respectively. The circle corresponds to the centromere.

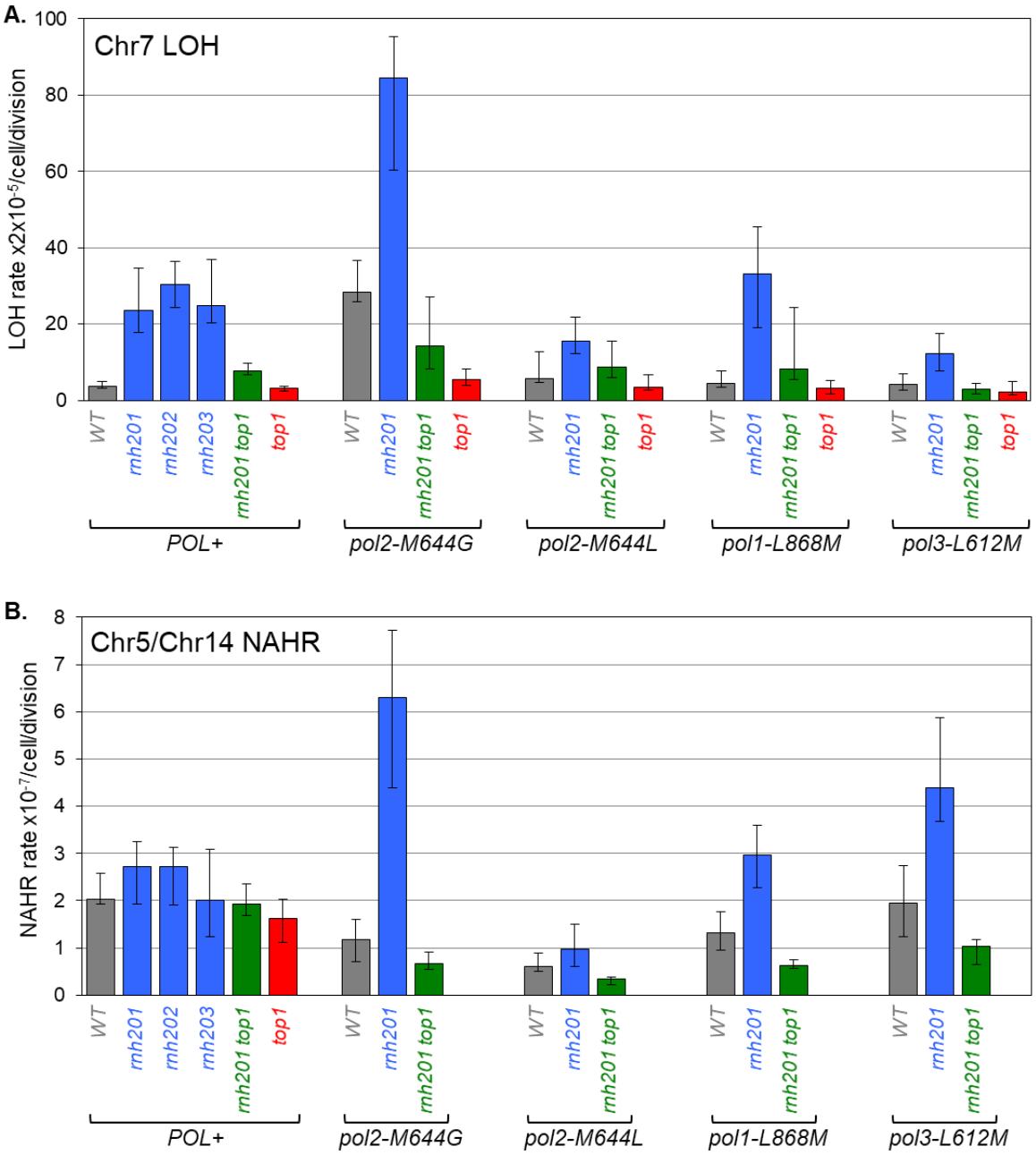


Figure 2. 2 Quantitative analyses of mitotic recombination.

The graphs show the recombination rates determined from the LOH (A) and NAHR (B) assays, described in Fig. 2.1 and Fig. 2.3, respectively. The bars correspond to the median recombination rates and the error bars represent the 95% confidence intervals. The results are grouped according to DNA polymerase genotype, and the bars are color-coded according to RNAase H2 and topoisomerase 1 genotypes. The same numerical values presented graphically in A and B are reproduced in Table S2.2 for reader reference, also including the number of cultures assayed for each genotype. Statistical significance analyses of specific pairwise comparisons between genotypes are shown in Table S2.3.

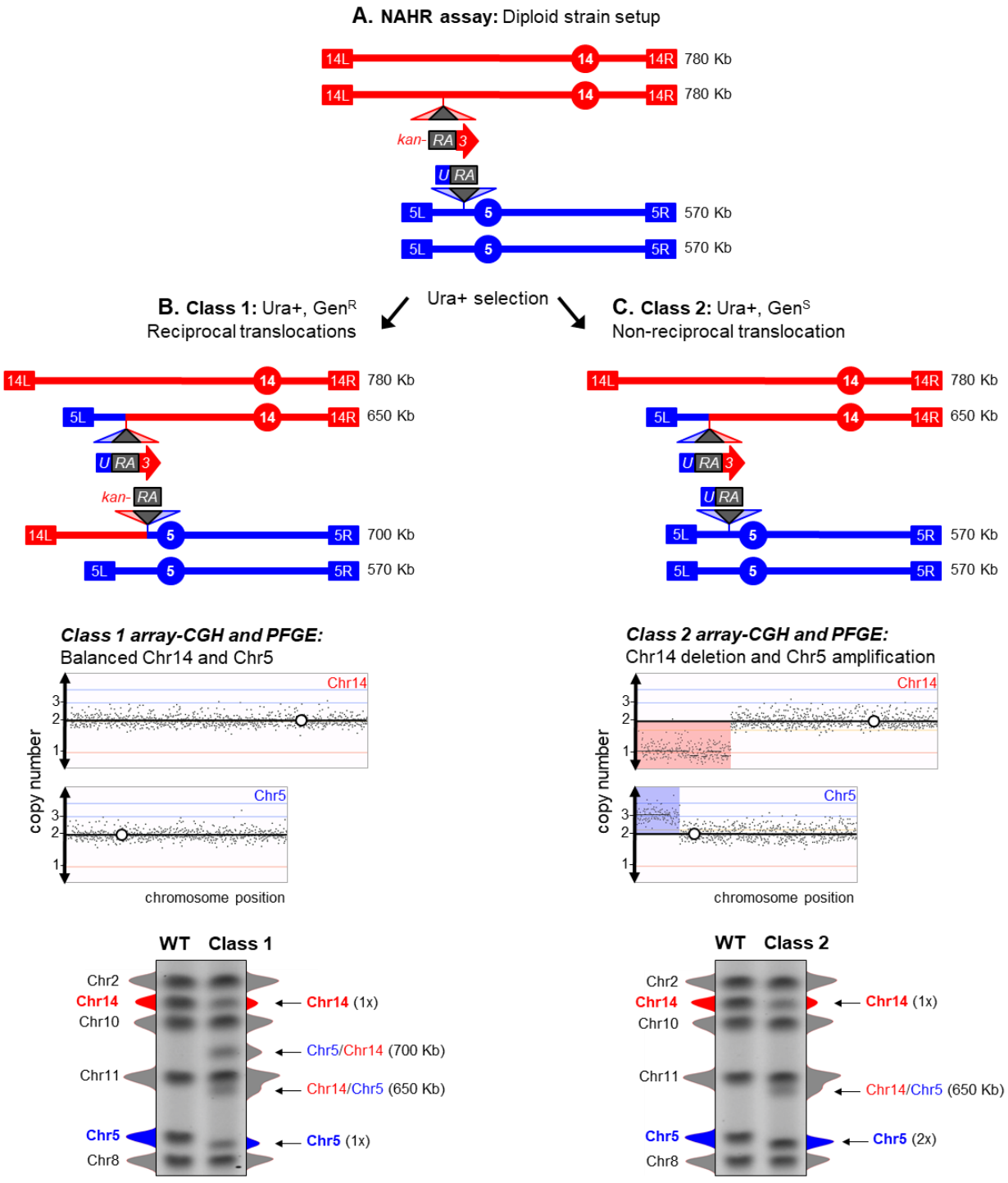


Figure 2.3. Non-allelic homologous recombination (NAHR) assay and predicted recombination outcomes.

Figure 2.3. Non-allelic homologous recombination (NAHR) assay and predicted recombination outcomes.

(A) Schematic representation of the karyotype of the diploid strain used in the NAHR assay. Chr5 (blue) and Chr14 (red) are drawn to approximate scale. Terminal boxes labeled L and R correspond to the left and right telomeres, respectively, and the numbered circles correspond to the centromeres. Only one of the Chr14 homologs contains a *Kan-RA3* insertion, and only one of the Chr5 homologs contains a *URA* sequence. Both recombination substrates are present in Watson orientation and no other *URA3* sequences are present in the genome. Recombination between the *RA* sequences can regenerate a full-length functional copy of *URA3*, selectable on uracil drop-out medium.
(B) The class 1 reciprocal crossover outcome is shown to the left, with the respective sizes and structure of the associated chromosomal rearrangements. The array-CGH plots for Chr5 and Chr14 and PFGE from a representative class 1 clone are shown. The array-CGH plots Y-axis corresponds to copy number (Log2 (Cy5-labeled Ura^+ clone DNA / Cy3-labeled parental DNA)). The X-axis corresponds to the probe coordinates along the respective chromosomes. The white circles indicate the positions of *CEN5* and *CEN14*. The gray dots indicate the Log2 Cy5/Cy3 values and chromosome position of each array probe. The copy number profile of class 1 clones was fully balanced, with no gains or losses relative to the parental diploid. The PFGE was cropped for emphasis, showing only the region from Chr8 (540 Kb) to Chr2 (815 Kb), with lane quantification traces flanking the image. Chr5 and Chr14 trace peaks are shaded in blue and red, respectively. (C) The class 2 non-reciprocal karyotype outcome is shown to the right. The array-CGH plots for a representative class 2 clone show a deletion (1 copy; pink-shaded) on the left arm of Chr14 from *TEL14L* to *PEX17*, and an amplification (3 copies; purple-shaded) on the left arm of Chr5 from *TEL05L* to *URA3*. The class 2 PFGE profile and its quantitative analysis are also shown. Note that the WT diploid strain used as reference in the PFGEs in B and C, and in Fig. S2.2, has a slightly longer Chr5 band because it is homozygous for the *ura3-52* allele (Ty1 insertion) rather than the *ura3Δ 3'* and *ura3Δ 0* alleles present in the NAHR parent diploid strain. The overall difference in Chr5 sizes caused by the Ty1 insertion and *ura3* deletions is approximately 7 Kb

Tables

Table 2.1 LOH events detected by whole genome sequencing of mutation accumulation lines.

Diploid homozygous genotype	*pol2-M644G RNH201*	*pol2-M644G rnh201Δ*
Number of isolates sequenced	8	7
Total generations elapsed	7200	6300
Total mutation count	436	912
LOH count	14	111
Mean LOH rate (per Mbp per generation)	38	140
Standard deviation (per Mbp per generation)	33	41
Fold difference in rate	3.7	
Welch's *t*-test *p*-value	0.00019	

All rates are lower bound estimates per million base pairs (Mbp) per generation. Lower bound LOH rates are estimated for each isolate thusly: (homozygous mutation count)/(mutation count)/(generations elapsed).

Table 2.2 Classes of Ura$^+$ NAHR clones.

NAHR class	Phenotype		PFGE Karyotype [a]				Clones in each Ura$^+$ class [b]	
	Ura	Gen	Chr14/Chr5 650 Kb	Chr5/Chr14 700 Kb	Chr5 570 Kb	Chr14 780 Kb	*RNH201*	*rnh201Δ*
Parent diploids	-	R	NA	NA	2	2	NA	NA
1	+	R	1	1	1	1	1	1
2	+	S	1	0	2	1	14	$17^{(1)}$
3	+	R	1	0	2	1	9	$15^{(4)}$
4	+	R	1	0	2	2	5	1
5	+	R	1	0	1	1	1	0
6	+	R	0	0	1	1	1	$1^{(1)}$
7	+	S	1	0	1	1	1	0
Total							32	$35^{(6)}$

a. The numbers (0, 1, or 2) indicate the number of copies of the indicated chromosomes (parental size Chr5 and Chr14, and the two translocations) determined by measuring the ethidium bromide staining intensity of the corresponding band in PFGE relative to the parental bands in the parent strains (representative examples shown in Fig. 2.3 and Fig. S2.2).
b. The number of Ura$^+$ clones in each class is shown, and when present, the superscript number between parenthesis indicates the number of clones that contained additional chromosomal rearrangements of size other than 650 Kb or 700 Kb. Six Ura$^+$ clones derived from *rnh201Δ* diploids contained such uncharacterized rearrangements as observed by PFGE; no rearrangements of unexpected sizes were observed in the wild type control. Note that all six rearrangements in question had sizes below 1200 Kb, therefore were unlikely to correspond to contractions of the ribosomal DNA tandem repeat cluster on Chr12.
NA. Not Applicable.

CHAPTER THREE: BOTH R-LOOP REMOVAL AND RIBONUCLEOTIDE EXCISION REPAIR ACTIVITIES OF RNASE H2 CONTRIBUTE SUBSTANTIALLY TO CHROMOSOME STABILITY[3]

Summary

Cells carrying deletions of genes encoding H-class ribonucleases display elevated rates of chromosome instability. The role of these enzymes is to remove RNA-DNA associations including persistent mRNA-DNA hybrids (R-loops) formed during transcription, and ribonucleotides incorporated into DNA during replication. RNases H1 and H2 can degrade the RNA component of R-loops, but only RNase H2 can initiate accurate ribonucleotide excision repair (RER). In order to examine the specific contributions of these activities to chromosome stability, we measured rates of loss-of-heterozygosity (LOH) in diploid *Saccharomyces cerevisiae* yeast strains carrying the *rnh201-RED* separation-of-function allele, encoding a version of RNase H2 that is RER-defective, but partly retains its other activity. The LOH rate in *rnh201-RED* was intermediate between *RNH201* and *rnh201Δ*. In strains carrying a mutant version of DNA polymerase ε (*pol2-M644G*) that incorporates more ribonucleotides than normal, the LOH rate in *rnh201-RED* was as high as the rate measured in *rnh201Δ*. Topoisomerase 1 cleavage at sites of ribonucleotide incorporation has been recently shown to produce DNA double strand breaks.

[3] *This chapter is an adaptation of previously published material, the figures have been renumbered to indicate both chapter and figure number.*

Reference for the full article:
**Indicates shared first authorship.*

Sedam, H.N.*, Cornelio, D.A.*, Ferrarezi, J., Sampaio, N.M.V, Argueso, J.L. 2017 Both R-loop removal and ribonucleotide excision repair activities of RNase H2 contribute substantially to chromosome stability. DNA Repair. DOI: 20.1016/j.dnarep.2017.02.012.

Contributions to this research are as follows:
LOH Assay: **HNS**, DAC, JF, NMVS, Karyotype Analysis: **HNS**, DAC, JLA, Data Analysis: JLA, **HNS**, DAC, Strain Construction: **HNS**, DAC, JF Manuscript Preparation: JLA, DAC, **HNS**

Accordingly, in both the *POL2* and *pol2-M644G* backgrounds, the LOH elevation in *rnh201-RED* was suppressed by *top1Δ*. In contrast, in strains that incorporate fewer ribonucleotides (*pol2-M644L*) the LOH rate in *rnh201-RED* was low and independent of topoisomerase 1. These results suggest that both R-loop removal and RER contribute substantially to chromosome stability, and that their relative contributions may be variable across different regions of the genome. In this scenario, a prominent contribution of R-loop removal may be expected at highly transcribed regions, whereas RER may play a greater role at hotspots of ribonucleotide incorporation.

Introduction

RNA strands, as well as single and tandem ribonucleotides (rNMPs) can in some cases be transiently associated with chromosomal DNA (SANTOS-PEREIRA and AGUILERA 2015; WILLIAMS *et al.* 2016). For example, during transcription the nascent mRNA chain may remain associated as a hybrid duplex with the DNA template strand (R-loops). During replication, a substantial number of single rNMPs are incorporated into newly synthesized DNA by the replicative polymerases, and delay or failure in the removal of tandem rNMPs used to prime lagging strand synthesis may cause these structures to persist. Any of these RNA-DNA associations may interfere with normal DNA transactions and thereby destabilize chromosomes. Two H-class ribonucleases are tasked with removing these RNA structures from DNA (CERRITELLI and CROUCH 2009) (Fig. 3.1A). RNases H1 and H2 have a redundant role in degrading the RNA component of RNA-DNA hybrids and tandem rNMPs in DNA, with RNase H2 likely playing a more prevalent role (ZIMMER and KOSHLAND 2016). In addition, RNase H2, but not RNase H1, has a second distinct activity, which is to initiate the accurate removal of single rNMPs incorporated into DNA (ribonucleotide excision repair - RER) (SPARKS *et al.* 2012).

Multiple studies have shown that in the absence of these enzymes (RNase H2 in particular) eukaryotic genomes can become destabilized in different ways. For example, 2-5 bp deletions accumulate at low complexity regions in RNase H2 mutants (CHO and JINKS-ROBERTSON 2016). This phenotype stems from a defect in normal RER initiation, which then affords an opportunity for mutagenic processing

by topoisomerase 1 at regions containing rNMPs (KIM *et al.* 2011; CHON *et al.* 2013; SPARKS and BURGERS 2015; HUANG *et al.* 2016; WILLIAMS *et al.* 2017). In addition to nucleotide-level mutations, the absence of RNase H2 has also been shown to destabilize chromosome structure, leading to increased rates of gene conversion, gene duplication, chromosomal rearrangements, chromosome loss, and loss-of-heterozygosity (LOH) (AGUILERA and KLEIN 1988; WAHBA *et al.* 2011; REIJNS *et al.* 2012; ALLEN-SOLTERO *et al.* 2014; POTENSKI *et al.* 2014; CONOVER *et al.* 2015; O'CONNELL *et al.* 2015). These structural mutations are presumably triggered by DNA lesions resulting from the accumulation and improper processing of RNase H2 substrates. For example, persistent R-loops stalling replication fork progression, leading to their eventual collapse (SANTOS-PEREIRA and AGUILERA 2015); and processing of single rNMPs by topoisomerase 1, resulting in DNA double strand breaks (HUANG *et al.* 2016). The relative contributions of these different sources to chromosome instability has not been fully ascertained, and contrasting views favoring one source over the other have been proposed (CONOVER *et al.* 2015; O'CONNELL *et al.* 2015).

This problem has been difficult to study directly because most prior studies have used full deletions of genes encoding essential subunits of RNase H2, in which R-loop, tandem and single rNMP removal activities are concomitantly eliminated. An approach that can illuminate this issue has been proposed (CERRITELLI and CROUCH 2016) through the use of a separation-of-function allele of the *RNH201* gene that encodes the catalytic subunit of RNase H2. This mutant, *rnh201-RED* (Ribonucleotide Excision Defective) (CHON *et al.* 2013), is completely unable to remove single rNMPs, but retains partial enzymatic activity for the removal of tandem rNMPs, and presumably also R-loops (Fig. 3.1A). Two recent studies (EPSHTEIN *et al.* 2016; ZIMMER and KOSHLAND 2016) reported mildly elevated chromosome instability phenotypes in *rnh201-RED* compared to wild type, but the elevation was not as high as that seen in *rnh201Δ*. In each case the authors concluded that RER was not a substantial contributor to chromosome stability. In this report, we measured chromosome instability in the *rnh201-RED* mutant, and in a comprehensive panel of double and triple mutant combinations with *rnh1Δ*, *top1Δ*

and two DNA Pol ε (*pol2*) alleles. We also observed elevated chromosome instability in *rnh201-RED*, and it was dependent on the presence of topoisomerase 1 and on the level of rNMP incorporation by Pol ε. We favor an interpretation of these results in which the RER activity of RNase H2 does play a meaningful role in promoting chromosome stability, particularly at regions of the genome that are prone to frequent rNMP incorporation. An alternative model in which RER plays a lesser role is also discussed.

Materials and Methods

Culture media, plasmids and yeast strains.

Saccharomyces cerevisiae yeast cells were cultured in conventional YPD and dropout media at 30C (ROSE *et al.* 1990). The strains used in this study were isogenic with the CG379 background (MORRISON *et al.* 1991), with modifications described previously (ARGUESO *et al.* 2008; ZHANG *et al.* 2013a). The genotypes of all yeast strains used are detailed in Table S3.1. The *rnh201-RED* allele (*rnh201-P45D-Y219A*) (CHON *et al.* 2013) was integrated at the native *RNH201* locus using a two-step allele replacement strategy using the pRS306 vector (SIKORSKI and HIETER 1989) (construction details in Table S3.1 footnotes). The *RNH1* gene was deleted by PCR-mediated integration of the *KanMX4* cassette (WACH *et al.* 1994) into strain JAY1161.

Loss-of-heterozygosity (LOH) assay and statistical analyses.

The LOH assays were conducted as previously described (CONOVER *et al.* 2015). Part of the data presented in Figure 3.1C and Table S3.2 are composites of rates reported previously plus additional cultures generated from the same strains for the present study. Recombination rates and 95% confidence intervals (CI) were determined using the method of the median (LEA and COULSON 1949; SPELL and JINKS-ROBERTSON 2004). Specific pairwise comparisons between LOH rates were assessed for statistical significance through a non-parametric Mann-Whitney test using GraphPad Prism software (Table S3.3).

Results and Discussion

Initial observations and experimental approach

We recently reported results from three different experimental approaches in diploid yeast cells (whole genome unselected LOH; right arm of chromosome 7 [Chr7] selected LOH; and Chr5/Chr14 translocations mediated by non-allelic homologous recombination) that suggested that a defect in RER resulting from absence of the catalytic subunit of RNase H2 (*rnh201Δ*) may contribute significantly to chromosome instability (CONOVER *et al.* 2015). Our conclusion was based on the observations that (1) the absence of topoisomerase 1 (*top1Δ*) suppressed the high rates of recombination measured in *rnh201Δ*; and (2) the rates of recombination in *rnh201Δ* could be modulated up or down in mutant Pol *Δ* backgrounds more or less prone to incorporate ribonucleotides into DNA (*pol2-M644G* or *pol2-M644L*, respectively). However, we were not able to estimate the specific extent of RER's contribution to these phenotypes, as all enzymatic activities of RNase H2 were eliminated in *rnh201Δ* (Fig. 3.1A).

Of the three approaches listed above, the Chr7 LOH assay was both the most straightforward from a technical standpoint, and also showed the largest phenotypic differentials between the tested genotypes. In this assay (Fig. 3.1B), diploid cells have a hemizygous insertion of a double *URA3* counter-selectable cassette near the right end of Chr7 (*MAL13::CORE2*). Mutations leading to loss of the cassette can be selected for by resistance to 5-fluoroorotic acid (5-FOA). Our previous characterization of the spectrum of 5-FOA resistance mechanisms (CONOVER *et al.* 2015), showed that most clones examined (26/27) were consistent with allelic interhomolog recombination leading to LOH, with breakpoints in the right arm anywhere in the long 575 Kb interval between *CEN7* and *MAL13*. None of the 27 examined clones were due to double point mutations in the *URA3* genes or to chromosome loss. The high specificity of the 5-FOA selection for LOH is likely observed because the rate of double *URA3* inactivating point mutations is predicted to be below our detection level ($<10^{-12}$), and monosomy of large chromosomes

(*e.g.* Chr7 or Chr4) might be deleterious and thus difficult to recover, whereas monosomy of small chromosomes (*e.g.* Chr1 or Chr3) is tolerated relatively well (ARGUESO *et al.* 2008; MCCULLEY and PETES 2010; COVO *et al.* 2014).

Chromosome stabilization role of RNase H2 in the presence of RNase H1.

We revisited the Chr7 LOH assay (CONOVER *et al.* 2015); graphic data representation in Fig. 3.1C; numerical data in Table S3.2; statistical analyses in Table S3.3), now including data for the *rnh201-RED* separation-of-function allele. We initially measured Chr7 LOH in the *RNH1 POL2* background, thus cells retain some tandem rNMP and R-loop removal ability, and have basal levels of rNMP incorporation (first set of columns from the left in Fig. 3.1C). We found that the rate of LOH in *rnh201-RED* was significantly higher than in *RNH201* ($p<0.0001$), and slightly more than half of the phenotype stimulation measured in *rnh201Δ* (3.1-fold and 5.8-fold higher than wild type, respectively). In addition, the LOH stimulation seen in the *rnh201-RED* single mutant was significantly lowered ($p<0.0001$) in the *rnh201-RED top1Δ* double mutant, thus the chromosome instability phenotype was dependent on the action of topoisomerase 1, possibly through double strand break formation at sites of single rNMP incorporation (HUANG *et al.* 2016). Since cells with the *RNH1 rnh201-RED* genotype are equipped to remove tandem rNMPs and R-loops, using the full RNase H1 and partial RNase H2 activities, we suggest that perturbation of proper RER initiation plays a meaningful role in destabilizing yeast chromosomes. According to this interpretation, our results would indicate that, in the absence of other genetic defects, the RER and tandem rNMP/R-loop removal activities of RNase H2 make approximately equal contributions to preventing LOH on the right arm of Chr7.

A plausible alternative interpretation would be that RER does not play a role in preventing LOH, and that the mild phenotype observed in *rnh201-RED* is caused by the partial defect in cleavage of tandem rNMPs (CHON *et al.* 2013). This model was favored by Epshtein *et al*. in their recent analysis of gene duplication at two independent reporter loci (EPSHTEIN *et al.* 2016). These authors reported

intermediate recombination rates in *rnh201-RED*, reaching approximately one third of the stimulation seen in *rnh201Δ*, and concluded that the RER defect was not a prominent contributor to locus-specific instability, since changing the frequency of rNMP incorporation with mutant Pol ε alleles did not significantly influence the duplication rates. These Pol ε results were in contrast to our earlier and present measurements of regional chromosomal instability (*i.e.* 575 Kb Chr7 LOH; see item 3.4 below).

Chromosome stabilization role of RNase H2 in the absence of RNase H1

Next we investigated how the absence of tandem rNMP and R-loop removal carried out by RNase H1 can influence the phenotypes of RNase H2 mutants (second set of columns from the left in Fig. 3.1C). While in principle RNases H1 and H2 both have redundant tandem rNMP and R-loop removal activities, RNase H2's activity is thought to be more prevalent, having a global genomic role, whereas RNase H1's role might be confined to a subset of R-loop forming regions (ZIMMER and KOSHLAND 2016). Thus, loss of RNase H1 by itself should have a limited impact on chromosome instability. Consistently, the rate of Chr7 LOH in the *rnh1Δ* single mutant did not significantly change relative to wild type ($p=0.8133$), but increased by twelve fold in the *rnh1Δ rnh201Δ* double ($p<0.0001$), in agreement with previous reports (O'CONNELL *et al.* 2015; ZIMMER and KOSHLAND 2016). This pronounced elevation in LOH when both RNase H's are absent underscores how important tandem rNMP/R-loop removal is for chromosome stabilization as a whole. However, the rate of LOH was elevated 3-fold in the *rnh1Δ rnh201-RED* double mutant compared to *rnh1Δ RNH201* ($p<0.0001$), and was similar to the rate in *RNH1 rnh201-RED* ($p=0.4111$). Likewise, the degree of suppression of the LOH rate was also similar between *rnh1Δ rnh201-RED top1Δ* and *RNH1 rnh201-RED top1Δ* ($p=0.6190$). Therefore, if we assume that the residual tandem rNMP removal activity present in *rnh201-RED* is sufficient to adequately process these substrates and R-loops, then an approximately 3-fold elevation in Chr7 LOH can be assigned to a specific defect in RER. Following this rationale, about 25% of the elevation in Chr7 LOH in the *rnh1Δ rnh201Δ* double mutant can be attributed to a defect in RER, while the remaining 75% was due to a defect in tandem rNMP/R-loop removal.

These results align well with those reported recently for a Chr3 instability assay that detects LOH within a relatively narrow region (98 Kb) plus Chr3 monosomy, each class accounting for about half of the overall instability rate (ZIMMER and KOSHLAND 2016). However, that study did not report chromosome instability rates for the *RNH1 rnh201-RED* single mutant, and therefore was unable fully contextualize the difference between the chromosome stabilization contributed by the RER and tandem rNMP/R-loop removal activities of RNase H2, versus the contributions of RER and the combined tandem rNMP/R-loop removal activities from RNase H1 and H2.

The abundance of rNMPs in DNA modulates the contribution of RER to chromosome stability

We interpret the results above to mean that RER does play a role in promoting chromosome stability. If this were indeed the case, then its contribution should be modulated up or down depending on the frequency of rNMP incorporation during DNA replication. We generated *rnh201-RED* mutant combinations with *pol2-M644G* and *pol2-M644L*, which respectively, encode versions of Pol ε that incorporate ~11-fold more or ~3-fold fewer rNMPs compared to the wild type enzyme (third and fourth set of columns from the left in Fig. 3.1C). As we reported previously, the *pol2-M644G* mutation causes an overall elevation in Chr7 LOH (CONOVER *et al.* 2015). In this study, we found that the LOH rate in the *rnh201-RED pol2-M644G* was very high and was indistinguishable from *rnh201Δ pol2-M644G* (p=0.9330). In both *rnh201-RED pol2-M644G* and *rnh201Δ pol2-M644G* backgrounds, similarly robust reductions in LOH resulted when combined with *top1Δ*. These data suggested that under conditions in which DNA is overloaded with rNMPs, the contribution of RNase H2 to chromosome stability is strongly dependent on its RER role, and the contribution from tandem rNMP/R-loop removal becomes relatively minor.

We also asked whether the absence of RNase H1 could influence the LOH rates in the *RNH201 pol2-M644G* and *rnh201-RED pol2-M644G* backgrounds. The LOH rate in *rnh1Δ RNH201 pol2-M644G* was marginally lower (-25%) than the rate in *RNH1 RNH201 pol2-M644G* (p=0.0288, but with

overlapping CIs). In the case of *rnh1Δ rnh201-RED pol2-M644G*, the LOH rate was slightly higher (+50%, p=0.0068, also with overlapping 95% CIs) than the rate in *RNH1 rnh201-RED pol2-M644G*. This mild elevation might suggest the possibility that a small number of tandem rNMPs are occasionally incorporated by the mutant Pol ε. These rNMP cluster regions could lead to transcription stalling and thereby to stable R-loops.

In contrast to the *pol2-M644G* background, the LOH rate in the *rnh201-RED pol2-M644L* was very low, and actually somewhat lower than in *RNH201 pol2-M644L* (p=0.0007, with overlapping CIs), but no further LOH decrease was observed between *rnh201-RED pol2-M644L TOP1* and *rnh201-RED pol2-M644L top1Δ* (p=0.3795). This indicated that when rNMPs are not incorporated into DNA at an appreciable frequency, the RER activity of RNase H2 no longer plays a meaningful role in chromosome stability, shifting the balance completely toward its role in tandem rNMP/R-loop removal.

Conclusions

Taken together, our results support the conclusion that both enzymatic activities of RNase H2 contribute substantially to the preservation of chromosomal integrity. We showed that the relative contribution of the RER activity is higher when more rNMPs are incorporated into the DNA, and is negligible when the ribonucleotide load is light. A similar modulation of the RER role might also be observed if the ribo- to deoxyribo- ratio in the cell's nucleotide pool is altered, for example by inhibition of ribonucleotide reductase. Likewise, the relative contribution of tandem rNMP/R-loop removal activities to chromosomal stability can be modulated by the frequency of R-loop formation, as had been demonstrated previously and further elucidated recently (SANTOS-PEREIRA and AGUILERA 2015; ZIMMER and KOSHLAND 2016). This flexibility suggests that both activities might, in fact, be critically and simultaneously important. Long and functionally diverse regions of chromosomes, such as the right arm of Chr7 examined here, include loci that are both prone to stable R-loop formation (WAHBA *et al.* 2016) and hotspots for ribonucleotide incorporation (JINKS-ROBERTSON and KLEIN 2015). Therefore, in order to

globally maintain whole chromosomes as stable structures, the two RNase H's must coordinate their enzymatic activities to efficiently remove each type of RNA structure from DNA according to the local context in which they form.

Figures

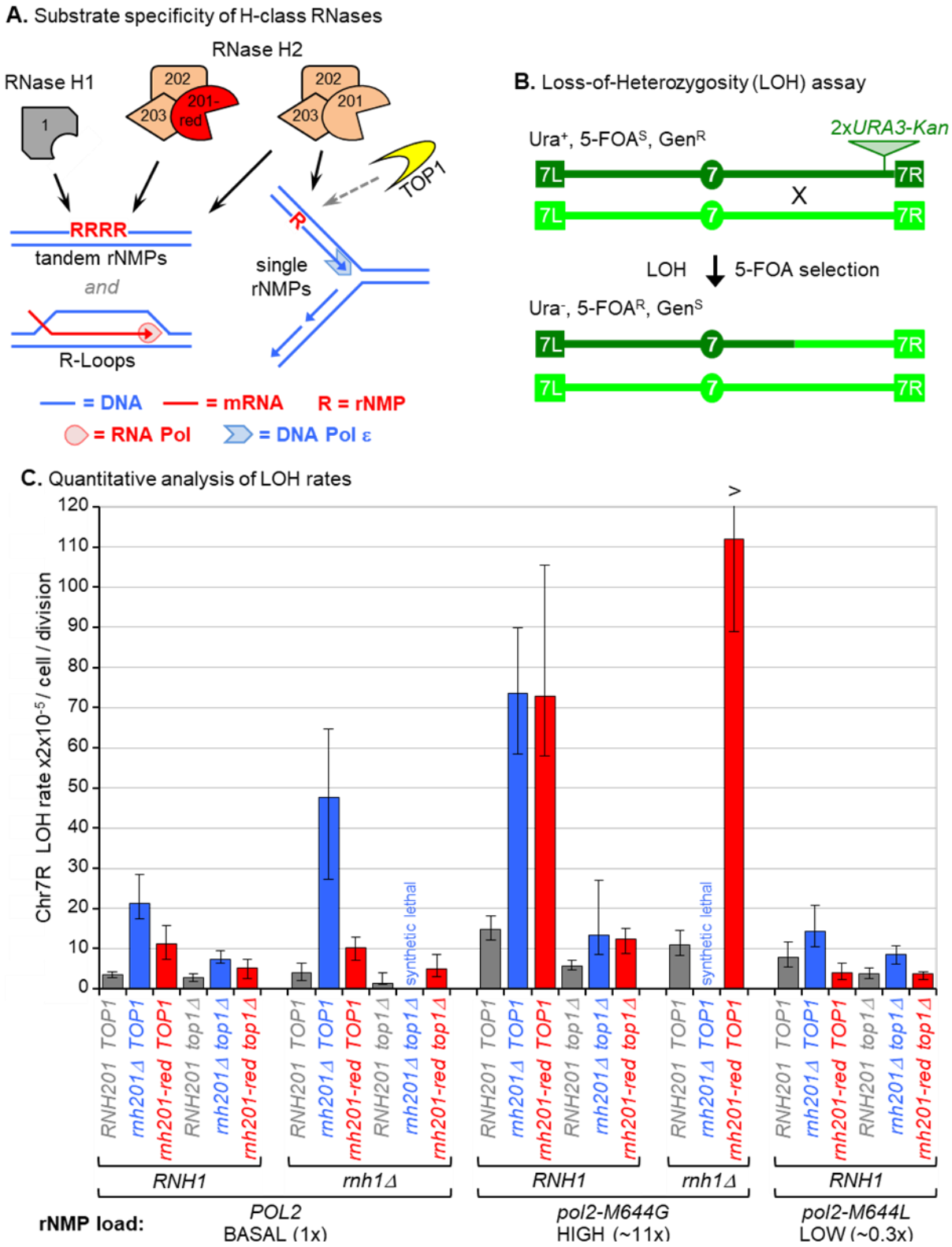


Figure 3.1 Substrates of H-Class RNases, experimental system, and LOH rate analysis.

Figure 3.1 Substrates of H-Class RNases, experimental system, and LOH rate analysis.

(A) Schematic representation of the substrate specificity of H-Class RNases. The black arrow lines link the various enzymes to their RNA-DNA substrates. The gray dashed arrow line indicates that evidence of topoisomerase 1 cleavage at ribonucleotides has only been observed in the absence of functional RER. (B) Depiction of the hemizygous chromosomal configuration used in the LOH assay. One of the homologs of Chr7 (dark green) has an insertion of the counter-selectable CORE2 cassette (*$_{Kl}$URA3-$_{Sc}$URA3-KanMX4*) (ZHANG *et al.* 2013a) downstream of the *MAL13* gene (distal side), ~20 kb from the right telomere (7R); the second homolog (light green) does not. A DNA lesion on the right arm of the dark green homolog may initiate an allelic mitotic recombination event leading to LOH, making the distal portion of the chromosome homozygous for the light green DNA sequence, and thus rendering that cell resistant to 5-FOA. (C) Quantitative analysis of LOH. The columns represent the median Chr7 right arm LOH rate for each genotype, and the error bars represent 95% confidence intervals (CI). The same data are presented in numeric form in Table S3.2, and statistical significance of pairwise comparisons are available in Table S3.3. All genotypes indicated in the X-axis are homozygous in the experimental diploids. The *greater than* (>) symbol indicates that the Y-axis was cropped to save space and to facilitate visualization of differences between the lower rates. The upper 95% CI limit of the *pol2-M644G rnh1Δ rnh201-RED TOP1* LOH rate was 182.48 x2x10^{-5}/cell/division. Intentional gaps were left in the data columns for the *POL2 rnh1Δ rnh201Δ top1Δ* and the *pol2-M644G rnh1Δ rnh201Δ TOP1* genotypes to emphasize the fact that these triple mutant combinations are synthetic lethal as reported previously (EL HAGE *et al.* 2010; LAZZARO *et al.* 2012; WILLIAMS *et al.* 2015).

CHAPTER FOUR: A NEW EXPERIMENTAL SYSTEM TO STUDY MEIOTIC NON-ALLELIC HOMOLOGOUS RECOMBINATION IN *S. CEREVISIAE*[4]

Summary

In humans, *de novo* recurrent copy number variations (CNVs) arise during meiosis from non-allelic homologous recombination (NAHR) between low copy repeat elements (LCRs). These chromosomal rearrangements represent a significant source of genetic variation and are responsible for a wide variety of genomic disorders. However, the precise factors that steer cells toward this detrimental recombination pathway are not well known. To create a model for investigation of LCR-mediated CNV pathways, we developed a diploid prototype experimental system in *Saccharomyces cerevisiae* by modifying the right arm of chromosome 5 (Chr5) through the introduction of engineered LCRs: duplicated 5 to 35 Kb segments of yeast DNA flanking single copy interstitial spacers, which simulate the meiotic NAHR substrates that exist in humans. Within the interstitial spacer are allelic insertions of phenotypic markers. The segregation of these markers in the haploid cell progeny was used to identity and classify recurrent CNV events. This system allowed us to not only measure the effects of LCR size on overall frequency of *de novo* meiotic recurrent CNVs, but also to determine the relative occurrence of each of the unique NAHR classes: interchromosome, interchromatid, and intrachromatid. The rate of CNV increased as the LCRs became larger, and this increase remained biased toward interchromosomal NAHR. We show that this experimental system directly mimics the features of *de novo* recurrent CNVs reported in human disease, suggesting that it has great potential to become a valuable tool for the discovery and characterization of cellular and environmental factors that control CNV formation.

[4]A version of this chapter will be submitted for publication.

Introduction

Genomic disorders are diseases caused by structural rearrangements of the human genome (LUPSKI 1998). Such rearrangements are often the result of meiotic non-allelic homologous recombination (NAHR) leading to *de novo* recurrent copy number variation events (CNVs) in which a disorder gene or locus is duplicated or deleted (CARVALHO and LUPSKI 2016; CONOVER and ARGUESO 2016). The majority of meiotic homologous recombination (HR) occurs between allelic sequences of properly aligned chromosomes, and does not form structural rearrangements (Fig 4.1 A). However, occasionally meiotic HR can occur between non-allelic substrates when large regions of highly homologous DNA, such as low copy repeats (LCRs), cause chromosomes to misalign (Fig. 4.1B-E) (SASAKI *et al.* 2010; LIU *et al.* 2012; KIM *et al.* 2016). There are three modalities through which meiotic NAHR between directly-oriented paralogous LCRs can occur: between homologous chromosomes (interchromosome), between sister chromatids of the same chromosome (interchromatid), and within one chromatid (intrachromatid) (Fig. 4.1 B-D, respectively) (SASAKI *et al.* 2010; LIU *et al.* 2012).

It has long been known that the presence of LCRs, large (>10Kb), repeated, regions of highly homologous DNA (>97% identity), mediate NAHR leading to CNVs (HAREL and LUPSKI 2018). These rearrangements have been shown to be important in primate evolution, human diversity, and genomic disorders (LUPSKI 1998; STANKIEWICZ *et al.* 2004; CARVALHO and LUPSKI 2016; CONOVER and ARGUESO 2016). Early studies established the role of LCRs flanking critical genomic disorder loci in recurrent *de novo* CNV formation (LUPSKI 1998; LUPSKI 2009; CARVALHO and LUPSKI 2016). More recently, variations in the size, distance, and identity between LCRs were found to influence the frequency of NAHR leading to *de novo* CNV events (LIU *et al.* 2011b; VERGES *et al.* 2017). Additional studies determined a role of variation in meiotic hotspot proteins such as PRDM9 (PRDM9 [MIM 609760]) in recurrent *de novo* CNV formation (BOREL *et al.* 2012). The variability between LCR size, distance, and identity as well as meiotic recombination hotspots suggest possible mechanisms for inter-individual NAHR risk (Fig 4.1 E).

In 2011 the Lupski group used the reciprocal recurrent CNV syndromes Smith-Magenis microdeletion syndrome (SMS [MIM 182290]) and Potocki-Lupski microduplication syndrome (PTLS [MIM 610883]) that span the *RAI1* gene (RAI1 [MIM 607642]) on chromosome 17p11.2, to investigate the effects of LCR size and distance on recurrent CNV formation (LIU *et al.* 2011b). This region is a paradigm for the investigation of patterns of recurrent CNV because both the reciprocal duplication and deletion cause a disease phenotype, the region is flanked by multiple pairs of LCRs that vary in size and distance to each other, and a relatively large cohort of cases was available for *de novo* CNV breakpoint characterization (LIU *et al.* 2011b). By analyzing the size of the LCR homologies and the distance between them in patients with either the recurrent *de novo* deletion (SMS) or duplication (PTLS), they were able to demonstrate a strong positive correlation between LCR length and CNV occurrence (R^2=0.85627) (LIU *et al.* 2011b). The results from this study were remarkable, but remained restricted by the constraints of patient cohort research: limited cohort size and potential phenotypic variability introduced the possibility of a skewed patient representation.

Although it is clear that LCR architecture and recombination hotspots are integral in NAHR frequency, in 2014 a striking and elegant sperm study conducted by MacArthur *et al.* revealed that additional factors may be modifiers of NAHR frequency (MACARTHUR *et al.* 2014). They showed that NAHR frequency was variable within a cohort with similar LCR architecture at the *CMT1A-REP* locus. Strikingly, while the frequency of NAHR was variable across individuals, it was highly correlated between monozygotic twins. In addition, genotypes at PRDM9 and SNPs associated with allelic recombination could not explain the NAHR frequency variability they observed. The results of this study indicated that there are likely additional undiscovered genetic and environmental factors involved in inter-individual NAHR frequency variation.

The combined results over the past decade have implicated LCR size, distance, and identity as well as recombination factors such as PRDM9 in NAHR risk leading to *de novo* recurrent CNVs and genomic disorders. Additionally, there is evidence for undiscovered genetic and environmental factors contributing to heritable inter-individual NAHR risk (Fig. 4.1E) (MACARTHUR *et al.* 2014; LUPSKI 2015b; CONOVER

and ARGUESO 2016; HAREL and LUPSKI 2018). Factors that influence the mechanisms of global NAHR irrespective of CNV locus, such as enzymes involved in meiotic recombination and chromosome pairing, are potential candidates, but have not yet been experimentally tested. To fully understand which genes and environmental factors are involved in NAHR risk, an assay able to mimic the meiotic recurrent NAHR found in human disease is of utmost importance (LUPSKI 2015b; YAUK *et al.* 2015a; CONOVER and ARGUESO 2016).

In this report, we developed a model assay system to experimentally test the role of LCR size on meiotic *de novo* recurrent CNV formation, which can later be applied to interrogate other aspects of the meiotic NAHR mechanism, including genetic and environmental pre-dispositions. We show that this system allows accurate phenotypic identification of cells carrying *de novo* duplications and deletions mediated by meiotic NAHR. In addition, we showed experimentally that LCR size is indeed strongly correlated to CNV formation frequency (R^2=0.9558, Fig. 4.2F-G), directly paralleling the SMS/PTLS study by Liu *et. al* (LIU *et al.* 2011b). Together these results validate the use of this yeast model as a germane approach to experimentally investigate the fundamental mechanisms that govern meiotic *de novo* recurrent CNV formation, including those that may affect inter-individual NAHR frequencies in humans.

Results and Discussion

Our meiotic CNV assay system consisted of a series of *Saccharomyces cerevisiae* strains harboring a chromosomal locus with engineered LCRs of varying sizes and phenotypic markers that allowed detection of recurrent CNVs produced by meiotic NAHR (Fig. 4.2, Table 4.1). We started with a haploid strain containing a 59.6 Kb tandem segmental duplication mediated by mitotic NAHR between two Ty1 retrotransposon element insertions (FCR8 clone) (Fig. 4.2 A) (STANTON 2012). We knocked out the proximal portion of the proximal duplicated segment of FCR8, resulting in two identical 35 Kb, directly-oriented repeats (LCRs), separated by a 12 Kb single copy interstitial spacer (IS) sequence containing the *SFA1*^{V208I}*-CUP1-Kan*MX4 copy number reporter cassette (Fig. 4.2 A-B). The 35 Kb LCRs correspond to regions 2-5 in Fig. 4.2 A-B, and the single copy IS corresponds to region 1.

We validated the structure of the 35 Kb LCR strain through pulse field gel electrophoresis (PFGE) and array CGH (Fig. 4.2 C). PFGE revealed a downward shift of chromosome 5 in the 35 Kb LCR strain when compared to FCR8, corresponding to loss of one copy of region 1 (Fig. 4.2 C). Array CGH further confirmed regions 1-5 were duplicated in the FCR8 starting strain, and in the 35 Kb LCR strain region 1 reverted back to single copy while regions 2-5 remained duplicated (Fig. 4.2 C).

We then switched the mating type of the haploid containing 35 Kb LCR Chr5, and swapped *Kan*MX4 for *Hph*MX4. The *Kan*MX4 and *Hph*MX4 markers confer resistance Geneticin (Gen) or Hygromycin B (Hyg), respectively, in a binary way. The *SFA1*^{V208I}*-CUP1-Kan*MX4 is a dosage sensitive reporter whereby having a duplication of the *SFA1*^{V208I}*-CUP1* locus allows for growth on media containing high concentrations of copper sulfate (Cu) and formaldehyde (FA). Cells carrying the parental configuration of Chr5, with a single copy of the reporter are resistant to intermediate Cu+FA concentrations, while cells with a deletion of the reporter are sensitive. These different growth patterns can be used to determine the copy number of the reporter cassette, and to detect CNVs (ZHANG *et al.* 2013b). Finally, the haploids were mated to each other to create a diploid which was homozygous for the 35 Kb LCRs and the *SFA1*^{V208I}*-CUP1* copy number reporter within the IS region, but hemizygous for either *Kan*MX4 or *Hph*MX4 (Fig. 4.1A-D).

We induced meiosis of the diploid 35 Kb LCR strain, leading to the formation of the characteristic yeast tetrad asci that contain the four haploid progeny of a single meiotic division. These haploid spores are analogous to human sperm or eggs, with the advantage that they remain together inside the yeast ascus, allowing sibling cells to be recovered and studied in the context of a single meiosis. The assay was designed so that allelic recombination and each of the NAHR categories conferred a different combination of Cu+FA, Gen and Hyg resistance and sensitivity between sister cells from the same meiotic division. Three hundred and twenty-three tetrads were dissected and each haploid spore was allowed to germinate and grow mitotically into a colony in which the presence of CNV was assessed phenotypically. Overall spore viability was high (90.2%), yielding 245 tetrads in which the complete set of four sibling spores germinated and formed a colony. Normal allelic recombination produced normal

cells with parental phenotypes that expressed one of each drug resistance marker (never both together), and intermediate Cu+FA resistance (Fig. 4.1 B-D). For the purpose of this assay, we diagnosed CNVs as the loss or gain of the reporter cassette where a loss leads to the inability to grow on all selective media and a gain leads to the ability to grow on media containing high Cu+FA concentrations (Fig. 4.1 B-D, Fig. 4.6). Interchromosome NAHR produced tetrads with a pair of normal parental phenotype cells displaying either parental drug resistance and resistance to intermediate Cu+FA concentrations, one cell carrying a duplication and displaying double drug resistance and hyper-resistance to Cu+FA, and one cell carrying the reciprocal deletion with double drug sensitivity and hyper-sensitivity to Cu+FA (Fig. 4.1B). As an additional phenotype, deletion of the *COX15* gene present in the IS, rendered cells unable to grow on media containing a non-fermentable carbon source (YPGE) (pink arrow, Fig. 4.2 A). Interchromatid NAHR produced tetrads containing a pair of normal parental phenotype cells with single drug resistance of the same type, one cell containing the duplication with hyper-resistance to Cu+FA and single drug resistance opposite of the two parental phenotype cells, and one cell containing the reciprocal deletion with sensitivity to both drugs, inability to grow on YPGE, and hyper-sensitivity to Cu+FA (Fig. 4.1C). Finally, intrachromatid NAHR produced tetrads containing a pair of normal parental phenotype cells with single drug resistance of the same type, another parental phenotype cell with single drug resistance of the opposite type, and one cell containing a deletion leading to sensitivity to both drugs, inability to grow on YPGE, and hyper-sensitivity to Cu+FA (Fig. 4.1D). The reciprocal circular fragment formed by intrachromatid NAHR does not contain any origins of DNA replication or a centromere, thus it was not propagated and did not contribute to phenotype (Fig. 4.1D).

To validate our phenotypic analyses, we randomly selected tetrads from which CNVs had been phenotypically called interchromosome, interchromatid, or intrachromatid, and analyzed colonies from all four sibling spores via PFGE, array-CGH, and digital droplet PCR (ddPCR) (Fig. 4.3). The PFGEs of each NAHR class are shown in figure 4.3A. As expected, the interchromosome and interchromatid classes had two colonies with parental sized Chr5, one contained a longer Chr5 due to a duplication, and one contained a shorter Chr5 due to a deletion (Fig. 4.3 A). The intrachromatid class showed three

colonies contained parental sized Chr5, and one contained a deletion within Chr5, without a reciprocal duplication (Fig. 4.3 A). We analyzed one normal, duplication, and deletion containing colony from one interchromosome CNV via array CGH and confirmed the presence of the correct Chr5 signal predicted by phenotype and PFGE (Fig 4.3 B). The normal colony had a single copy of the IS region 1 and two copies of the LCR regions 2-5, the duplication colony had two copies of the IS region 1 and three copies of the LCR regions 2-5, and the deletion colony had lost all signal from the region 1 probes (zero copies of the IS region 1) and only one copy of the LCR region 2-5 (Fig. 4.3 B). Finally, we also performed quantitative digital droplet PCR using primers for the *SFA1*, *Kan*MX4, *Hph*MX4 markers- all of which were inside the single copy IS region 1, and for a region proximal to the proximal LCR predicted to not be involved in the NAHR event, which was used as the reference for copy number normalization (Fig. 4.3 C). We found that in all normal spores the proximal region, *SFA1*, and either *Kan*MX4 or *Hph*MX4, but never both together, existed in the expected one copy (Fig. 4.3 C). All deletions revealed a single copy of the proximal region, and loss of all signal from IS ddPCR markers (Fig. 4.3 C). The interchromosome duplication revealed duplication of the *SFA1* marker, and the presence of one copy of each *Kan*MX4 and *Hph*MX4 (Fig. 4.3 C). In contrast, the interchromatid duplication revealed duplication of the *SFA1* marker and of the *Kan*MX4 marker, as expected from a Gen resistant interchromatid duplication (Fig. 4.3 C). Taken together, these analyses confirmed that for each meiotic NAHR class the growth phenotypes of the spores directly reflected what was observed in the genome (Chr5 copy number gain or loss of the IS and LCR regions). No additional chromosomal rearrangements were detected elsewhere in the genome.

It is well established that there is an intrinsic bias to use the homolog (interchromosome) for repair of double strand breaks during allelic meiotic recombination. This is believed to be necessary for establishing physical connections between homologs (*i.e.*, chiasmata) which are essential for proper chromosomal disjunction of during the first meiotic division (SCHWACHA and KLECKNER 1997). However, it is unclear whether this bias is also maintained during non-allelic meiotic recombination, for example between LCRs. This is an extraordinarily difficult problem to study in *de novo* recurrent CNV human genomic disorders, as it requires phased haplotype information from the chromosomes of the

parents and of the affected patient. In one example, when haplotype data at the 16p11.2 locus was restricted to high confidence haplotypes, there appeared to be a trend towards the presence of an interchromosome NAHR bias (DUYZEND *et al.* 2016). However, when all phasing data was combined, the 16p11.2 interchromosome NAHR trend was abolished (DUYZEND *et al.* 2016). We investigated the presence of a bias within our system by comparing the observed frequency of tetrads containing a CNV from each NAHR category to the frequency expected if the selection of a recombination donor sequence were random. If no bias were present, the interchromosome NAHR frequency should be 50% (two non-allelic recombination donors available in the homolog). Interchromatid and intrachromatid NAHR frequency should each be 25% (one non-allelic donor available in each sister chromatid). Instead, in the 35 Kb strain we measured interchromosome NAHR frequency at 63.6%, interchromatid at 22.1%, and intrachromatid at 14.3% (Fig. 4.3 A, Table 4.1). This observed distribution (49 interchromosome: 17 interchromatid: 11 intrachromatid) was statistically different from the expected distribution if NAHR between any of the non-allelic LCR donor sequences was equally likely to occur (X^2= 6.6623, p-value = 0.03575). This result suggests that, at least in this system and for the 35 Kb LCR configuration, NAHR interactions are strongly biased toward interchromosome at the expense of intrachromatid events.

There is evidence for rare *de novo* recurrent tandem triplication events of dose-sensitive genes in human genomic disorders (LIU *et al.* 2011a; HAREL and LUPSKI 2018). Interestingly, we were able detect phenotypically five examples of double-NAHR leading to triplication events within single meiotic divisions. Those tetrads contained spores carrying one normal copy of Chr5, two deletions, and one triplication. We characterized one of these tetrads by PFGE and array-CGH (Fig. 4.4, Table 4.1) to confirm the structural rearrangements. Figure 4.4 A shows a PFGE of sibling spores from a single tetrad in which one contained a parental sized Chr5, one contained a much longer Chr5 indicating a triplication CNV, and two contained a deletion on Chr5. We further confirmed this triplication through array-CGH which showed the triplication contained two copies of the IS region 1, and four copies of the LCR regions 2-5 (Fig. 4.4 B-C). Additionally, we identified three tetrads with complex patterns of marker segregation

consistent with pre-meiotic mitotic chromosomal rearrangements (Table 4.1). Even though these events were detectable, their frequency was low (<1%), therefore they were not likely to have interfered with the meiotic NAHR frequency measurements.

These combined results demonstrate that our experimental model system allows for faithful phenotypic identification as well as classification of recurrent CNVs of all three meiotic NAHR modalities. In addition, this initial analysis suggested that meiotic NAHR is likely subject to the same interchromosome bias that is well described for allelic HR, and also showed that it is suitable for the detection of complex and rare recombination events such as triplications and pre-meiotic mitotic CNVs.

As an initial use of this experimental approach, we decided to revisit the question of whether the size of LCRs can influence CNV frequency, and if so, if a correlation could be traced analogously to that shown to exist at the human 17p11.2 locus for the formation of CNVs that cause SMS and PTLS (LIU *et al.* 2011b). To do so, we used a similar chromosome engineering approach to the one used to create the 35 Kb LCR strain. In this case we altered the size of the proximal deletion of the FCR8 segmental duplication, which allowed for the creation of varied LCR sizes, while maintaining a constant 52.9 Kb distance between homologies (Fig. 4.2 B). The meiotic recombination hotspots in *S. cerevisiae* have been thoroughly mapped via sequencing of DNA fragments bound to the conserved meiosis-specific meiotic recombination initiator, Spo11 (PAN *et al.* 2011). Using these data, we were able to design LCR strains such that they would contain at least one predicted meiotic DSB hot spot (Fig. 4.2 A). Though each LCR also contained a 6 Kb Ty1 element at their distal end, it is important to note that the recombination properties of yeast Ty1 elements have been extensively studied and shown to be repressed for meiotic recombination (KUPIEC and PETES 1988). Accordingly, we saw very low CNV in a control strain that contained only the Ty1 elements but lacked engineered LCRs (<1%). Since the Ty1 insertions behaved as essentially inert sequences, each LCR-containing strain was identified by the length of their respective unique DNA (regions 2-5), but not including the length of the Ty1 element. We produced four additional experimental strains through this approach: 0 Kb (control strain; Ty1 only), 5 Kb, 15 Kb, and 24 Kb LCRs (Fig. 4.2 B-C). We then switched the mating type and drug resistance marker of each of the LCR

containing haploids and mated them to create diploids which were homozygous for each LCR and the *SFA1*^{V208I}*-CUP1* copy number reporter, but hemizygous for either *Kan*MX4 or *Hph*MX4.

Approximately 300 tetrads were dissected for each new LCR strain and were phenotypically scored for the presence of CNV via the *SFA1*^{V208I}*-CUP1* dosage dependent reporter, and the *KanMX4* and *HphMX4* markers (Fig. 4.1A-D, 4.6). The large proximal deletions required to create the smaller LCRs had the consequence of moving genes essential for yeast viability into the IS region (black arrows in regions 2, 4, 5, Fig. 4.2 A). Therefore, unlike the 35 Kb LCR described earlier in this study, deletion of the IS region in these new strains lead to inviability of spores carrying the deletion CNVs (Fig. 4.2A-B). Due to inability to recover deletions, in these strains we were unable to positively call the intrachromatid NAHR class and therefore diagnosed only duplication CNVs as the gain of the reporter cassette leading to the ability to grow on Cu+FA media. However, we were still able to differentiate between the two duplication NAHR modalities, based on the presence of double drug resistance (Gen and Hyg) for interchromosome, versus single drug resistance (Gen or Hyg) for interchromatid (Fig. 4.1 B-C). We performed karyotypic validation of our phenotypic analyses by randomly selecting interchromosome and interchromatid duplication events from our 15 Kb LCR strain and analyzing them via PFGE and ddPCR (Fig. 4.5). Although deletion events were inviable and thus were not recovered from this strain, we did observe the expected upward shift of the Chr5 band in both the interchromosome and interchromatid duplication containing colonies, indicative of duplication CNVs (Fig. 4.5 A). Additionally, ddPCR revealed a single copy of the proximal, *SFA1* and *Kan*MX4 or *Hph*MX4 markers in all parental phenotype colonies and a duplication of *SFA1* in both duplication colonies, with a single copy of *Kan*MX4 and *Hph*MX4 in the interchromosome duplication, and a duplication of *Hph*MX4 in the interchromatid duplication (Fig. 4.5 B). The molecular analyses of examples from both meiotic NAHR categories again confirmed that the drug resistance phenotypes paralleled what was observed in the genome (gain of the Chr5 CNV cassette and LCR), with no additional detectable gross chromosomal rearrangements (Fig. 4.5).

To interrogate the correlation between LCR size and CNV frequency, we calculated the total CNV frequency for each LCR size by adding spore interchromosome and interchromatid duplication frequencies. Similar to the Liu *et al.* 2011 SMS/PTLS findings, we saw a clear positive linear correlationbetween the size of the LCRs and the frequency of recurrent duplication CNVs (Fig. 4.1G, R^2=0.9558, p-value= 0.02235). These experimental results agree with clinical observations of larger LCRs significantly increasing the frequency of NAHR events.

The presence of deletion and duplication phenotypes of the 17p11.2 locus allowed Liu *et. al* to obtain a robust R^2 despite very few patient data points for the two uncommon recurrent CNV classes. Our study only recovered deletions from one LCR size and was therefore unable to calculate linear correlation between LCR size and deletion CNV. Nevertheless, our study produced more data points within each LCR size and therefore saw an even stronger positive linear correlation between LCR size and CNV frequency. Furthermore, our study was able to isolate LCR size as the only variable affecting NAHR, whereas Liu *et. al* were constrained by the fluctuating distance between homologies at the 17p11.2 locus.

Interestingly, increased frequency with increased LCR size applied to both interchromosome and interchromatid NAHR within our system. We presume that this tendency may also apply to intrachromatid NAHR, although our current system did not allow that classification. Our data showed an excess of interchromosome over interchromatid NAHR within each LCR population supporting the hypothesis that the homolog bias is maintained in all cases.

Although LCRs have been observed to play a large role in recurrent CNV formation, these observations had not previously been matched by experimental data, due to lack of an experimental system with which to perform such tests. We report here the development of a new prototype assay for detection of *de novo* meiotic recurrent CNVs produced by NAHR between engineered LCRs that simulate the NAHR that leads to human genomic disorders. Introduction of the *SFA1*^{V208I}*-CUP1* reporter cassette and selectable maternal and paternal markers allowed for the detection of single *de novo* meiotic recurrent copy number gains or losses of our reporter locus. In addition to detecting copy number gains and losses between LCRs, our system took advantage of the unique *S. cerevisiae* tetrads formed through meiotic

division which keep meiotic sister cells together, allowing dissection of the reciprocal events that occur during CNV formation. Our measurements of NAHR between LCRs indicate that CNV rates are influenced by the presence and size of LCRs, and that larger LCRs increase the likelihood of meiotic NAHR while maintaining the interchromosome bias. These results corroborate what has been observed in previous studies of human genomic disorders caused by meiotic recurrent CNVs, and suggest that normal variance in the size of genomic architecture elements between individuals can have significant impacts on rates of meiotic recombination. Additionally, these results establish our system as a pertinent model for formation of *de novo* recurrent CNVs frequently observed in human genomic disorders.

The results from this study show that meiotic NAHR functions similarly in yeast and humans, and provide a model system for further studies. In recent years the need for such an assay system has become undeniable, as evidence mounts that the factors involved in recurrent and non-recurrent CNV formation are likely different (CONOVER and ARGUESO 2016; COSTA *et al.* 2018). Therefore, we will now develop a second generation of this system to allow for further probing of environmental and genetic factors that may influence NAHR. This prototype along with our second generation system will provide a method for investigation of factors involved in meiotic NAHR such as LCR size, distance and identity, as well experimental interrogation of environmental and conserved genetic factors that may be involved in parental transmission of *de novo* recurrent CNVs to their offspring.

Materials and Methods

Yeast Strains and Plasmids:

All *Saccharomyces cerevisiae* strains used in this study were derived from CG379 strain background containing noted locus-specific changes introduced by PCR-based transformation (Fig. 4.2). The precursor strain for the engineered LCRs was the FCR8 clone, a copper and formaldehyde hyper-resistant strain that had a mitotically-derived tandem segmental duplication on the right arm of chromosome 5 (Chr5) mediated by recombination between *YERCTy1-1* and *YERCTy1-2*. FCR8 was

obtained by selection for Cu+FA resistance from haploid carrying the *SFA1*^{V208I}*-CUP1-Kan*MX4 reporter cassette inserted downstream of *DDI1* on Chr5 (Fig. 4.2 A-B) (STANTON 2012).

Through PCR-based, homologous recombination-mediated targeted deletion of the FCR8 proximal *SFA1*^{V208I}*-CUP1* cassette with the *Nat*MX4 cassette (GOLDSTEIN and MCCUSKER 1999), we were able to generate strains with one *SFA1*^{V208I}*-CUP1-Kan*MX4 cassette flanked by repeated regions of yeast DNA of varying sizes (Fig. 4.2). The homology segments used to target the knockout PCR products are shown in Fig. 4.2B (left column) and specific DNA sequences are shown in Table S4.2. By increasing the region knocked out by PCR, we were able to alter the size of the LCR inversely with the size of the interstitial spacer (IS) while keeping the distance between homologous regions constant. We used this method to create four haploid LCR containing strains: 1. 5 Kb LCRs flanking a 41 Kb IS with 52.9 Kb separating LCR homologies 2. 15 Kb LCRs flanking a 33 Kb IS with 52.9 Kb separating LCR homologies 3. 24 Kb LCRs flanking a 23 Kb IS with 52.9 Kb separating LCR homologies, and 4. 35 Kb LCRs flanking a 12 Kb IS with 52.9Kb separating LCR homologies (Fig. 4.2 B). We then swapped the mating type of the LCR containing haploids. Next, we used a PCR based transformation to swap the *Kan*MX4 Geneticin (Gen) resistance marker for *Hph*MX4 Hygromycin B (Hyg) resistance marker in one homolog. We then mated the opposite resistance marker containing strains so their diploid progeny were homozygous for the LCRs and the *SFA1*^{V208I}*-CUP1* reporter and hemizygous for either *Kan*MX4 or *Hph*MX4.

Culture media and CNV selection conditions:

Yeast diploid cells were induced to sporulate by first growing overnight 5 mL liquid pre-sporulation media cultures (8 g/L yeast extract, 3 g/L peptone, 100 g/L glucose, 10 g/L complete drop out mix, and 5g/L methionine). Next they were centrifuged, cell pellets were washed twice with sterile distilled water and half the culture was put into 4 mL of liquid sporulation media (1 g/L yeast extract, 10 g/L potassium acetate, 0.5 g/L glucose, 2.5 g/L complete drop out mix, and 3.8 g/L methionine). Cells were incubated in liquid sporulation media at 25C while shaking for four days. On the fourth day tetrads

were removed from the shaker and kept at 25C to be dissected for up to one week. Tetrads were dissected on YPD agar (rich medium) and incubated for 2 days at 30C. Plates were then replica plated onto YPD plus 400 mg/L Gen, 400 mg/L Hyg, YPGE, and SC supplemented with a complete drop-out mix and a range of concentrations of copper sulfate and formaldehyde (200 μM $CuSO_4$ / 2.3 mM FA, 250 μM $CuSO_4$ / 2.5 mM FA, 300 μM $CuSO_4$ / 2.7 mM FA). Cu+FA concentrations were optimized for tetrad replica plating based on parameters described earlier (ZHANG *et al.* 2013b). Formaldehyde-containing plates were poured fresh the day before they were used. 1 M dilutions [101.5 μL of 37% by weight Formaldehyde methanol stabilized stock (Fisher) in 1148.5 μL sterile water] of formaldehyde were made and placed opposite copper sulfate inside an Erlenmeyer flask immediately before media was added to the flask and plates were poured. Cells were incubated on YPD + Gen and YPD + Hyg plates at 30 C for 2 days, Cu+FA plates were incubated at 30 C for 4 days.

Analysis of NAHR Mediated CNVs:

To detect *de novo* meiotic recurrent CNVs caused by chromosome architecture we needed reporters capable of differentiating the expected single copy of our region of interest (akin to a human disorder locus) versus a copy number loss or gain. Recently, Stanton *et al.* improved a CNV assay allowing for detection of single copy increase of the *SFA1*^{V208I}*-CUP1* reporter cassette in mitotic cells, where a single duplication of the *SFA1*^{V208I}*-CUP1* reporter cassette allows for growth on media containing copper sulfate and formaldehyde (NARAYANAN *et al.* 2006; STANTON 2012; ZHANG *et al.* 2013b). Additionally, because NAHR can occur between LCRs on the homologous chromosome, the sister chromatid, or within the same chromatid (Fig. 4.1 B-D), detection of the maternal allele versus the paternal allele was essential for a comprehensive recurrent CNV reporter assay (LIU *et al.* 2012). By adding the either the *Kan*MX4 gene or the *Hph*MX4 gene to the *SFA1*^{V208I}*-CUP1* reporter, we were able to differentiate between which LCRs the CNV causing NAHR occurred based on growth on Gen or Hyg containing media, respectively

(Fig. 4.1 A-D). The possible phenotypic combinations of each rearrangement as well as Pulse Field Gel Electrophoresis and array-CGH of the Chr5 region for each type of NAHR event can be viewed in Figures 4.6 and 4.3, respectively.

Molecular Karyotype Analysis and Digital Droplet PCR:

Sister spores carrying NAHR mediated duplications and deletions of the *SFA1*^{V208I}*-CUP1* reporter cassette were analyzed via their growth patterns on the different media (Fig. 4.1 A-D, Fig. 4.6). Candidate sister spores from each CNV class were randomly selected for pulse field gel electrophoresis, digital droplet PCR and gene microarray to validate CNV events captured by differing drug resistance (Fig. 4.3, 4.5).

Full length chromosomal DNA was prepared in agarose and was separated based on length via Pulse Field Gel Electrophoresis (PFGE) to reveal size changes of Chr5. Genomic DNA was prepared from PFGE agarose plugs as previously described and used for comparative genomic hybridization microarrays (array-CGH) using both methods and microarray design described previously (ZHANG *et al.* 2013b). These DNA samples were also used for quantitative digital droplet (ddPCR). Sheared DNA was then digested with *Mfe*I-HF, a restriction enzyme that cuts Chr5 frequently and between LCRs, but did not cut between ddPCR primers. Digested DNA was diluted to 0.05 ng/μL and 2 μL of this dilution was used as a template for each ddPCR reaction. The diluted DNAs were analyzed using a combination of four ddPCR primer sets, each for a specific region of Chr5. Specific ddPCR primer sequences are shown in Table S4.2. Biorad's QX200 EvaGreen Supermix and protocol were used for all ddPCR reactions. For each experimental ddPCR reaction, the same DNA underwent ddPCR for a control single copy region directly proximal to the LCR. The signal from this proximal region was used to normalize the concentration of the DNA internally for each DNA template, allowing comparison of gene copy number with each experimental probe. One haploid parent was selected as a control for each ddPCR reaction and a no-DNA control was run with each experimental ddPCR reaction. Before normalization or standardization any residual fluorescence seen in the control was subtracted from all experimental values.

Statistics Employed:

We performed a linear regression analysis on our CNV frequency versus LCR size using the lm function in R studio version 3.4.0. CNV frequency was calculated as the total number of spores containing *SFA1*^{V208I}*-CUP1* duplications divided by the total number of living spores analyzed for each strain. Spore data did not include intrachromatid events as deletions were only called in the 35 Kb LCR strain.

Additionally, we performed a Chi-Squared Test for Given Proportions on the ratios of each NAHR modality used within the 35 Kb LCR strain using the chisq.test function in R studio version 3.4.0. Proportions were calculated as the number of complete tetrads containing an interchromosome, interchromatid, or interchromatid recombination event versus the total number of complete tetrads containing recombination events. The expected proportions were ½ interchromosome, ¼ interchromatid, and ¼ intrachromatid.

Figures

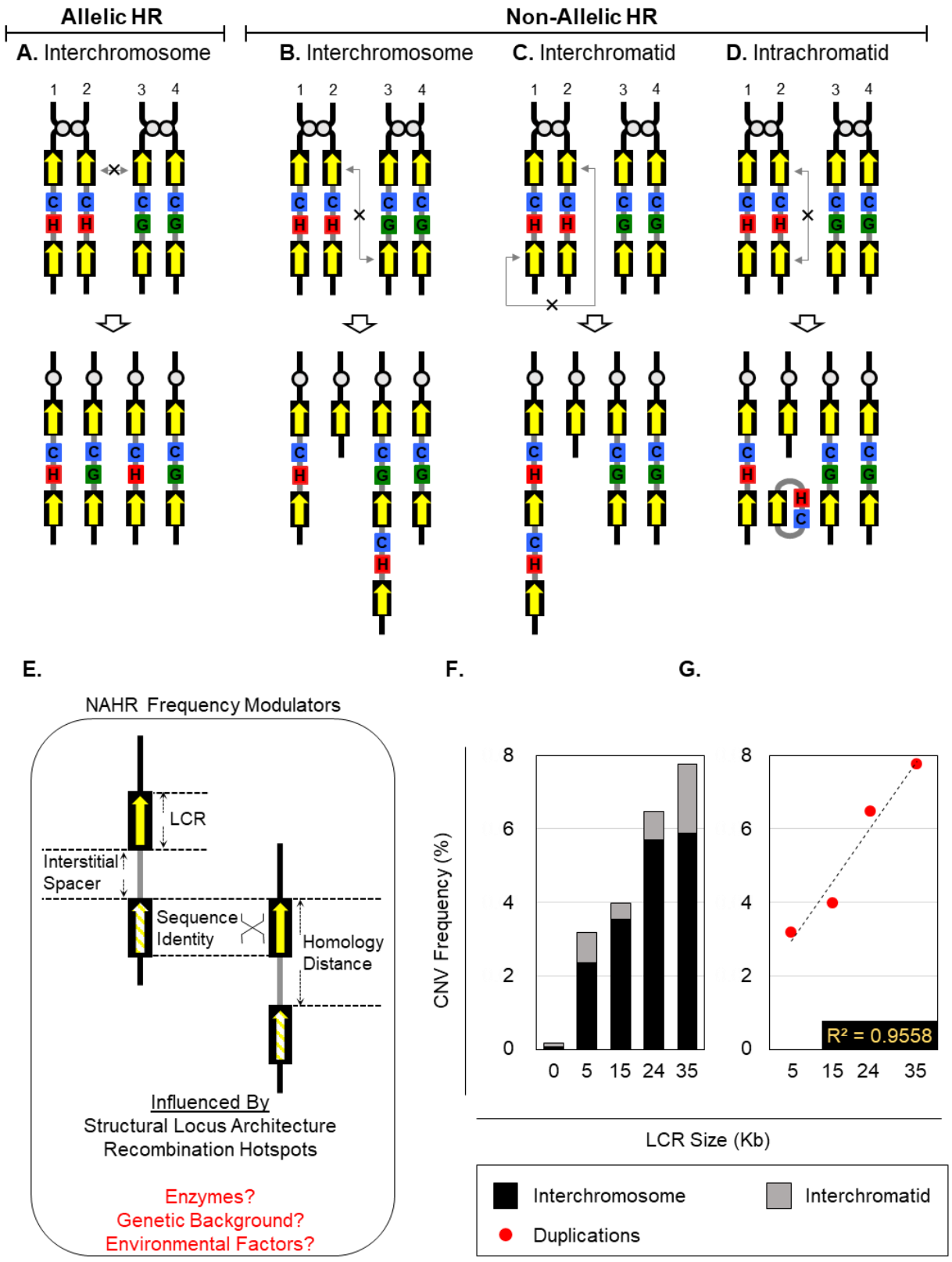
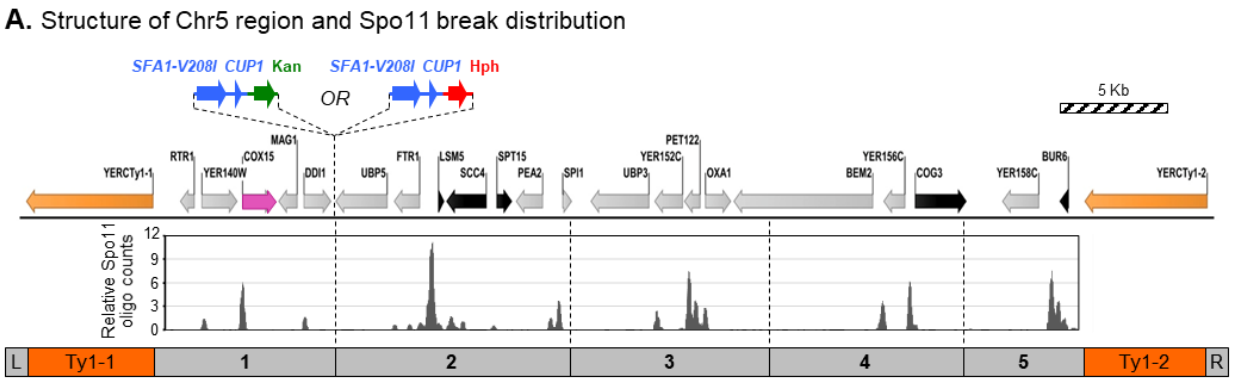
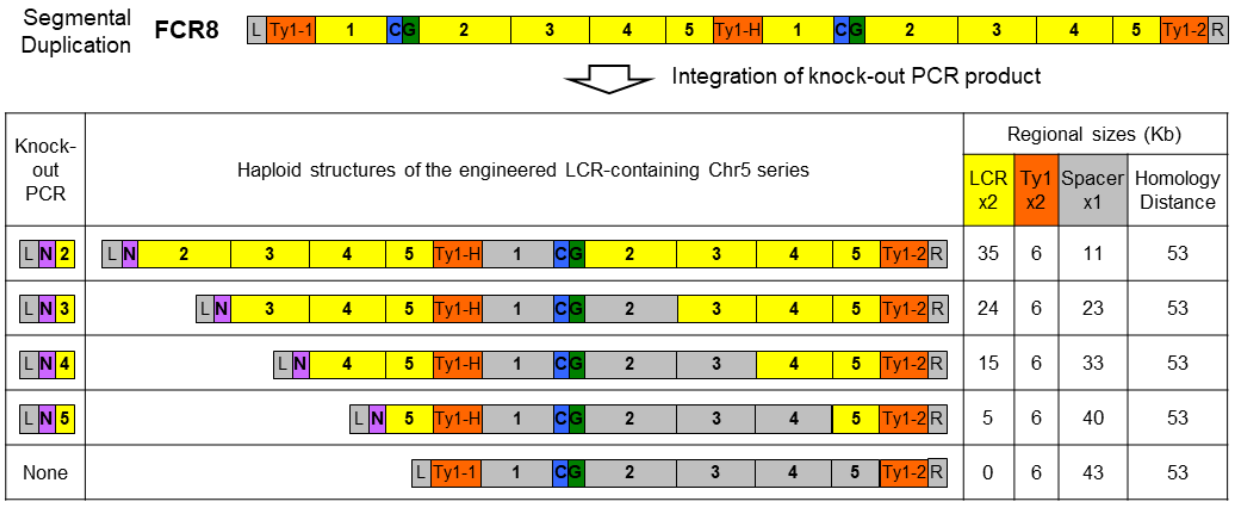


Figure 4.1 Phenotypic Analysis of Meiotic Recombination

Figure 4. 1 Phenotypic Analysis of Meiotic Recombination

(A-D) Schematic representation of the two replicated diploid Chr5 homologs at the start of meiotic recombination, not to scale. Four figures at the bottom of each panel represent four sister haploid cells after meiosis to be phenotypically examined. The red H represents the *Hph*MX4 gene conferring resistance to Hygromycin B, the green G represents *Kan*MX4 gene conferring resistance to Geneticin. The blue C represents the dosage dependent *SFA1*^{V208I}*-CUP1* reporter cassette represent which, when duplicated, confers hyper-resistance to copper and formaldehyde. Segregation patterns distinguish between different non-allelic crossover events represented in Panels B-D. **(A)** Non-crossovers and allelic crossovers produce four normal spores with parental phenotype. **(B)** Non-Allelic interchromosome crossovers produce two parental phenotype spores of opposite resistance, one duplication containing spore with double drug and copper formaldehyde resistance, and one deletion with no resistance **(C)** Non-Allelic interchromatid recombination produces two parental phenotype spores of the same resistance, one duplication containing spore single resistance opposite the two parental phenotype spores and copper and formaldehyde resistance, and one deletion with no resistance. **(D)** Non-Allelic intrachromatid crossovers produces three parental phenotype spores, one deletion with no resistance, and one acentric chromosome which will be lost. **(E)** Factors involved in NAHR. **(F)** Frequency of CNV events in each LCR population based on single spore analysis. **(G)** Linear regression of data from (F) where the X axis is LCR size and the Y axis is interchromosome+interchromatid duplication CNV frequency. Coordinates for (F) and (G) were calculated with data from Table 4.1.





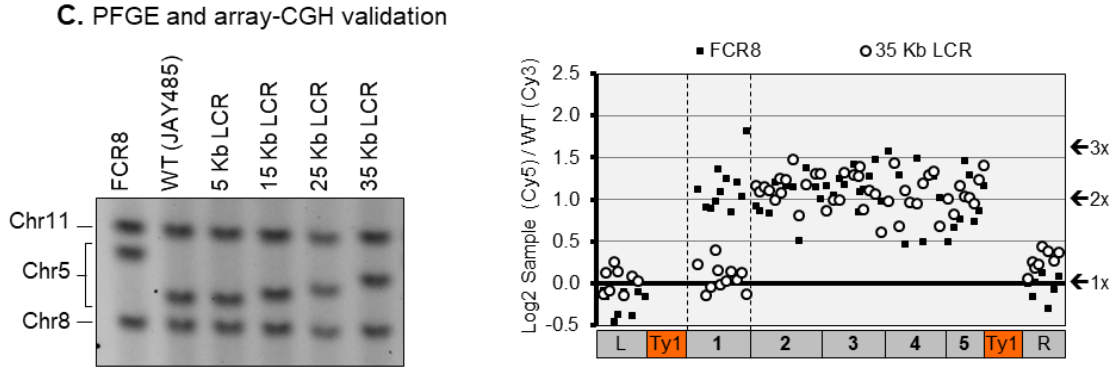


Figure 4. 2 Construction of LCR-containing *S. cerevisiae* strains

(A) Structure of FCR8 Chr5 region and *Spo11* break distribution. Orange arrows indicate meiotic recombination repressed Ty1 elements. Blue arrows indicate the *SFA1*^{V208I}*-CUP1* dosage dependent reporter cassette. Green and red arrows indicate Geneticin and Hygromycin B resistance genes, respectively. Black arrows indicate lethal deletion regions. Gray arrows indicate all other genes present in our experimental region. The pink arrow indicates the *COX15* gene. Peaks in the *Spo11* graph indicate regions where Spo11 is predicted to bind creating meiotic recombination hotspots in *S. cerevisiae*, data

from Pan *et. al* 2011 **(B)** Construction of LCR-containing *S. cerevisiae* Chr5 derivatives. Orange boxes indicate recombination repressed Ty1 elements, yellow arrows indicate LCR segments, gray boxes indicate the interstitial spacer area, blue boxes indicate the *SFA1*^{V208I}*-CUP1* dosage dependent reporter cassette, green boxes indicate resistance to Geneticin (homologs were resistant to Hygromycin B). The purple N indicates Neothrosin resistance which was used to knock out the proximal portion of the duplication region to produce smaller LCRs. L and R indicate the left and right orientation of the CNV region. LCR, Ty1, spacer, and homology distance are provided in Kb. **(C)** PFGE and array-CGH validation of LCR integration into Chr5 region. Chr11 and Chr 8 are also shown for comparison purposes.

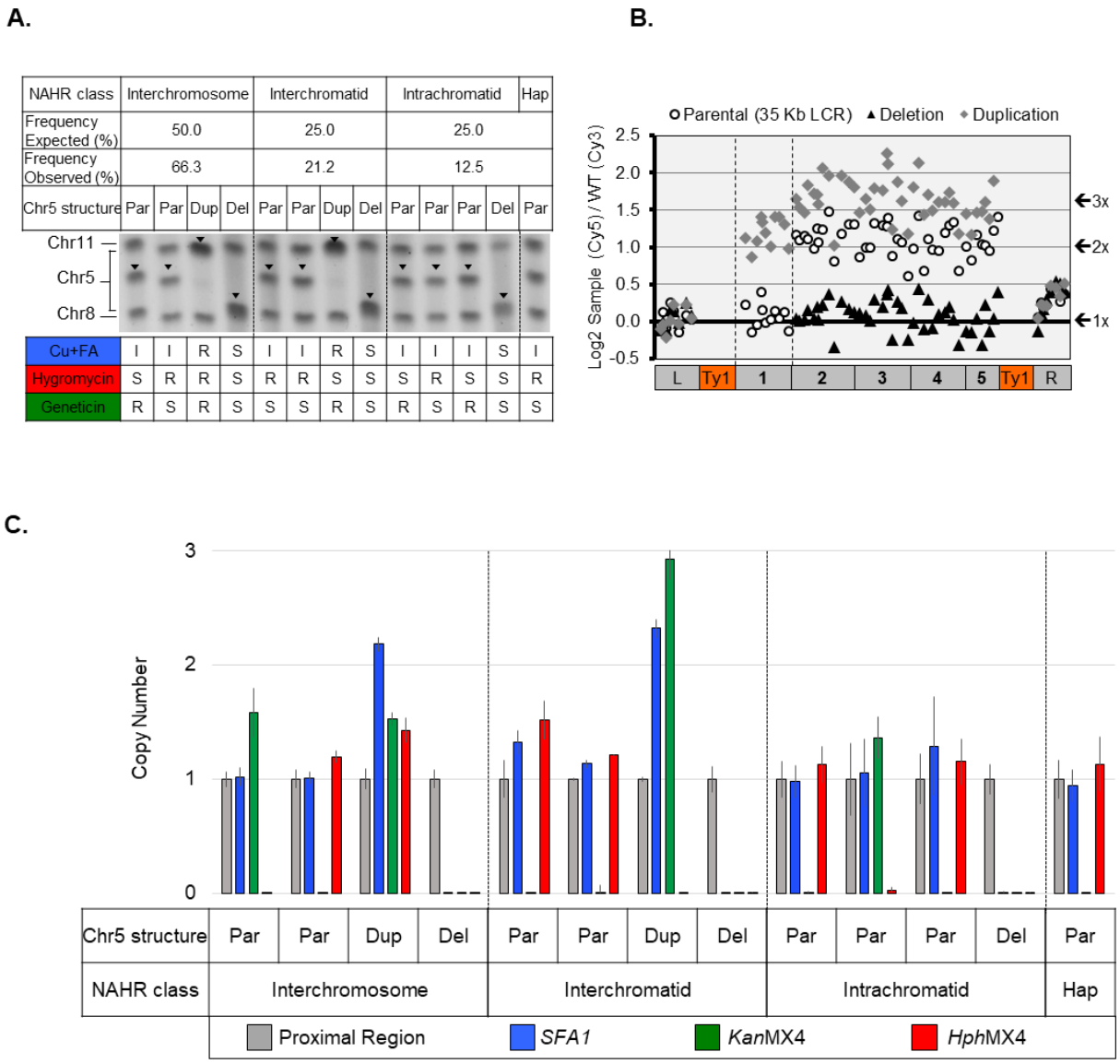


Figure 4. 3 Phenotypic Consequences of Meiotic NAHR: 35Kb LCR Strain

(A) Pulse Field Gel Electrophoresis of the four haploid spores from one tetrad from each class of NAHR. The frequency of tetrads detected within each class and the expected frequency if recombination partner selection were random are displayed above the gel. Phenotypic drug resistance of each spore is indicated below the pulse field lane, S=sensitivity to drug media and R=resistance, I= resistance to intermediate copper sulfate and formaldehyde concentrations. Par=parental phenotype and Chr5 size, Dup=duplication of the CNV region, Del=deletion of the CNV region. Black arrows indicate Chr5. **(B)** Array CGH of a parental phenotype spore, a duplication containing spore, and a deletion containing spore. **(C)** Graphical representation of digital droplet PCR results measuring normalized copy number of the *SFA1*^{V208I}, *Hph*MX4, and *Kan*MX4 reporter genes as well a proximal region not involved in the rearrangement used for normalization. Data corresponds to the spores represented in (A).

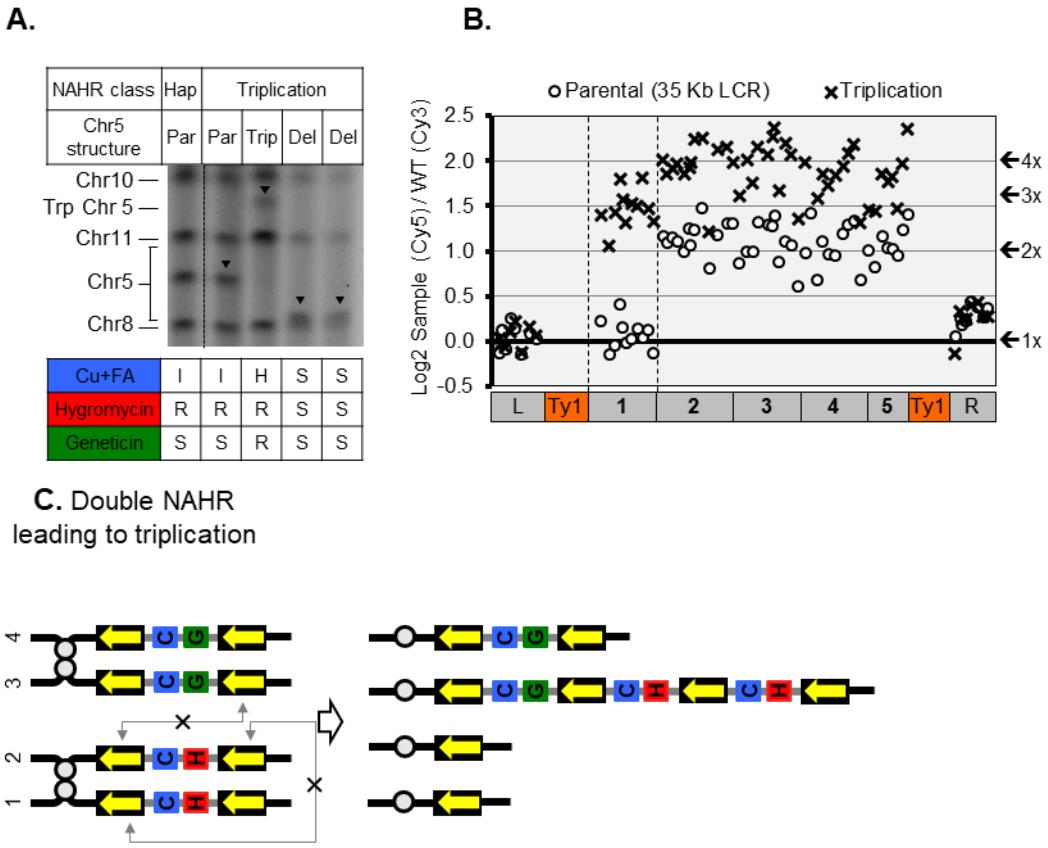


Figure 4. 4 Assessment of Triplication Event

(A) Pulse Field Gel Electrophoresis of four spores recovered from a triplication event. Black arrows indicate Chr5. Chr10, Chr11, and Chr8 are shown for comparison purposes. Par= parental sized Chr5, Trip= Chr5 containing the triplication event, Del= Chr5 containing a deletion. Phenotypic drug resistance of each spore is indicated below the pulse field lane, S=sensitivity to the drug media and R=resistance, I= intermediate resistance to copper and formaldehyde concentrations, H=hyper-resistance to copper and formaldehyde concentrations. (B) Array CGH shows a parental phenotype spore and a spore containing a triplication event. (C) Schematic representation of triplication event.

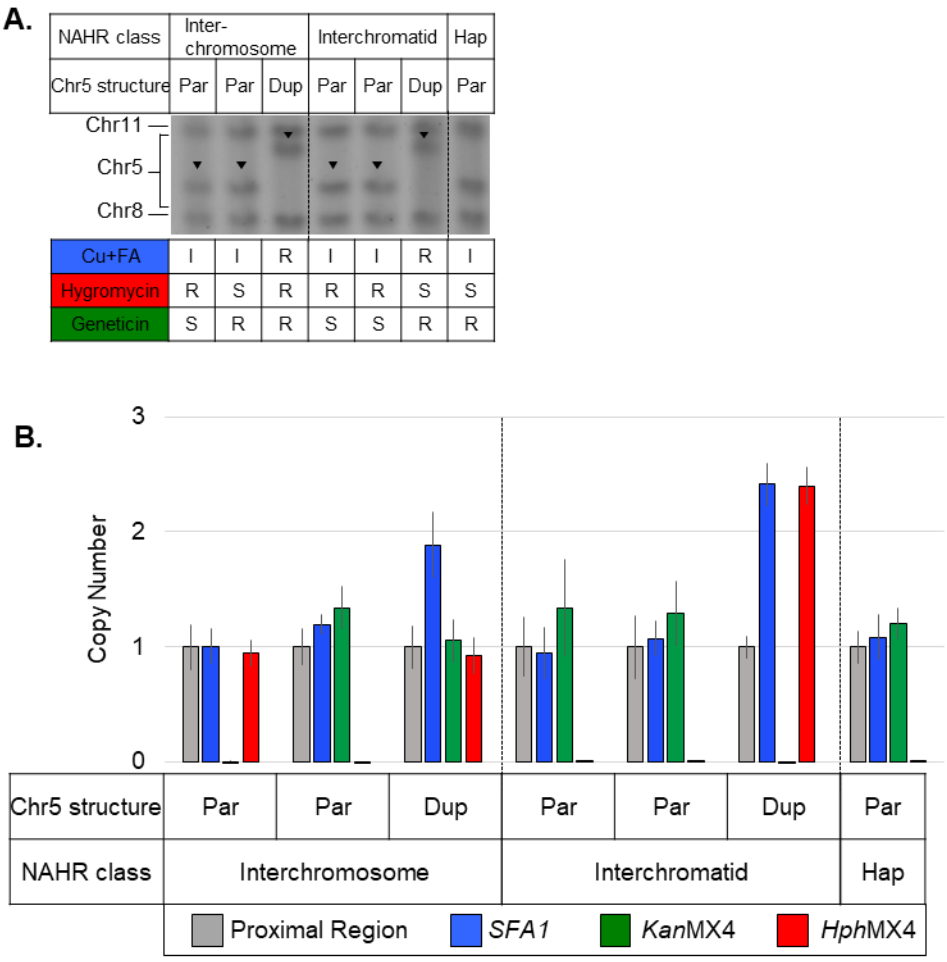


Figure 4. 5 Phenotypic Consequences of Meiotic NAHR: 15Kb LCR Strain

(A) Pulse field karyotypes showing representatives of interchromosome and interchromatid NAHR from the 15Kb LCR strain. Phenotypic drug resistance of each spore is indicated below the pulse field lane, S=sensitivity to the drug media and R=resistance, I=resistance to intermediate copper sulfate and formaldehyde concentrations. Par=parental phenotype and Chr5 size, Dup=a duplication of the CNV region. **(B)** Graphical representation of digital droplet PCR results measuring copy number of the *SFA1*^{V208I}, *Hph*MX4, and *Kan*MX4 reporter genes as well a proximal region not involved in the rearrangement used for normalization. Data corresponds to the spores represented in (A).

Media	Phenotype															
	Parental				Interchromosome				Interchromatid				Intrachromatid			
	P	P	P	P	P	P	Dup	Del	P	P	Dup	Del	P	P	P	Del
Cu+FA	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-
Hygromycin	-	-	+	+	-	+	+	-	-	-	+	-	-	-	+	-
Geneticin	+	+	-	-	+	-	+	-	+	+	-	-	+	+	-	-
YPGE	+	+	+	+	+	+	+	-	+	+	+	-	+	+	+	-
YPD	+	+	+	+	+	+	+	+/-	+	+	+	+/-	+	+	+	+/-

-OR-

Media	Phenotype															
	Parental				Interchromosome				Interchromatid				Intrachromatid			
	P	P	P	P	P	P	Dup	Del	P	P	Dup	Del	P	P	P	Del
Cu+FA	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-
Hygromycin	-	-	+	+	-	+	+	-	+	+	-	-	+	+	-	-
Geneticin	+	+	-	-	+	-	+	-	-	-	+	-	-	-	+	-
YPGE	+	+	+	+	+	+	+	-	+	+	+	-	+	+	+	-
YPD	+	+	+	+	+	+	+	+/-	+	+	+	+/-	+	+	+	+/-

Figure 4. 6 Phenotypic Analysis of Meiotic Recombination

Possible phenotypic patterns used to classify spores as parental, interchromosome, interchromatid, or intrachromatid events. Par= parental phenotype, Dup= duplication, Del= deletion.+ indicates growth, - indicates sensitivity, +/- indicates potential inviability.

Tables

Table 4. 1 CNV Counts, Frequencies, and Viabilities

	0Kb	5Kb	15Kb	24Kb	35Kb
Complete Interchromosome Tetrads	-	-	-	-	49
Complete Interchromatid Tetrads	-	-	-	-	17
Complete Intrachromatid Tetrads	-	-	-	-	11
Interchromosome Duplication Spores	1	26	39	58	69
Interchromatid Duplication Spores	1	9	5	8	22
Triplications	-	-	-	-	5
Mitotic Events	-	-	-	-	3
Complete Tetrads	-	-	-	-	245
Total Tetrads	308	310	318	299	323
Total Spores	1146	1098	1104	1019	1173
Spore Viability (%)	92.9	88.5	86.8	85.2	90.2
Frequency of Complete Tetrads with Interchromosome (%)	-	-	-	-	20.0
Frequency of Complete Tetrads with Interchromatid (%)	-	-	-	-	6.9
Frequency of Complete Tetrads with Intrachromatid (%)	-	-	-	-	4.5
Frequency of Spores with Interchromosome (%)	0.09	2.4	3.5	5.7	5.9
Frequency of Spores with Interchromatid (%)	0.09	0.8	0.5	0.8	1.9
Complete Tetrad CNV (%)	-	-	-	-	31.4
Total Spore CNV (%)	0.18	3.2	4.0	6.5	7.8

Coordinates for bars and linear regressions from Fig. 4.1 F,G are contained in this table.

CONCLUSIONS AND PROSPECTIVE DIRECTIONS

Endogenous sources of genomic instability are important risk factors for copy number variation. Just as pathogenic variants of genes are often found to cause genetic disease, copy number variation of genes involved in development and function can also have large impacts on health and disease. We chose to interrogate conserved cellular processes that contribute to endogenous genome instability. Specifically, we were interested in factors that could affect the frequency of copy number variation.

Both Ribonucleotide and R-loop incorporation into DNA affect chromosomal stability

Conclusions

We set out to answer three questions regarding the role of ribonucleotide incorporation on genome instability: Do ribonucleotides incorporated during nuclear DNA replication in RER-defective yeast strains elevate the rates of LOH and NAHR? If so, do elevated LOH or NAHR rates depend on ribonucleotides incorporated by Pol ε, Pol α, or Pol δ? Do elevated LOH or NAHR rates depend on *TOP1*? We demonstrated that ribonucleotides incorporated during DNA replication in RER-defective yeast strains do, in fact, elevate the rates of LOH and NAHR. Further, we found that this mutagenicity of ribonucleotide incorporation was asymmetric based on incorporation by Polε and dependent on inappropriate processing by topoisomerase 1. Finally, we demonstrated that both the ribonucleotide and R-loop removal activities of RNase H2 contribute significantly to chromosome stability.

Prospective Directions

RNase H2- R-loop Excision Deficient Mutant

Our studies interrogated chromosomal instability when both ribonucleotides and R-loops remained associated with the DNA as well as when ribonucleotides alone remained integrated in the DNA. Design of a separation of function allele that allows for proper ribonucleotide excision repair but

inability to repair R-loops would allow for further interrogation of the independent roles of R-loops and ribonucleotides independently on genome instability. Whole genome unselected LOH studies using the RNase H2-RED allele and an R-loop repair defective allele could provide an opportunity to test the hypothesis that R-loop removal occurs more frequently at highly transcribed regions and RER occurs at hotspots of ribonucleotide incorporation.

An experimental system to investigate *de novo* recurrent meiotic CNV

Conclusions

Within the introduction of this thesis three key questions regarding CNV formation were discussed: *"Do current mutagenesis assays measure CNV formation?; Can we design such an assay?; Are we introducing compounds into our environment that induce CNV mutagenesis?"* We provided the opinion that current mutagenesis assays only partly measure CNV formation due to the different mechanisms of recurrent and non-recurrent CNV formation. I set out to design an experimental assay system with which to study *de novo* recurrent meiotic CNV. To do this, I engineered LCRs flanking a dosage sensitive CNV reporter cassette into my experimental *S. cerevisiae* strain. Recapitulation of the known linear correlation between LCR size and recurrent CNV frequency demonstrated the ability of my system to faithfully mimic the recurrent CNVs seen in human disease. This correlation confirmed experimentally that larger LCRs promote more recurrent CNV events. Further, the segregation of the drug resistance cassettes allowed for classification of recurrent CNVs by the NAHR mechanism through which they were formed; interchromosome, interchromatid or intrachromatid. Importantly, the majority of NAHR events occurred between LCRs on homologous chromosomes. While an interchromosome bias is well known under normal meiotic conditions, it is unclear whether this bias persists under conditions of NAHR leading to CNV involved in human genomic disorders. The presence of an interchromosome bias within my experimental system suggests that this bias is maintained during NAHR, and provides an additional avenue to model human *de novo* recurrent CNV.

Prospective Directions

The final question discussed in my introduction was whether we are releasing CNV mutagens into the environment. Human sperm studies of *de novo* recurrent CNV frequency have indicated that certain individuals are more prone to meiotic NAHR than others (MOLINA *et al.* 2010; MOLINA *et al.* 2011; MACARTHUR *et al.* 2014; VERGES *et al.* 2014). However, there is currently no experimental system with which to investigate the factors behind interindividual risk through screens of broad panels of gene and environmental candidates. The development of my meiotic NAHR CNV assay provides the basis for the development of an assay for detection of environmental mutagens and conserved genetic factors involved in meiotic NAHR frequency fluctuation. Described below are the current improvements I am making to my preliminary CNV assay and prospective applications of this model system. The results of this work will provide clarity regarding environmental meiotic recurrent CNV mutagens and will be published beyond this thesis.

A High-Throughput Second Generation to Study Meiotic Non-Allelic Homologous Recombination

My preliminary assay system was developed to determine the effects of LCR size on recurrent CNV formation as confirmation of faithful reproduction of the meiotic NAHR results seen in human genomic disorders. However, this system relied on time consuming tetrad dissection, spore growth, as well as spore viability for accurate interpretation of CNV frequency. To better address the combinatory effects of environment, genetics, and genome architecture on meiotic NAHR, it will be useful to adapt my preliminary system for use as a high-throughput assay.

Traditionally, *S. cerevisiae* meiosis studies are performed in the SK1 strain background due to its robust sporulation trait. As such, there is a wealth of knowledge surrounding the factors involved in SK1 meiosis. My preliminary system was built in the CG379 laboratory genetic background of *S. cerevisiae*, however CG379 is a less robust sporulator, and has not been as thoroughly characterized for meiotic studies. I will redesign a similar NAHR reporter construct in the SK1 *S. cerevisiae* background. As a

strong sporulator, building a secondary system in the SK1 background will reduce the time required for sporulation and increase the percentage of diploids who sporulate, thus reducing time spent dissecting tetrads.

Additionally, due to the necessity of deletion growth for intrachromatid CNV detection, deletion viability is essential for uncovering the contributions of each NAHR mechanism to CNV frequency. Design of a high-throughput system in the SK1 background will allow for a redesigned interstitial spacer region lacking essential genes, similar to the current 35Kb LCR strain, thus allowing deletion survival and recovery.

The necessity of tetrad dissection eliminates the possibility of truly high-throughput studies. Addition of spore-autonomous fluorescent proteins, such as those described by Thacker *et. al*, could introduce the ability of automated CNV measurement with flow cytometry, fluorescent cell sorting, or fluorescent microscopy (THACKER *et al.* 2011). In light of the recently described flow cytometry method for high throughput detection of *S. cerevisiae* recombination rates, I have integrated a similar fluorescent NAHR detection system in my secondary system by integrating hemizygous spore-autonomous GFP or tdTomato into the interstitial spacer region of each homolog in addition to the *SFA1*^{V208I}*-CUP1* and *Kan*MX4 or *Hph*MX4 reporters (pMC5, pMC7) (THACKER *et al.* 2011; RAFFOUX *et al.* 2018).

Meiotic Allelic Homologous Recombination

We expect to find different classes of recurrent CNV mutagens. Some mutagens may affect total recombination rates leading to fluctuations in both NAHR and allelic homologous recombination. Other mutagens may affect only NAHR while leaving allelic recombination unchanged. Still other mutagens may affect only the interchromosome bias. While all classes of mutagens will be important to study, their mechanisms of action will likely be different. It will be beneficial to introduce allelic markers that span the CNV region as well markers outside of the CNV region. Segregation patterns of these allelic markers will provide understanding of perturbation of meiotic allelic recombination, allowing distinction of different meiotic mutagens.

Candidate Genetic and Environmental Screen

It is hypothesized that there are undiscovered genetic and environmental factors specific to meiotic NAHR (CONOVER and ARGUESO 2016). The predicted mode of action of such a mutagen and reasons behind this position are discussed in depth in the introduction of this thesis (CONOVER and ARGUESO 2016). It will be interesting to use my assay to perform a genetic and environmental screen for potential NAHR risk factors. The high throughput improvements to my system will allow for large panels of potential environmental mutagens to be tested. Additionally, without the prerequisite of spore viability, genetic mutations or chemicals which reduce spore germination and viability can be interrogated for their effects on NAHR frequency.

REFERENCES

ADEWOYE, A. B., S. J. LINDSAY, Y. E. DUBROVA and M. E. HURLES, 2015 The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline. Nat Commun **6:** 6684.

AGUILERA, A., and H. L. KLEIN, 1988 Genetic control of intrachromosomal recombination in Saccharomyces cerevisiae. I. Isolation and genetic characterization of hyper-recombination mutations. Genetics **119:** 779-790.

ALLEN-SOLTERO, S., S. L. MARTINEZ, C. D. PUTNAM and R. D. KOLODNER, 2014 A *Saccharomyces cerevisiae* RNase H2 interaction network functions to suppress genome instability. Mol Cell Biol **34:** 1521-1534.

ARGUESO, J. L., M. CHAPMAN, G. STEWART and D. MARANON, 2014 A Yeast Bioassay to Uncover Environmental Mediators of Germ Cell Copy Number Mutagenesis. Environmental and Molecular Mutagenesis **55:** S34-S34.

ARGUESO, J. L., J. WESTMORELAND, P. A. MIECZKOWSKI, M. GAWEL, T. D. PETES *et al.*, 2008 Double-strand breaks associated with repetitive DNA can reshape the genome. Proc. Natl Acad Sci USA **105:** 11845-11850.

ARLT, M. F., J. G. MULLE, V. M. SCHAIBLEY, R. L. RAGLAND, S. G. DURKIN *et al.*, 2009 Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. Am J Hum Genet **84:** 339-350.

ARLT, M. F., A. C. OZDEMIR, S. R. BIRKELAND, R. H. LYONS, JR., T. W. GLOVER *et al.*, 2011a Comparison of constitutional and replication stress-induced genome structural variation by SNP array and mate-pair sequencing. Genetics **187:** 675-683.

ARLT, M. F., A. C. OZDEMIR, S. R. BIRKELAND, T. E. WILSON and T. W. GLOVER, 2011b Hydroxyurea induces *de novo* copy number variants in human cells. Proc Natl Acad Sci USA **108:** 17360-17365.

ARLT, M. F., S. RAJENDRAN, S. R. BIRKELAND, T. E. WILSON and T. W. GLOVER, 2012a De novo CNV formation in mouse embryonic stem cells occurs in the absence of Xrcc4-dependent nonhomologous end joining. PLoS Genet **8:** e1002981.

ARLT, M. F., S. RAJENDRAN, S. R. BIRKELAND, T. E. WILSON and T. W. GLOVER, 2014 Copy number variants are produced in response to low-dose ionizing radiation in cultured cells. Environ Mol Mutagen **55:** 103-113.

ARLT, M. F., T. E. WILSON and T. W. GLOVER, 2012b Replication stress and mechanisms of CNV formation. Curr Opin Genet Dev **22:** 204-210.

BAILEY, J. A., Z. GU, R. A. CLARK, K. REINERT, R. V. SAMONTE *et al.*, 2002 Recent segmental duplications in the human genome. Science **297:** 1003-1007.

BATTAGLIA, A., V. DOCCINI, L. BERNARDINI, A. NOVELLI, S. LODDO *et al.*, 2013 Confirmation of chromosomal microarray as a first-tier clinical diagnostic test for individuals with developmental delay, intellectual disability, autism spectrum disorders and dysmorphic features. Eur J Paediatr Neurol **17:** 589-599.

BEAUDET, A. L., 2014 Reaching a CNV milestone. Nat Genet **46:** 1046-1048.

BETERMIER, M., P. BERTRAND and B. S. LOPEZ, 2014 Is non-homologous end-joining really an inherently error-prone process? PLoS Genet **10:** e1004086.

BOFFETTA, P., 2006 Human cancer from environmental pollutants: the epidemiological evidence. Mutat Res **608:** 157-162.

BOONE, P. M., B. YUAN, I. M. CAMPBELL, J. C. SCULL, M. A. WITHERS *et al.*, 2014 The Alu-rich genomic architecture of SPAST predisposes to diverse and functionally distinct disease-associated CNV alleles. Am J Hum Genet **95:** 143-161.

BOREL, C., F. CHEUNG, H. STEWART, D. A. KOOLEN, C. PHILLIPS *et al.*, 2012 Evaluation of PRDM9 variation as a risk factor for recurrent genomic disorders and chromosomal non-disjunction. Hum Genet **131:** 1519-1524.

BOSE, P., K. E. HERMETZ, K. N. CONNEELY and M. K. RUDD, 2014 Tandem repeats and G-rich sequences are enriched at human CNV breakpoints. PLoS One **9:** e101607.

BURNS, K. H., and J. D. BOEKE, 2012 Human transposon tectonics. Cell **149:** 740-752.

CAMPBELL, C. D., and E. E. EICHLER, 2013 Properties and rates of germline mutations in humans. Trends Genet **29:** 575-584.

CAMPBELL, I. M., T. GAMBIN, P. DITTWALD, C. R. BECK, A. SHUVARIKOV *et al.*, 2014 Human endogenous retroviral elements promote genome instability via non-allelic homologous recombination. BMC Biol **12:** 74.

CAMPBELL, I. M., C. A. SHAW, P. STANKIEWICZ and J. R. LUPSKI, 2015 Somatic mosaicism: implications for disease and transmission genetics. Trends Genet **31:** 382-392.

CARVALHO, C. M., and J. R. LUPSKI, 2016 Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet **17:** 224-238.

CARVALHO, C. M., S. VASANTH, M. SHINAWI, C. RUSSELL, M. B. RAMOCKI *et al.*, 2014 Dosage changes of a segment at 17p13.1 lead to intellectual disability and microcephaly as a result of complex genetic interaction of multiple genes. Am J Hum Genet **95:** 565-578.

CERRITELLI, S. M., and R. J. CROUCH, 2009 Ribonuclease H: the enzymes in eukaryotes. FEBS J **276:** 1494-1505.

CERRITELLI, S. M., and R. J. CROUCH, 2016 The Balancing Act of Ribonucleotides in DNA. Trends Biochem Sci **41:** 434-445.

CHEN, C., and R. D. KOLODNER, 1999 Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants. Nature genetics **23:** 81-85.

CHEN, L., W. ZHOU, C. ZHANG, J. R. LUPSKI, L. JIN *et al.*, 2015 CNV instability associated with DNA replication dynamics: evidence for replicative mechanisms in CNV mutagenesis. Hum Mol Genet **24:** 1574-1583.

CHERRY, J. M., E. L. HONG, C. AMUNDSEN, R. BALAKRISHNAN, G. BINKLEY *et al.*, 2012 Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res **40:** D700-705.

CHO, J. E., and S. JINKS-ROBERTSON, 2016 Ribonucleotides and Transcription-Associated Mutagenesis in Yeast. J Mol Biol.

CHON, H., J. L. SPARKS, M. RYCHLIK, M. NOWOTNY, P. M. BURGERS *et al.*, 2013 RNase H2 roles in genome integrity revealed by unlinking its activities. Nucleic Acids Res **41:** 3130-3143.

CIRIELLO, G., M. L. MILLER, B. A. AKSOY, Y. SENBABAOGLU, N. SCHULTZ *et al.*, 2013 Emerging landscape of oncogenic signatures across human cancers. Nat Genet **45:** 1127-1133.

CLARK, A. B., S. A. LUJAN, G. E. KISSLING and T. A. KUNKEL, 2011 Mismatch repair-independent tandem repeat sequence instability resulting from ribonucleotide incorporation by DNA polymerase epsilon. DNA Repair (Amst) **10:** 476-482.

CONOVER, H. N., and J. L. ARGUESO, 2016 Contrasting mechanisms of de novo copy number mutagenesis suggest the existence of different classes of environmental copy number mutagens. Environ Mol Mutagen **57:** 3-9.

CONOVER, H. N., S. A. LUJAN, M. J. CHAPMAN, D. A. CORNELIO, R. SHARIF *et al.*, 2015 Stimulation of Chromosomal Rearrangements by Ribonucleotides. Genetics **201:** 951-961.

COSTA, E. O. A., I. P. PINTO, M. W. GONÇALVES, J. F. DA SILVA, L. G. OLIVEIRA *et al.*, 2018 Small de novo CNVs as biomarkers of parental exposure to low doses of ionizing radiation of caesium-137. Scientific Reports **8:** 5914.

COVO, S., C. M. PUCCIA, J. L. ARGUESO, D. A. GORDENIN and M. A. RESNICK, 2014 The sister chromatid cohesion pathway suppresses multiple chromosome gain and chromosome amplification. Genetics **196:** 373-384.

DEMARINI, D. M., 2012 Declaring the existence of human germ-cell mutagens. Environ Mol Mutagen **53:** 166-172.

DITTWALD, P., T. GAMBIN, P. SZAFRANSKI, J. LI, S. AMATO *et al.*, 2013 NAHR-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and Mendelizing traits. Genome Res **23:** 1395-1409.

DUYZEND, M. H., X. NUTTLE, B. P. COE, C. BAKER, D. A. NICKERSON *et al.*, 2016 Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. Am J Hum Genet **98:** 45-57.

EL HAGE, A., S. L. FRENCH, A. L. BEYER and D. TOLLERVEY, 2010 Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev **24:** 1546-1558.

EPSHTEIN, A., C. J. POTENSKI and H. L. KLEIN, 2016 Increased spontaneous recombination in RNase H2-deficient cells arises from multiple contiguous rNMPs and not from single rNMP residues incorporated by DNA polymerase epsilon. Microbial Cell **3:** 248-254.

FERREIRA, D. W., and P. ALLARD, 2015 Models of germ cell development and their application for toxicity studies. Environ Mol Mutagen **10.1002/em.21946**.

GAMBUS, A., R. C. JONES, A. SANCHEZ-DIAZ, M. KANEMAKI, F. VAN DEURSEN *et al.*, 2006 GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol **8:** 358-366.

GHODGAONKAR, M. M., F. LAZZARO, M. OLIVERA-PIMENTEL, M. ARTOLA-BORAN, P. CEJKA *et al.*, 2013 Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. Mol Cell **50:** 323-332.

GIRIRAJAN, S., M. Y. DENNIS, C. BAKER, M. MALIG, B. P. COE *et al.*, 2013 Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am J Hum Genet **92:** 221-237.

GIRIRAJAN, S., and E. E. EICHLER, 2010 Phenotypic variability and genetic susceptibility to genomic disorders. Hum Mol Genet **19:** R176-187.

GOLDMAN, A. S., and M. LICHTEN, 1996 The efficiency of meiotic recombination between dispersed sequences in Saccharomyces cerevisiae depends upon their chromosomal location. Genetics **144:** 43-55.

GOLDMAN, A. S., and M. LICHTEN, 2000 Restriction of ectopic recombination by interhomolog interactions during *Saccharomyces cerevisiae* meiosis. Proc Natl Acad Sci U S A **97:** 9537-9542.

GOLDSTEIN, A. L., and J. H. MCCUSKER, 1999 Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. Yeast **15:** 1541-1553.

GRUSHCOW, J. M., T. M. HOLZEN, K. J. PARK, T. WEINERT, M. LICHTEN *et al.*, 1999 *Saccharomyces cerevisiae* checkpoint genes *MEC1, RAD17* and *RAD24* are required for normal meiotic recombination partner choice. Genetics **153:** 607-620.

HABER, J. E., W. Y. LEUNG, R. H. BORTS and M. LICHTEN, 1991 The frequency of meiotic recombination in yeast is independent of the number and position of homologous donor sequences: implications for chromosome pairing. Proc Natl Acad Sci U S A **88:** 1120-1124.

HALL, B. M., C. X. MA, P. LIANG and K. K. SINGH, 2009 Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics **25:** 1564-1565.

HAREL, T., and J. R. LUPSKI, 2018 Genomic disorders 20 years on-mechanisms for clinical manifestations. Clin Genet **93:** 439-449.

HASTINGS, P. J., G. IRA and J. R. LUPSKI, 2009 A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS genetics **5:** e1000327.

HEHIR-KWA, J. Y., B. RODRIGUEZ-SANTIAGO, L. E. VISSERS, N. DE LEEUW, R. PFUNDT *et al.*, 2011 De novo copy number variants associated with intellectual disability have a paternal origin and age bias. J Med Genet **48:** 776-778.

HERMETZ, K. E., S. NEWMAN, K. N. CONNEELY, C. L. MARTIN, B. C. BALLIF *et al.*, 2014 Large inverted duplications in the human genome form via a fold-back mechanism. PLoS Genet **10:** e1004139.

HERMETZ, K. E., U. SURTI, J. D. CODY and M. K. RUDD, 2012 A recurrent translocation is mediated by homologous recombination between HERV-H elements. Mol Cytogenet **5:** 6.

HOFFMANN, G. R., 1994 Induction of genetic recombination: consequences and model systems. Environ Mol Mutagen **23 Suppl 24:** 59-66.

HUANG, S. N., J. S. WILLIAMS, M. E. ARANA, T. A. KUNKEL and Y. POMMIER, 2016 Topoisomerase I-mediated cleavage at unrepaired ribonucleotides generates DNA double-strand breaks. EMBO J **36:** 361-373.

II, M., T. II, L. I. MIRONOVA and S. J. BRILL, 2011 Epistasis analysis between homologous recombination genes in *Saccharomyces cerevisiae* identifies multiple repair pathways for Sgs1, Mus81-Mms4 and RNase H2. Mutat Res **714:** 33-43.

JINKS-ROBERTSON, S., and H. L. KLEIN, 2015 Ribonucleotides in DNA: hidden in plain sight. Nat Struct Mol Biol **22:** 176-178.

KEENEY, S., J. LANGE and N. MOHIBULLAH, 2014 Self-organization of meiotic recombination initiation: general principles and molecular pathways. Annu Rev Genet **48:** 187-214.

KIM, N., S. N. HUANG, J. S. WILLIAMS, Y. C. LI, A. B. CLARK *et al.*, 2011 Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. Science **332:** 1561-1564.

KIM, S., S. E. PETERSON, M. JASIN and S. KEENEY, 2016 Mechanisms of germ line genome instability. Semin Cell Dev Biol **54:** 177-187.

KIRCHNER, J. M., H. TRAN and M. A. RESNICK, 2000 A DNA polymerase epsilon mutant that specifically causes +1 frameshift mutations within homonucleotide runs in yeast. Genetics **155:** 1623-1632.

KONG, A., M. L. FRIGGE, G. MASSON, S. BESENBACHER, P. SULEM *et al.*, 2012 Rate of de novo mutations and the importance of father's age to disease risk. Nature **488:** 471-475.

KUPIEC, M., and T. D. PETES, 1988 Meiotic recombination between repeated transposable elements in Saccharomyces cerevisiae. Mol Cell Biol **8:** 2942-2954.

LAZZARO, F., D. NOVARINA, F. AMARA, D. L. WATT, J. E. STONE *et al.*, 2012 RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. Mol Cell **45:** 99-110.

LEA, D. E., and C. A. COULSON, 1949 The distribution of the numbers of mutants in bacterial populations. Journal of Genetics **28:** 491-511.

LIU, P., C. M. CARVALHO, P. J. HASTINGS and J. R. LUPSKI, 2012 Mechanisms for recurrent and complex human genomic rearrangements. Curr Opin Genet Dev **22:** 211-220.

LIU, P., A. EREZ, S. C. NAGAMANI, W. BI, C. M. CARVALHO *et al.*, 2011a Copy number gain at Xp22.31 includes complex duplication rearrangements and recurrent triplications. Hum Mol Genet **20:** 1975-1988.

LIU, P., M. LACARIA, F. ZHANG, M. WITHERS, P. J. HASTINGS *et al.*, 2011b Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over. Am J Hum Genet **89:** 580-588.

LUJAN, S. A., A. R. CLAUSEN, A. B. CLARK, H. K. MACALPINE, D. M. MACALPINE *et al.*, 2014 Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. Genome Res **24:** 1751-1764.

LUJAN, S. A., J. S. WILLIAMS, A. R. CLAUSEN, A. B. CLARK and T. A. KUNKEL, 2013 Ribonucleotides are signals for mismatch repair of leading-strand replication errors. Mol Cell **50:** 437-443.

LUPSKI, J. R., 1998 Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet **14:** 417-422.

LUPSKI, J. R., 2009 Genomic disorders ten years on. Genome Med **1:** 42.

LUPSKI, J. R., 2015a Structural variation mutagenesis of the human genome: Impact on disease and evolution. Environ Mol Mutagen **10.1002/em.21943**.

LUPSKI, J. R., 2015b Structural variation mutagenesis of the human genome: Impact on disease and evolution. Environ Mol Mutagen **56:** 419-436.

LYNCH, A. M., J. C. SASAKI, R. ELESPURU, D. JACOBSON-KRAM, V. THYBAUD *et al.*, 2011 New and emerging technologies for genetic toxicity testing. Environ Mol Mutagen **52:** 205-223.

MACARTHUR, J. A., T. D. SPECTOR, S. J. LINDSAY, M. MANGINO, R. GILL *et al.*, 2014 The rate of nonallelic homologous recombination in males is highly variable, correlated between monozygotic twins and independent of age. PLoS Genet **10:** e1004195.

MACRAE, R., 2014 On the shoulders of worms. Trends Genet **30:** 475.

MALKOVA, A., and G. IRA, 2013 Break-induced replication: functions and molecular mechanism. Curr Opin Genet Dev **23:** 271-279.

MCCULLEY, J. L., and T. D. PETES, 2010 Chromosome rearrangements and aneuploidy in yeast strains lacking both Tel1p and Mec1p reflect deficiencies in two different mechanisms. Proc Natl Acad Sci U S A **107:** 11465-11470.

MEFFORD, H. C., and E. E. EICHLER, 2009 Duplication hotspots, rare genomic disorders, and common disease. Curr Opin Genet Dev **19:** 196-204.

MOLINA, O., E. ANTON, F. VIDAL and J. BLANCO, 2011 Sperm rates of 7q11.23, 15q11q13 and 22q11.2 deletions and duplications: a FISH approach. Hum Genet **129:** 35-44.

MOLINA, O., J. BLANCO and F. VIDAL, 2010 Deletions and duplications of the 15q11-q13 region in spermatozoa from Prader-Willi syndrome fathers. Mol Hum Reprod **16:** 320-328.

MORRISON, A., J. B. BELL, T. A. KUNKEL and A. SUGINO, 1991 Eukaryotic DNA polymerase amino acid sequence required for 3'----5' exonuclease activity. Proc Natl Acad Sci U S A **88:** 9473-9477.

MYUNG, K., and R. D. KOLODNER, 2003 Induction of genome instability by DNA damage in Saccharomyces cerevisiae. DNA Repair (Amst) **2:** 243-258.

NARAYANAN, V., P. A. MIECZKOWSKI, H. M. KIM, T. D. PETES and K. S. LOBACHEV, 2006 The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. Cell **125:** 1283-1296.

NEWMAN, S., K. E. HERMETZ, B. WECKSELBLATT and M. K. RUDD, 2015 Next-Generation Sequencing of Duplication CNVs Reveals that Most Are Tandem and Some Create Fusion Genes at Breakpoints. Am J Hum Genet **96:** 208-220.

NICK MCELHINNY, S. A., D. KUMAR, A. B. CLARK, D. L. WATT, B. E. WATTS *et al.*, 2010a Genome instability due to ribonucleotide incorporation into DNA. Nat Chem Biol **6:** 774-781.

NICK MCELHINNY, S. A., B. E. WATTS, D. KUMAR, D. L. WATT, E. B. LUNDSTROM *et al.*, 2010b Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. Proc Natl Acad Sci U S A **107:** 4949-4954.

O'CONNELL, K., S. JINKS-ROBERTSON and T. D. PETES, 2015 Elevated Genome-Wide Instability in Yeast Mutants Lacking RNase H Activity. Genetics **201:** 963-975.

PAN, J., M. SASAKI, R. KNIEWEL, H. MURAKAMI, H. G. BLITZBLAU *et al.*, 2011 A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell **144:** 719-731.

PAVLOV, Y. I., D. NGUYEN and T. A. KUNKEL, 2001 Mutator effects of overproducing DNA polymerase eta (Rad30) and its catalytically inactive variant in yeast. Mutat Res **478:** 129-139.

PAYEN, C., R. KOSZUL, B. DUJON and G. FISCHER, 2008 Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. PLoS Genet **4:** e1000175.

PENG, Z., W. ZHOU, W. FU, R. DU, L. JIN *et al.*, 2015 Correlation between frequency of non-allelic homologous recombination and homology properties: evidence from homology-mediated CNV mutations in the human genome. Hum Mol Genet **24:** 1225-1233.

PETES, T. D., and C. W. HILL, 1988 Recombination between repeated genes in microorganisms. Annu Rev Genet **22:** 147-168.

POTENSKI, C. J., H. NIU, P. SUNG and H. L. KLEIN, 2014 Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms. Nature **511:** 251-254.

PRATTO, F., K. BRICK, P. KHIL, F. SMAGULOVA, G. V. PETUKHOVA *et al.*, 2014 DNA recombination. Recombination initiation maps of individual human genomes. Science **346:** 1256442.

RAFFOUX, X., M. BOURGE, F. DUMAS, O. C. MARTIN and M. FALQUE, 2018 High-throughput measurement of recombination rates and genetic interference in Saccharomyces cerevisiae. Yeast.

REIJNS, M. A., B. RABE, R. E. RIGBY, P. MILL, K. R. ASTELL *et al.*, 2012 Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. Cell **149:** 1008-1022.

RELIENE, R., A. J. BISHOP and R. H. SCHIESTL, 2007 Involvement of homologous recombination in carcinogenesis. Adv Genet **58:** 67-87.

RELIENE, R., and R. H. SCHIESTL, 2003 Mouse models for induced genetic instability at endogenous loci. Oncogene **22:** 7000-7010.

ROBBERECHT, C., T. VOET, M. Z. ESTEKI, B. A. NOWAKOWSKA and J. R. VERMEESCH, 2013 Nonallelic homologous recombination between retrotransposable elements is a driver of *de novo* unbalanced translocations. Genome Res **23:** 411-418.

ROSE, M. D., F. WINSTON and P. HIETER, 1990 *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

RUSSO, A., F. PACCHIEROTTI, D. CIMINI, N. J. GANEM, A. GENESCA *et al.*, 2015 Genomic instability: Crossing pathways at the origin of structural and numerical chromosome changes. Environ Mol Mutagen **10.1002/em.21945**.

SANTOS-PEREIRA, J. M., and A. AGUILERA, 2015 R loops: new modulators of genome dynamics and function. Nat Rev Genet **16:** 583-597.

SASAKI, M., J. LANGE and S. KEENEY, 2010 Genome destabilization by homologous recombination in the germ line. Nat Rev Mol Cell Biol **11:** 182-195.

SCHAAF, C. P., J. WISZNIEWSKA and A. L. BEAUDET, 2011 Copy number and SNP arrays in clinical diagnostics. Annu Rev Genomics Hum Genet **12:** 25-51.

SCHIESTL, R. H., 1989 Nonmutagenic carcinogens induce intrachromosomal recombination in yeast. Nature **337:** 285-288.

SCHWACHA, A., and N. KLECKNER, 1997 Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. Cell **90:** 1123-1135.

SEBAT, J., B. LAKSHMI, D. MALHOTRA, J. TROGE, C. LESE-MARTIN *et al.*, 2007 Strong association of de novo copy number mutations with autism. Science **316:** 445-449.

SEGUREL, L., M. J. WYMAN and M. PRZEWORSKI, 2014 Determinants of mutation rate variation in the human germline. Annu Rev Genomics Hum Genet **15:** 47-70.

SHINAWI, M., P. LIU, S. H. KANG, J. SHEN, J. W. BELMONT *et al.*, 2010 Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. J Med Genet **47:** 332-341.

SHINOHARA, M., and A. SHINOHARA, 2013 Multiple pathways suppress non-allelic homologous recombination during meiosis in Saccharomyces cerevisiae. PLoS One **8:** e63144.

SHUVARIKOV, A., I. M. CAMPBELL, P. DITTWALD, N. J. NEILL, M. G. BIALER *et al.*, 2013 Recurrent HERV-H-mediated 3q13.2-q13.31 deletions cause a syndrome of hypotonia and motor, language, and cognitive delays. Hum Mutat **34:** 1415-1423.

SIBBONS, C., J. K. MORRIS, J. A. CROLLA, P. A. JACOBS and N. S. THOMAS, 2012 De novo deletions and duplications detected by array CGH: a study of parental origin in relation to mechanisms of formation and size of imbalance. Eur J Hum Genet **20:** 155-160.

SIKORSKI, R. S., and P. HIETER, 1989 A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics **122:** 19-27.

SPARKS, J. L., and P. M. BURGERS, 2015 Error-free and mutagenic processing of topoisomerase 1-provoked damage at genomic ribonucleotides. EMBO J **34:** 1259-1269.

SPARKS, J. L., H. CHON, S. M. CERRITELLI, T. A. KUNKEL, E. JOHANSSON *et al.*, 2012 RNase H2-initiated ribonucleotide excision repair. Mol Cell **47:** 980-986.

SPELL, R. M., and S. JINKS-ROBERTSON, 2004 Determination of mitotic recombination rates by fluctuation analysis in Saccharomyces cerevisiae. Methods Mol Biol **262:** 3-12.

STANKIEWICZ, P., C. J. SHAW, M. WITHERS, K. INOUE and J. R. LUPSKI, 2004 Serial segmental duplications during primate evolution result in complex human genome architecture. Genome Res **14:** 2209-2220.

STANTON, J. L., 2012 A yeast-based assay system for the study of environmentally induced copy number variation. Dissertations & Theses @ Colorado State University.

STARTEK, M., P. SZAFRANSKI, T. GAMBIN, I. M. CAMPBELL, P. HIXSON *et al.*, 2015 Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination. Nucleic Acids Res **43:** 2188-2198.

STEWART, C., D. KURAL, M. P. STROMBERG, J. A. WALKER, M. K. KONKEL *et al.*, 2011 A comprehensive map of mobile element insertion polymorphisms in humans. PLoS Genet **7:** e1002236.

SYMINGTON, L. S., R. ROTHSTEIN and M. LISBY, 2014 Mechanisms and Regulation of Mitotic Recombination in Saccharomyces cerevisiae. Genetics **198:** 795-835.

THACKER, D., I. LAM, M. KNOP and S. KEENEY, 2011 Exploiting spore-autonomous fluorescent protein expression to quantify meiotic chromosome behaviors in Saccharomyces cerevisiae. Genetics **189:** 423-439.

THOMPSON, D. A., and F. W. STAHL, 1999 Genetic control of recombination partner preference in yeast meiosis. Isolation and characterization of mutants elevated for meiotic unequal sister-chromatid recombination. Genetics **153:** 621-641.

VASAN, S., A. DEEM, S. RAMAKRISHNAN, J. L. ARGUESO and A. MALKOVA, 2014 Cascades of genetic instability resulting from compromised break-induced replication. PLoS Genet **10:** e1004119.

VERGES, L., O. MOLINA, E. GEAN, F. VIDAL and J. BLANCO, 2014 Deletions and duplications of the 22q11.2 region in spermatozoa from DiGeorge/velocardiofacial fathers. Mol Cytogenet **7:** 86.

VERGES, L., F. VIDAL, E. GEAN, A. ALEMANY-SCHMIDT, M. OLIVER-BONET *et al.*, 2017 An exploratory study of predisposing genetic factors for DiGeorge/velocardiofacial syndrome. Sci Rep **7:** 40031.

WACH, A., A. BRACHAT, R. POHLMANN and P. PHILIPPSEN, 1994 New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast **10:** 1793-1808.

WAHBA, L., J. D. AMON, D. KOSHLAND and M. VUICA-ROSS, 2011 RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Mol Cell **44:** 978-988.

WAHBA, L., L. COSTANTINO, F. J. TAN, A. ZIMMER and D. KOSHLAND, 2016 S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev **30:** 1327-1338.

WECKSELBLATT, B., K. E. HERMETZ and M. K. RUDD, 2015 Unbalanced translocations arise from diverse mutational mechanisms including chromothripsis. Genome Res **25:** 937-947.

WEI, W., J. H. MCCUSKER, R. W. HYMAN, T. JONES, Y. NING *et al.*, 2007 Genome sequencing and comparative analysis of Saccharomyces cerevisiae strain YJM789. Proc Natl Acad Sci U S A **104:** 12825-12830.

WELCH, B. L., 1947 The generalisation of student's problems when several different population variances are involved. Biometrika **34:** 28-35.

WILLIAMS, J. S., A. R. CLAUSEN, S. A. LUJAN, L. MARJAVAARA, A. B. CLARK *et al.*, 2015 Evidence that processing of ribonucleotides in DNA by topoisomerase 1 is leading-strand specific. Nat Struct Mol Biol **22:** 291-297.

WILLIAMS, J. S., D. B. GEHLE and T. A. KUNKEL, 2017 The role of RNase H2 in processing ribonucleotides incorporated during DNA replication. DNA Repair **in press**.

WILLIAMS, J. S., and T. A. KUNKEL, 2014 Ribonucleotides in DNA: origins, repair and consequences. DNA Repair **19:** 27-37.

WILLIAMS, J. S., S. A. LUJAN and T. A. KUNKEL, 2016 Processing ribonucleotides incorporated during eukaryotic DNA replication. Nat Rev Mol Cell Biol **17:** 350-363.

WILLIAMS, J. S., D. J. SMITH, L. MARJAVAARA, S. A. LUJAN, A. CHABES *et al.*, 2013 Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA. Mol Cell **49:** 1010-1015.

WILSON, T. E., M. F. ARLT, S. H. PARK, S. RAJENDRAN, M. PAULSEN *et al.*, 2015 Large transcription units unify copy number variants and common fragile sites arising under replication stress. Genome Res **25:** 189-200.

YAUK, C. L., M. J. AARDEMA, J. BENTHEM, J. B. BISHOP, K. L. DEARFIELD *et al.*, 2015a Approaches for identifying germ cell mutagens: Report of the 2013 IWGT workshop on germ cell assays(). Mutat Res Genet Toxicol Environ Mutagen **783:** 36-54.

YAUK, C. L., M. J. AARDEMA, J. VAN BENTHEM, J. B. BISHOP, K. L. DEARFIELD *et al.*, 2015b Approaches for identifying germ cell mutagens: Report of the 2013 IWGT workshop on germ cell assays. Mutat. Res.: Genet. Toxicol. Environ. Mutagen. **783:** 36-54.

YAUK, C. L., J. L. ARGUESO, S. S. AUERBACH, P. AWADALLA, S. R. DAVIS *et al.*, 2013 Harnessing genomics to identify environmental determinants of heritable disease. Mutat Res **752:** 6-9.

ZHANG, H., Y. BIAN, X. GOU, Y. DONG, S. RUSTGI *et al.*, 2013a Intrinsic karyotype stability and gene copy number variations may have laid the foundation for tetraploid wheat formation. Proc Natl Acad Sci U S A **110:** 19466-19471.

ZHANG, H., A. F. ZEIDLER, W. SONG, C. M. PUCCIA, E. MALC *et al.*, 2013b Gene copy-number variation in haploid and diploid strains of the yeast Saccharomyces cerevisiae. Genetics **193:** 785-801.

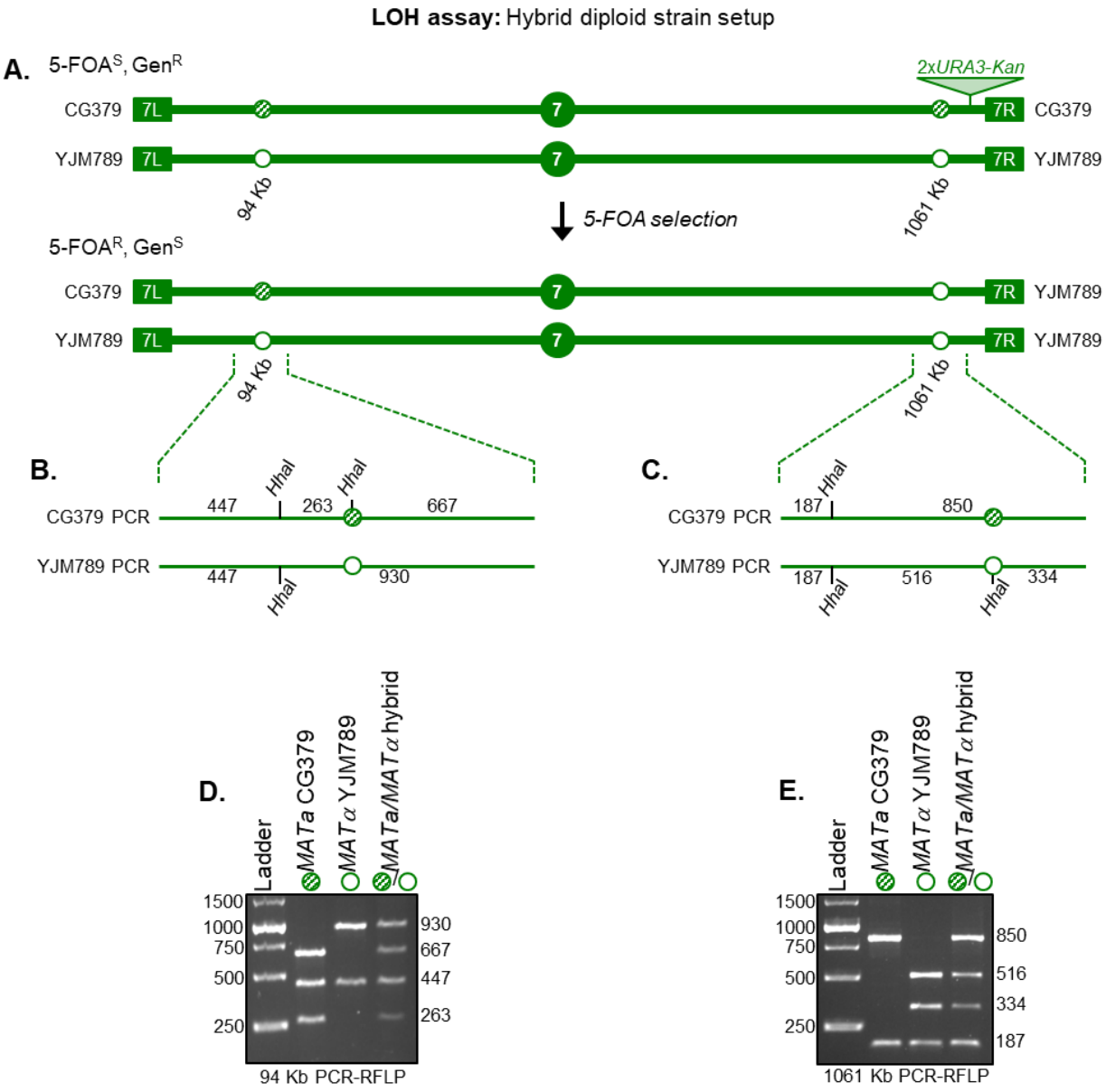
ZICKLER, D., and N. KLECKNER, 2015 Recombination, Pairing, and Synapsis of Homologs during Meiosis. Cold Spring Harb Perspect Biol **7:** a016626.

ZIMMER, A. D., and D. KOSHLAND, 2016 Differential roles of the RNases H in preventing chromosome instability. Proc Natl Acad Sci U S A **113:** 12220-12225.

ZIMMERMANN, F. K., 1992 Tests for recombinagens in fungi. Mutat Res **284:** 147-158.

SUPPLEMENTAL MATERIALS

Supplemental Figures



F.

5-FOA^R clones derived from	Genotype at PCR-RFLP marker (Heterozygous clones / Total clones)	
	left arm - 94 Kb	right arm - 1061 Kb
Wild type diploids	14 / 14	1 / 14
rnh201Δ diploids	13 / 13	0 / 13

Figure S2.1 Loss-of-heterozygosity (LOH) assay system in hybrid diploids.

Figure S2.1 Loss-of-heterozygosity (LOH) assay system in hybrid diploids.

(**A**) Schematic representation of the karyotype of the hybrid diploid strains used in LOH assays. A hemizygous counter-selectable CORE2 cassette was inserted near the right end of the Chr7 homolog from the CG379 parent haploid. The Chr7 homolog from the YJM789 parent haploid did not contain this insertion. Two heterozygous sites (position 94 Kb and 1061 Kb) that coincide with the recognition sequence for the *Hha*I restriction endonuclease in one of the two Chr7 homologs are shown as hatched circles corresponding to the CG379 alleles and open circles corresponding to the YJM789 alleles. Allelic homologous recombination between the two Chr7 homologs may result in homozygosity for the YJM789 region lacking the CORE2 insertion (position 1072 Kb), and often also for the nearby 1061 Kb site within the same LOH tract. (**B**) and (**C**) Representation of the *Hha*I restriction maps of PCR products obtained using primers flanking the 94 Kb and 1061 Kb sites, respectively. The numbers indicate the expected sizes in bp of the *Hha*I restriction fragments from each allele. (**D**) and (**E**) *Hha*I digested PCR products from the 94 Kb and 1061 Kb loci, respectively, resolved by agarose gel electrophoresis showing the banding patterns for the CG379 and YJM789 alleles, and the pattern found in the parent heterozygous hybrid diploids. The numbers indicate the band sizes of the ladder (left) and restriction fragments (right). (**F**) Genotypes at the 94 Kb and 1061 Kb markers from 27 independently selected 5-FOA^R clones derived from hybrid diploids. All 5-FOA^R clones remained heterozygous for the 94 Kb marker, and all but one clone became homozygous for the YJM789 allele.

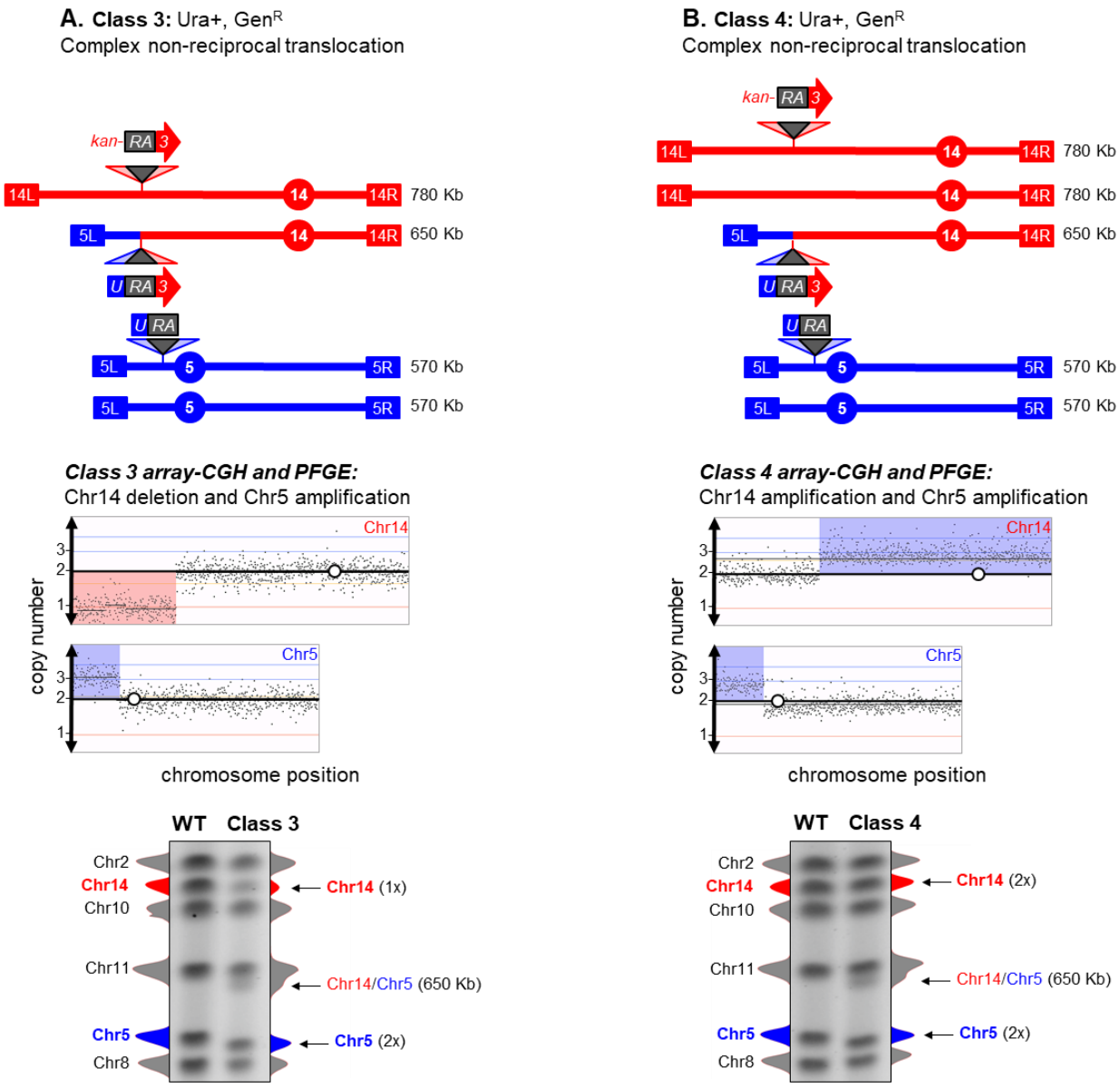


Figure S2.2 Examples of unexpected outcomes obtained from the NAHR assay.

Labeling is as described for Figure 2.3 in the main text. (**A**) The class 3 non-reciprocal outcome is shown to the left, with the respective sizes and structures of the associated chromosomes. The array-CGH plots and PFGE for a representative Class 3 clone shows a deletion (1 copy; pink-shaded) on the left arm of Chr14 from *TEL14L* to *PEX17*, and an amplification (3 copies; purple-shaded) on the left arm of Chr5 from *TEL05L* to *URA3*. (**B**) The class 4 non-reciprocal outcome is shown to the right, and the array-CGH plots and PFGE of a representative clone show an amplification (3 copies; purple-shaded) of part of the left arm of Chr14 from *PEX17* to *CEN14* and the entire right arm from *CEN14* to *TEL14R*. Class 4 clones have two complete copies of Chr14 and one copy of a Chr14/Chr5 translocation. The Chr5 array-CGH plot shows an amplification (3 copies; purple-shaded) on the left arm of Chr5 from *TEL05L* to *URA3*.

Supplemental Tables

Table S2.1 Yeast strains used in quantitative recombination assays.

		DIPLOIDS	HAPLOIDS PARENTS			DIPLOIDS
		NAHR assay	NAHR	Shared	LOH	LOH assay
EXPERIMENTAL GENOTYPES	*MAT*	*a/α*	*a*	α	*a*	*a/α*
	MAL13	*+/+*	*+*	*+*	*::CORE2*	*+/::CORE2*
	ura3	*Δ3'/Δ0*	*Δ3'*	*Δ0*	*Δ0*	*Δ0/Δ0*
	PEX17	*+/::Kan-ura3Δ5'*	*::Kan-ura3Δ5'*	*+*	*+*	*+/+*
MUTANT BACKGROUNDS		JAY# ↓	JAY# ↓	JAY# ↓	JAY# ↓	JAY# ↓
Wild Type		1296	1087	1167	1168	1201
rnh201Δ::Hph		1297,1563	1301	1161	1165	1487,1488
top1Δ::Nat		1401	1393	1163	1166	1402
rnh201Δ::Hph top1Δ::Nat		1299	1303	1164	1162	1200
rnh202Δ::Nat		1403,1404	1359, 1360	1361, 1362	1353, 1354	1405,1406
rnh203Δ::Nat		1424,1562	1314,1315	1308	1309	1310
pol1-L868M		1489,1490	1415	1482	1497, 1498	1516,1517
pol1-L868M rnh201Δ::Hph		1474,1475	1416	1389	1421	1476,1477
pol1-L868M rnh201Δ::Hph		1499,1505	1392	1400	1407,1408	1472,1473

top1Δ::Nat					
pol1-L868M top1Δ::Nat	NA	NA	1586,1588	1585,1587	1589,1590
pol3-L612M	1495,1496	1419	1411	1417	1446,1447
pol3-L612M rnh201Δ::Hph	1442,1443, 1597	1376	1387	1418	1440,1441
pol3-L612M rnh201Δ::Hph top1Δ::Nat	1444,1445	1414	1412	1450,1451	1483,1484
pol3-L612M top1Δ::Nat	NA	NA	1582,1584	1581,1583	1591,1592
pol2-M644G	1449,1758, 1761,1564	1334	1313	1318	1432
pol2-M644G rnh201Δ::Hph	1470,1471	1336	1320	1321	1428,1429
pol2-M644G rnh201Δ::Hph top1Δ::Nat	1438,1439	1343	1324	1325	1430,1431
pol2-M644G top1Δ::Nat	NA	NA	1574,1576	1573,1575	1593,1594
pol2-M644L	1468,1469	1335	1312	1319	1466,1467
pol2-M644L rnh201Δ::Hph	1462,1463	1425,1426	1322	1323	1464,1465
pol2-M644L rnh201Δ::Hph top1Δ::Nat	1458,1459	1427	1342	1422,1423	1515,1461
pol2-M644L top1Δ::Nat	NA	NA	1578,1580	1577,1579	1595,1596

Strain table notes: The numbers in the table represent the JAY strain collection numbers for all haploid parents and the resulting diploids used in the LOH and NAHR assays. Two or more independently built strains were used where more than one number appears. The diploids were made by mating a *MATa* assay-specific haploid to a shared *MAT*α haploid parent. For example, the JAY1296 NAHR assay wild type diploid was constructed by mating JAY1087 to JAY1167, while the JAY1201 LOH assay wild type diploid was constructed by mating JAY1168 to JAY1167.

Table S2.2 Quantitative analyses of recombination.

	Genotype (stID)	LOH assay			NAHR assay		
		Cultures	rate x2x10^{-5}/cell/division		Cultures	rate x10^{-7}/cell/division	
		(n)	Median	95% CI	(n)	Median	95% CI
POL+	*WT* (A)	40	3.71	3.06 - 4.84	28	2.04	1.92 - 2.59
	rnh201 (B)	45	23.65	17.85 - 34.66	50	2.73	1.92 - 3.24
	rnh202 (C)	20	30.45	24.29 - 36.39	35	2.73	1.91 - 3.12
	rnh203 (D)	29	24.90	20.24 - 36.87	35	2.02	1.23 - 3.1
	rnh201 top1 (E)	34	7.70	6.75 - 9.62	20	1.93	1.69 - 2.36
	top1 (F)	20	3.13	2.3 - 3.67	20	1.62	1.12 - 2.02
pol2-M644G	*WT* (G)	19	28.42	25.71 - 36.77	21	1.17	0.7 - 1.6
	rnh201 (H)	32	84.58	60.33 - 95.46	35	6.30	4.38 - 7.73
	rnh201 top1 (I)	21	14.24	8.18 - 27.02	28	0.67	0.55 - 0.9
	top1 (J)	21	5.38	3.81 - 8.25			
pol2-M644L	*WT* (K)	21	5.64	4.77 - 12.62	30	0.60	0.5 - 0.89
	rnh201 (L)	20	15.42	12.34 - 21.87	30	0.96	0.6 - 1.5
	rnh201 top1 (M)	21	8.81	5.81 - 15.52	30	0.34	0.21 - 0.39
	top1 (N)	21	3.54	2.56 - 6.62			
pol1-L868M	*WT* (O)	21	4.30	3.51 - 7.77	30	1.31	0.95 - 1.76
	rnh201 (P)	21	33.16	19.05 - 45.58	30	2.97	2.27 - 3.59
	rnh201 top1 (Q)	21	8.24	5.41 - 24.4	30	0.63	0.57 - 0.74

	top1 (R)	21	3.21	1.68 - 5.23			
pol3-L612M	*WT* (S)	21	4.04	2.61 - 6.94	30	1.94	1.24 - 2.75
	rnh201 (T)	21	12.20	7.81 - 17.52	45	4.39	3.67 - 5.87
	rnh201 top1 (U)	21	2.85	1.62 - 4.42	30	1.04	0.63 - 1.17
	top1 (V)	21	2.07	1.38 - 4.85			

The same numerical values presented in this table are shown graphically in Fig. 2.2 in the main manuscript. The genotype statistics identity (stID) for each genotype was used for reference in the pairwise comparisons shown in Table S2.3.

Table S2.3 Statistical significances of pertinent pairwise recombination rates comparisons.

Genotype 1	stID 1	Genotype 2	stID 2	LOH assay *p*-value	NAHR assay *p*-value
WT	A	*rnh201*	B	< 0.0001*	0.2817
rnh201	B	*rnh201 top1*	E	< 0.0001*	0.068
WT	A	*top1*	F	0.0853	0.0201*
pol2-M644G	G	*pol2-M644G rnh201*	H	< 0.0001*	< 0.0001*
pol2-M644G rnh201	H	*pol2-M644G rnh201 top1*	I	< 0.0001*	< 0.0001*
pol2-M644G	G	*pol2-M644G top1*	J	< 0.0001*	NA
pol2-M644L	K	*pol2-M644L rnh201*	L	0.0004*	0.0871
pol2-M644L rnh201	L	*pol2-M644L rnh201 top1*	M	0.0071*	< 0.0001*
pol2-M644L	K	*pol2-M644L top1*	N	0.0058*	NA
pol1-L868M	O	*pol1-L868M rnh201*	P	< 0.0001*	< 0.0001*
pol1-L868M rnh201	P	*pol1-L868M rnh201 top1*	Q	0.0001*	< 0.0001*
pol1-L868M	O	*pol1-L868M top1*	R	0.0181*	NA
pol3-L612M	S	*pol3-L612M rnh201*	T	< 0.0001*	< 0.0001*
pol3-L612M rnh201	T	*pol3-L612M rnh201 top1*	U	< 0.0001*	< 0.0001*
pol3-L612M	S	*pol3-L612M top1*	V	0.077	NA
WT	A	*pol2-M644G*	G	< 0.0001*	< 0.0001*
WT	A	*pol2-M644L*	K	0.011*	< 0.0001*
WT	A	*pol1-L868M*	O	0.3027	< 0.0001*
WT	A	*pol3-L612M*	S	0.7835	0.3928
pol2-M644G rnh201	H	*rnh201*	B	< 0.0001*	< 0.0001*
pol2-M644L rnh201	L	*rnh201*	B	0.0349*	< 0.0001*

pol1-L868M rnh201	P	*rnh201*	B	0.1821	0.4479
pol3-L612M rnh201	T	*rnh201*	B	0.0026*	< 0.0001*
pol3-L612M	S	*pol2-M644L*	K	0.0253*	< 0.0001*
pol3-L612M rnh201	T	*pol2-M644L rnh201*	L	0.1799	< 0.0001*
pol2-M644G rnh201 top1	I	*rnh201 top1*	E	0.0016*	< 0.0001*
pol2-M644L rnh201 top1	M	*rnh201 top1*	E	0.5972	< 0.0001*
pol1-L868M rnh201 top1	Q	*rnh201 top1*	E	0.4415	< 0.0001*
pol3-L612M rnh201 top1	U	*rnh201 top1*	E	< 0.0001*	< 0.0001*

* Asterisks indicate cases in which the pairwise comparison between the recombination rates obtained for genotype 1 were significantly different (p-value <0.05) from the recombination rates obtained from genotype 2. stID 1 and stID 2 correspond to the statistics identity for each genotype from Table S2.2. Statistical analyses of comparisons between recombination rates were performed using a two-sided nonparametric Mann Whitney test in GraphPad Prism.

Table S3.1 Yeast strains used in quantitative recombination assays.

		HAPLOIDS PARENTS				LOH ASSAY DIPLOIDS
GENOTYPES	*MAT*	*a*	α	*a*	α	*a/α*
	MAL13	+	+	*::CORE2*	*::CORE2*	*+/::CORE2*
MUTANT BACKGROUNDS		JAY# ↓	JAY# ↓	JAY# ↓	JAY# ↓	JAY# ↓
Wild Type			1167	1168		1201
rnh201Δ::Hph			1161	1165		1487,1488
rnh201-RED			1883,1884	1889,1890		1897,1898
top1Δ::Nat			1163	1166		1402
rnh201Δ::Hph top1Δ::Nat			1164	1162		1200
rnh201-RED top1Δ::Nat		1892			1893,1894	1899,1900
rnh1Δ::Kan			1938	1939		1958,1959
rnh1Δ::Kan rnh201Δ::Hph			1881,1882	1886		1896,1909
rnh1Δ::Kan rnh201-RED			1920	1921		1918
rnh1Δ::Kan top1Δ::Nat			1937	1926		1942,1943
rnh1Δ::Kan rnh201-RED top1Δ::Nat		1922			1923	1919
pol2-M644G			1313	1318		1432
pol2-M644G rnh201Δ::Hph			1320	1321		1428,1429
pol2-M644G rnh201-RED		1901			1903,1904	1912,1913
pol2-M644G top1Δ::Nat			1574,1576	1573,1575		1593,1594
pol2-M644G rnh201Δ::Hph top1Δ::Nat			1324	1325		1430,1431
pol2-M644G rnh201-RED top1Δ::Nat			1905,1906	1908		1914

pol2-M644G rnh1Δ	2036			2035	2040
pol2-M644G rnh1Δ rnh201-RED	2038			2037	2041
pol2-M644L		1312	1319		1466,1467
pol2-M644L rnh201Δ::Hph		1322	1323		1464,1465
pol2-M644L rnh201-RED	1928			1927	1929,1946
pol2-M644L top1Δ::Nat		1578,1580	1577,1579		1595,1596
pol2-M644L rnh201Δ::Hph top1Δ::Nat		1342	1422,1423		1515,1461
pol2-M644L rnh201-RED top1Δ::Nat	1924			1925	1917

Strain table footnotes: The numbers in the table represent the JAY strain collection numbers for all haploid parents and the resulting diploids used in the LOH assays. In addition to the indicated genotypes, all strains have the following markers: *ade5-1*, *his7-2*, *leu2-3,112* Leu$^+$, *trp1-289*, *cup1Δ*, *RSC30*, *sfa1D::hisG*, *lys2Δ0*, *ura3Δ0*. Two or more independently built strains were used where more than one number appears. The diploids were made by mating haploids from the corresponding row. For example, the JAY1201 LOH assay wild type diploid was constructed by mating JAY1168 to JAY1167. The strains in rows indicated in red color contain the *rnh201-RED* and/or the *rnh1Δ:Kan* mutations and were generated for the present study. The rows in black color correspond to strains previously described in Conover *et al.* 2015 [17]. The *rnh201-RED* allele (*rnh201-P45D-Y219A*) mutant sequence was cloned into the *URA3* integrative shuttle vector pRS306, generating plasmid pDC4. The DNA sequence of *rnh201-RED* in pDC4 is identical to that in the SGD reference S288c genome, with the exception of two double nucleotide mutations: The P45D amino acid substitution is the result of a GA to CC mutation that eliminates a *Ban*II restriction site present in the wild type allele; the Y219A amino acid substitution is the result of a CG to TA mutation that creates a *Ban*II restriction site in the *rnh201-RED* allele. Therefore, PCR followed by *Ban*II digestion can be used to distinguish the genotype at *RNH201* locus. *Sph*I-linearized pDC4 was transformed into the JAY1167 strain, and Ura$^+$ transformants were selected. Ura$^+$ candidates were screened by PCR-*Ban*II to identify clones carrying integration of pDC4 at the *RNH201* locus and presence of the *rnh201-RED* mutations. Independent "pop-in" Ura$^+$ strains were used to obtain independent "pop-out" 5-FOA resistant derivatives, screened by PCR-*Ban*II, yielding strains JAY1883 and JAY1884 that carried the *rnh201-RED* allele. These were crossed to various isogenic strains, and following tetrad dissection, all haploid genotype combinations necessary for the LOH assays were obtained. Additional detailed information about DNA sequences, plasmid constructions and PCR primer sequences are available upon request.

Table S3.2 Quantitative analysis of loss-of-heterozygosity.

	Genotype (stID)	stID	Total cultures	LOH rate x2x10^{-5}/cell/division	
				Median	95% CI
POL2	*RNH201 RNH1 TOP1*	A	53	3.63	2.83 - 4.28
	rnh201Δ RNH1 TOP1	B	54	21.25	17.41 - 28.42
	rnh201-RED RNH1 TOP1	C	32	11.27	7.29 - 15.65
	RNH201 RNH1 top1Δ	D	29	2.71	1.82 - 3.64
	rnh201Δ RNH1 top1Δ	E	42	7.46	6.37 - 9.62
	rnh201-RED RNH1 top1Δ	F	32	5.13	2.59 - 7.27
	RNH201 rnh1Δ TOP1	G	20	3.98	2.11 - 6.44
	rnh201Δ rnh1Δ TOP1	H	38	47.60	27.20 - 64.60
	rnh201-RED rnh1Δ TOP1	I	29	10.21	7.07 - 12.88
	RNH201 rnh1Δ top1Δ	J	19	1.35	1.06 - 3.96
	rnh201-RED rnh1Δ top1Δ	K	19	4.93	3.08 - 8.55
pol2-M644G	*RNH201 RNH1 TOP1*	L	52	14.68	12.12 - 18.20
	rnh201Δ RNH1 TOP1	M	49	73.58	58.36 - 89.86
	rnh201-RED RNH1 TOP1	N	40	72.27	57.96 - 105.43
	RNH201 RNH1 top1Δ	O	33	5.68	4.73 - 7.24
	rnh201Δ RNH1 top1Δ	P	27	13.25	8.65 - 27.02
	rnh201-RED RNH1 top1Δ	Q	31	12.39	8.72 - 14.98

	RNH201 rnh1Δ TOP1	X	50	10.98	8.41 - 14.46
	rnh201-RED rnh1Δ TOP1	Y	29	111.95	88.90 - 182.48
pol2-M644L	*RNH201 RNH1 TOP1*	R	30	7.75	5.38 - 11.70
	rnh201Δ RNH1 TOP1	S	31	14.22	10.41 - 20.67
	rnh201-RED RNH1 TOP1	T	31	3.93	2.38 - 6.48
	RNH201 RNH1 top1Δ	U	30	3.66	2.55 - 5.28
	rnh201Δ RNH1 top1Δ	V	33	8.49	6.14 - 10.75
	rnh201-RED RNH1 top1Δ	W	31	3.65	2.24 - 4.26

Data table footnotes: The same numerical values presented in this table are shown graphically in Fig. 3.1C in the main manuscript. The statistics identity (stID) for each genotype was used for reference in the pairwise comparisons shown in Table S3.3. The data in rows indicated in red color were newly generated for the present study. The rows in black color present composite data incorporating cultures previously reported in Conover *et al.* 2015 and additional cultures newly generated for the present study. In all cases, the LOH rates in the newly generated data were not statistically different from those reported previously. The composite data, with their higher number of cultures, often had narrower 95% CIs and provided higher resolution for statistical comparisons between genotypes. The only exception was for strain JAY1432 (*pol32-M644G RNH201 RNH1 TOP1*) for which we previously reported a rate of 28.42 (25.71-36.77) x2x10^{-5}/cell/division obtained from 19 cultures. The rate reported here from 52 new cultures is roughly half of that measured earlier. This higher confidence LOH rate measurement does not affect the conclusions presented in Conover *et al.* 2015.

Table S3.3 Statistical significances of pertinent pairwise recombination rates comparisons.

Genotype 1	stID 1		Genotype 2	stID 2	p-value
POL2 RNH201 RNH1 TOP1	A	*vs.*	*POL2 rnh201Δ RNH1 TOP1*	B	< 0.0001*
POL2 RNH201 RNH1 TOP1	A	*vs.*	*POL2 rnh201-RED RNH1 TOP1*	C	< 0.0001*
POL2 rnh201Δ RNH1 TOP1	B	*vs.*	*POL2 rnh201-RED RNH1 TOP1*	C	0.0003*
POL2 RNH201 RNH1 TOP1	A	*vs.*	*POL2 RNH201 RNH1 top1Δ*	D	0.05
POL2 rnh201Δ RNH1 TOP1	B	*vs.*	*POL2 rnh201Δ RNH1 top1Δ*	E	< 0.0001*
POL2 rnh201-RED RNH1 TOP1	C	*vs.*	*POL2 rnh201-RED RNH1 top1Δ*	F	< 0.0001*
POL2 RNH201 RNH1 TOP1	A	*vs.*	*POL2 RNH201 rnh1Δ TOP1*	G	0.8133
POL2 rnh201-RED RNH1 TOP1	C	*vs.*	*POL2 rnh201-RED rnh1Δ TOP1*	I	0.4111
POL2 rnh201-RED RNH1 top1Δ	F	*vs.*	*POL2 rnh201-RED rnh1Δ top1Δ*	K	0.6190
POL2 RNH201 rnh1Δ TOP1	G	*vs.*	*POL2 rnh201Δ rnh1Δ TOP1*	H	< 0.0001*
POL2 RNH201 rnh1Δ TOP1	G	*vs.*	*POL2 rnh201-RED rnh1Δ TOP1*	I	<0.0001*

POL2 rnh201Δ rnh1Δ TOP1	H	*vs.*	*POL2 rnh201-RED rnh1Δ TOP1*	I	< 0.0001*
POL2 RNH201 rnh1Δ TOP1	G	*vs.*	*POL2 RNH201 rnh1Δ top1Δ*	J	0.0300*
POL2 RNH201 RNH1 top1Δ	D	*vs.*	*POL2 RNH201 rnh1Δ top1Δ*	J	0.2099
POL2 rnh201-RED rnh1Δ TOP1	I	*vs.*	*POL2 rnh201-RED rnh1Δ top1Δ*	K	0.0117*
POL2 RNH201 RNH1 TOP1	A	*vs.*	*pol2-M644G RNH201 RNH1 TOP1*	L	< 0.0001*
POL2 rnh201Δ RNH1 TOP1	B	*vs.*	*pol2-M644G rnh201Δ RNH1 TOP1*	M	< 0.0001*
POL2 rnh201-RED RNH1 TOP1	C	*vs.*	*pol2-M644G rnh201-RED RNH1 TOP1*	N	< 0.0001*
pol2-M644G RNH201 RNH1 TOP1	L	*vs.*	*pol2-M644G rnh201Δ RNH1 TOP1*	M	< 0.0001*
pol2-M644G RNH201 RNH1 TOP1	L	*vs.*	*pol2-M644G rnh201-RED RNH1 TOP1*	N	<0.0001*
pol2-M644G rnh201Δ RNH1 TOP1	M	*vs.*	*pol2-M644G rnh201-RED RNH1 TOP1*	N	0.9330
pol2-M644G RNH201 RNH1 TOP1	L	*vs.*	*pol2-M644G RNH201 RNH1 top1Δ*	O	< 0.0001*
pol2-M644G rnh201Δ RNH1 TOP1	M	*vs.*	*pol2-M644G rnh201Δ RNH1 top1Δ*	P	< 0.0001*
pol2-M644G rnh201-RED RNH1	N	*vs*	*pol2-M644G rnh201-RED RNH1*	Q	<

TOP1		.	*top1Δ*		0.0001*
pol2-M644G RNH201 RNH1 TOP1	L	*vs*.	*pol2-M644G RNH201 rnh1Δ TOP1*	X	0.0288*
pol2-M644G rnh201-RED RNH1 TOP1	N	*vs*.	*pol2-M644G rnh201-RED rnh1Δ TOP1*	Y	0.0068*
pol2-M644G RNH201 rnh1Δ TOP1	X	*vs*.	*pol2-M644G rnh201-RED rnh1Δ TOP1*	Y	<0.0001
POL2 RNH201 RNH1 TOP1	A	*vs*.	*pol2-M644L RNH201 RNH1 TOP1*	R	< 0.0001*
POL2 rnh201Δ RNH1 TOP1	B	*vs*.	*pol2-M644L rnh201Δ RNH1 TOP1*	S	0.0036*
POL2 rnh201-RED RNH1 TOP1	C	*vs*.	*pol2-M644L rnh201-RED RNH1 TOP1*	T	< 0.0001*
pol2-M644L RNH201 RNH1 TOP1	R	*vs*.	*pol2-M644L rnh201Δ RNH1 TOP1*	S	0.0086*
pol2-M644L RNH201 RNH1 TOP1	R	*vs*.	*pol2-M644L rnh201-RED RNH1 TOP1*	T	0.0007*
pol2-M644L RNH201 RNH1 TOP1	R	*vs*.	*pol2-M644L RNH201 RNH1 top1Δ*	U	0.0002*
pol2-M644L rnh201Δ RNH1 TOP1	S	*vs*.	*pol2-M644L rnh201Δ RNH1 top1Δ*	V	0.0051*
pol2-M644L rnh201-RED RNH1 TOP1	T	*vs*.	*pol2-M644L rnh201-RED RNH1 top1Δ*	W	0.3795

Statistics table footnotes: * Asterisks indicate cases in which the pairwise comparison between the recombination rates obtained for genotype 1 were significantly different (p-value <0.05) from the recombination rates obtained from genotype 2. stID 1 and stID 2 correspond to the statistics identity for each genotype from Table S3.2. Statistical analyses of comparisons between recombination rates were performed using a two-sided nonparametric Mann Whitney test in GraphPad Prism software.

Table S4.1Yeast strains used in meiotic recurrent non-allelic homologous recombination assays.

LCR Strain	MAT	*Kan*MX4	*Hph*MX4	Experimental Diploid
FCR8	α	871	NA	NA
0Kb	*a*	-	2008	2025
	α	2026	-	
5Kb	*a*	1952	-	2117, 2118
	α	-	2113, 2114	
15Kb	*a*	880, 881	-	1849, 1852
	α	-	879	
24Kb	*a*	1954	-	1965
	α	-	1950	
35Kb	*a*	-	1843	1910, 1911
	α	1888	-	

Common Genotype: *ade5-1 his7-2 leu2-3,112 LEU+ ura3-52 trp1-289 cup1D RSC30 sfa1D::hisG DDI1::SFA1-V208I-CUP1 prox::Nat*

Table S4.2 Meiotic recurrent non-allelic homologous recombination primer sequences.

JAO	5'-3' Sequence	Description
697	TTGGTTGTAGAGTGAGTAGCGACAGGCGGCAGC GGAATATAAGAAGGATAAAATGAGAATACTTAG AACG *AGGCCACTAGTGGATCTG*	Homology left of *Ty1-1*, downstream of MX4, used to create all deletions in FCR8
698	GCGCACATTGAGGCACAAATACAATCTGCATAT TATATACTTAACAGAAGTACAATCATATACAAT ACAA *AGCTGAAGCTTCGTACGC*	Right of *SFA1-CUP1*, upstream of MX4, used to create 35Kb LCR Reverse
699	CGTGATGGAAATTTCAAACGTCAACAAAAGCAG TAGTGAAAGGTTAACTACTTGTAAAGTTATATT ACAA *AGCTGAAGCTTCGTACGC*	Inside *BEM2*, upstream of MX4, used to create 5Kb LCR, from FCR8 Reverse
700	ACATTGATTCTTAAGGATGATAATGG	Reverse primer to check 39kb deletion
1433	CTCTAGTTAAAGACTATTCGATCATAAAAGCAT CATAACTGCACCAGCCAAACCTAGGGCTCCTGG GATT *AGCTGAAGCTTCGTACGC*	Inside *SPI1*, upstream of MX4, used to create 15Kb LCR from FCR8 Reverse
1434	AACTCCCTATTGCACAACGGC	Reverse primer downstream of *SPI1* to check integration that makes 24Kb LCR
1435	GCGACTTGATCCTTAGCAAACTTTCCATTTTCT GCCAATTCAGTTATAGTTTCATAATATTTGGAA TATG *AGCTGAAGCTTCGTACGC*	Inside *COG3*, upstream of MX4, used to create 24KB LCR from FCR8 Reverse
1608	TCTGTGGTTGCCATCGATCC	ddPCR *SFA* Forward
1609	GAAGCGCCCCTTAACTTTGC	ddPCR *SFA* Reverse
1612	CAGCGATCGCGTATTTCGTC	ddPCR *Kan*MX4 Forward
1613	AGTGACGACTGAATCCGGTG	ddPCR *Kan*MX4 Reverse
1614	TGGCAAACTGTGATGGACG	ddPCR *Hph*MX4 Forward
1615	TTGGGAATCCCCGAACATCG	ddPCR *Hph*MX4 Reverse
1618	TGAACCAAGAGAAGACGGCTC	ddPCR Proximal to LCR Forward – single copy control
1619	CCCTCCAATTGCTGCTTTAGTC	ddPCR Proximal to LCR Reverse – single copy control