Title

The Roles of the RNases H and Chromosomal Sequences in DNA:RNA Hybrid Mediated Genome Instability

Author

Zimmer, Anjali Diana

Publication Date

2016

Peer reviewed|Thesis/dissertation

The Roles of the RNases H and Chromosomal Sequences in DNA:RNA Hybrid Mediated Genome Instability

By

Anjali Diana Zimmer

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Douglas Koshland, Chair
Professor Nicholas Ingolia
Professor Jasper Rine
Professor Daniel Zilberman

Summer 2016

Abstract

The Roles of the RNases H and Chromosomal Sequences in

DNA:RNA Hybrid Mediated Genome Instability

by

Anjali Diana Zimmer

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Douglas Koshland, Chair

DNA:RNA hybrids can lead to DNA damage and genome instability. This damage can be prevented by degradation of the RNA in the hybrid by two evolutionarily conserved enzymes, RNase H1 and RNase H2. Indeed, RNase H deficient cells have increased chromosomal rearrangements. However, the quantitative and spatial contributions of the individual enzymes to hybrid removal has been unclear. Additionally, RNase H2 can remove single ribonucleotides misincorporated into DNA during replication. The relative contribution of DNA:RNA hybrids and misincorporated ribonucleotides to chromosome instability was also uncertain. To address these issues, we studied the rate and location of loss of heterozygosity events on chromosome III in *Saccharomyces cerevisiae* that were defective for RNase H1, H2 or both. We showed that RNase H2 plays the major role in preventing chromosome III instability through its hybrid-removal activity. Furthermore, RNase H2 acts pervasively at many hybrids along the chromosome. In contrast, RNase H1 acts to prevent LOH within a small region of chromosome III, and this instability is dependent upon two hybrid prone regions. This restriction of RNase H1 activity to a subset of hybrids is not due to its constrained localization as we found it at hybrids genome wide. This result suggests that the genome protection activity of RNase H1 is regulated at a step after hybrid recognition. The global function of RNase H2 and the region specific function of RNase H1 provide insight on why these enzymes with overlapping hybrid-removal activities have been conserved throughout evolution.

Additionally we developed a novel method to map DNA:RNA hybrids in the yeast genome, termed S1-DRIP-seq. Hybrids were found to form at the same loci in wild-type and RNase H deficient cells and to form at higher levels in RNase H mutants, which corroborates previous observations in the field. This study allowed for the identification of factors that influence hybrid formation and found that high transcription is both strongly correlated with and sufficient for hybrid formation. We then used the information gained from this high-resolution map of hybrids to further probe the factors that influence hybrid formation. We investigated if the increased hybrid levels observed

in the RNase H mutants were due to changes in transcript levels. We found that the transcript levels at hybrid-forming loci were not increased in cells lacking the RNases H. We also investigated whether genes that are highly expressed and form hybrids at their endogenous locus can form hybrids at an ectopic locus. We found that hybrid-prone genes did form hybrids when inserted onto a yeast artificial chromosome (YAC), showing that hybrid-prone sequences can form hybrids outside their native chromosomal context. The presence of these hybrid-forming sequences did not cause instability of the YAC, even when overexpressed.

Together, the work presented in this dissertation identifies several properties of hybrids and the systems in place for their removal from the yeast genome. We found that the RNases H do not act equally to prevent chromosome instability, and they also act differentially at specific hybrids. We found that hybrid-prone sequences can form hybrids outside their native chromosomal context. Finally, our studies of hybrid-related instability on chromosome III and the YAC showed that not all hybrid-forming sequences lead to increased chromosome instability.

Table of Contents

Acknowledgements

I would like to thank the members of the Koshland Lab, past and present, for their invaluable training and assistance. Firstly I thank team hybrid who I have worked most closely with. Thank you to Lamia Wahba who trained me from my first day in the lab and initiated the transcription induced instability on the YAC experiments and the DNA:RNA hybrid genome-wide mapping project. She taught me many lab techniques, edited proposals and manuscripts, and kept many experiments running even after leaving the lab. Thank you to Jeremy Amon, my Koshland Lab "co-ton", who has been here with me, working on similar projects, talking through experiments, and helping me through stumbling blocks in and out of lab. His friendship has been a foundation throughout grad school. Also thank you to Lorenzo Costantino who has been wonderful to work with on the DRIP-seq mapping studies. He taught me many things including how to prioritize experiments, how to get things done, and which flour to use for pizza dough.

Many other members of the lab provided critical training and discussion. Thank you to Hugo Tapia, my bay mate who is always helpful to bounce ideas off of and who taught me many of the tools of yeast genetics. He made our bay the coolest bay with planets hanging above and good music always playing. Thank you to Vinny Guacci who has endless knowledge of yeast biology, reagents and techniques. He has also been a great source of productive, and occasionally unproductive, discussions. Thank you to Thomas Eng for teaching me many techniques and helping me navigate through grad school. Thank you to Fred Tan who helped me dip into the world of bioinformatics. Thank you to Aaron Welch for bringing laughter into the lab and helping me get started. Thank you to Gamze Camdre for her friendship, advice, and levelheadedness. Thank you to Michelle Bloom for being a wonderful friend to navigate through the storms of grad school. You have made me stronger, of mind and especially of body as an amazing workout buddy. Thank you to Brett Robison for being a source of both deep conversation and true enthusiasm. Thank you to Rebecca Lamothe for bringing optimism, friendship and arts and crafts into the lab. Thank you to Ryan Holly for bringing the perfect balance of fun and no-nonsense attitude. Thank you to Steven Gore, Weijia Chua and Jeremiah Stricklin for being wonderful colleagues and keeping the lab in ship shape. Thank you to Xuchen Hu, Amna Saeed, and Skylar Kim for their enthusiasm, levity and great talks. Finally, a special thank you to my advisor and our lab's enthusiastic leader, Doug Koshland. He taught me productive and positive interpretation of science. His guidance through the development of this project and myself as a scientist was essential.

Thank you to my classmates, MCB 2010, for being a great support network and friends with whom to celebrate the good times. Thank you to all my friends, friendsies, and extended family for their love, support and grounding. A huge thank you to my family, Mom, Appa, and Alex, who have loved, encouraged, and bolstered me my entire life. And finally, an enormous thank you to my loving husband Brian, without whom this degree would not have been possible. You have been my support, my cheerleader, my joy, and partner in life.

Chapter 1: Introduction to DNA:RNA hybrid formation and associated genome instability

The transmission of genetic information through generations is a fundamental biological process. If DNA is not faithfully replicated and segregated during cell division, mutations can arise. Mutations are the basis of evolution; some mutations can lead to evolutionarily advantageous changes in an organism. However, many mutations are deleterious to the organism and can lead to disease or death. In relation to human disease, a hallmark of cancer is a high burden of mutation and genome instability. For these reasons, there has been much interest and study of DNA damage and genome instability.

Genome instability can be induced by exogenous agents, such as irradiation and chemical mutagens, or by the malfunction of endogenous processes, such as DNA replication or chromosome segregation. In recent years, it has become more appreciated that the endogenous processes of transcription and RNA biogenesis are also potential instigators of genome instability. Work in the field has elucidated a connection between transcription and genome instability that involves the formation of a DNA:RNA hybrid intermediate leading to the misrepair of DNA double strand breaks (DSBs). This introductory chapter outlines our current understanding of how transcription contributes to genome instability.

Transcription can induce DNA damage and mutation

Transcription as a source of mutagenesis

The first hints that transcription is related to DNA damage have been termed transcription-associated mutagenesis (TAM) and transcription-associated recombination (TAR). TAM was first observed in *E. coli*, as the increased mutagenesis of genes upon induction of their transcription in the presence of chemical mutagens (Brock 1971; Herman & Dworkin 1971). Similar observations were made in budding yeast using a galactose inducible promoter to drive transcription of the LYS2 gene (Datta & Jinks-Robertson 1995).

Studies of genomic recombination and the transcription of repeated sequences revealed a similar phenomenon. TAR has been observed in lambda phage, *E. coli*, budding and fission yeast, and mammals (reviewed in (Aguilera 2002)). Both the integrity of the transcription machinery and the sequence being transcribed can affect the rate of TAR (Chávez & Aguilera 1997; Chávez et al. 2001). Indeed, the first observation of TAR in lambda was that mutants in transcription termination had increased recombination (Ikeda & Matsumoto 1979). The most thoroughly studied transcription machinery that protects against TAR is the multi-subunit THO/TREX complex, which is involved in transcription elongation in both yeast and humans (Gómez-González et al. 2011; Domínguez-Sánchez et al. 2011). Mutants lacking THO/TREX subunits exhibit up to 100 fold greater rates of recombination between direct repeats during impaired transcription elongation (Prado et al. 1997; Chávez et al. 2000). The mechanisms through which defects in transcription elongation and termination contribute to increased

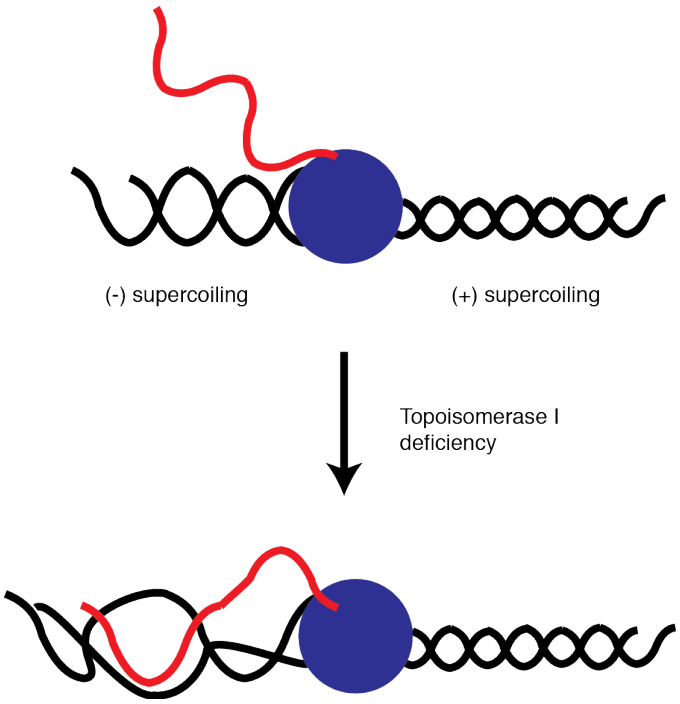


Figure 1.1
Model of supercoiling in the absence of topoisomerase I leading to R-loop formation

Actively transcribing RNA polymerase (diagramed as a blue circle) can induce positive supercoiling ahead of the polymerase and negative supercoiling behind. This supercoiling is normally relieved by topoisomerase I, but in its absence, supercoiling can build and be susceptible to invasion by RNA (diagrammed as the red strand) to form and R-loop.

---

recombination were initially unknown, but were ultimately revealed to involve the formation of DNA:RNA hybrids and R-loops in the genome.

R-loops are associated with genome instability

R-loops form when an RNA molecule anneals to a homologous DNA molecule, forming a DNA:RNA hybrid and a displaced single stranded DNA. The first studies of the formation of R-loops in relation to transcription were performed in *E. coli* cells deficient in topoisomerase I (*topA*). Topoisomerase I relaxes supercoiled DNA, and in *E. coli* it was observed that *topA* co-localized with sites of active transcription (Dürrenberger et al. 1988). Topoisomerase I can relieve the negative supercoils that build up behind an actively transcribing RNA polymerase (Liu & Wang 1987). Interestingly, *topA* mutants accumulated R-loops in plasmid DNA and exhibited a growth defect (Drolet et al. 1994; Drolet et al. 1995). The growth defect could be suppressed by

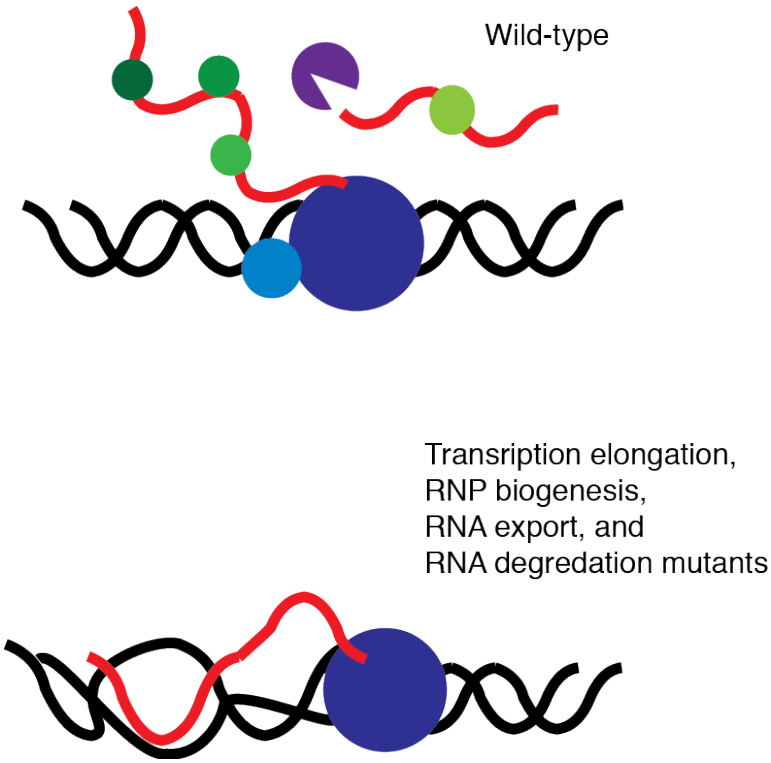


Figure 1.2
Model of R-loop formation in transcription and RNA biogenesis mutants

The top panel models RNA transcription and ribonucleoprotein formation in wild-type cells. Upon exiting from the actively elongating RNA polymerase (polymerase diagramed as a blue circle, elongation factors as light blue circle), the RNA molecule is bound by a variety of RNA binding proteins, including splicing factors and RNA export factors, diagrammed as green circles. Furthermore aberrant RNA molecules in the nucleus are degraded by RNA degradation factors, diagramed in purple. However, in the absence of these protective factors, the RNA molecule may be available to invade the DNA template and form and R-loop structure, as diagrammed in the lower panel.

---

overexpression of RNase H, which is an enzyme that degrades the RNA in a DNA:RNA hybrid. RNase H plays a key role in the prevention of R-loop formation and will be described in more detail below. These series of observations lead to the following model: if negative supercoils build up behind RNA polymerase and are not relieved by topoisomerase I, RNA can invade the DNA and form R-loops (Figure 1.1).

Another breakthrough in the field came when it was demonstrated that R-loops are the mutagenic intermediate in the TAR observed in transcription elongation mutants in budding yeast. Huertas and Aguilera (2003) (Huertas & Aguilera 2003) showed that in the THO complex mutant *hpr1*Δ, TAR was exacerbated in cells lacking the RNases H

and suppressed by overexpression of RNase H. R-loops as an intermediate in transcription-associated instability turned out to be common to many different transcription processes and conserved in many organisms. In chicken and human cells, the depletion of splicing factors lead to increased DNA damage, which again could be suppressed by RNase H (Li & Manley 2005; Paulsen et al. 2009).

R-loops have also been associated with a particularly disruptive form of genome instability: gross chromosomal rearrangements (GCRs). GCR describes large chromosomal aberrations including translocations and partial or complete loss of chromosomes. Two large-scale screens in yeast to find factors important for preventing GCRs identified many genes related to transcription and RNA biogenesis (Wahba et al. 2011; Stirling et al. 2012). Loss of genes involved in processes including global transcription repression, transcription initiation, RNA degradation, RNA export, mRNA cleavage and polyadenylation increased chromosomal instability. The mechanism of this instability involves the formation of an R-loop intermediate, and could be suppressed by the overexpression of RNase H. A model to explain how these factors suppress R-loop formation is presented in Figure 1.2. Immediately after transcription, the nascent mRNA is contacted by various RNA binding proteins, spliced, packaged, and exported from the nucleus. If this process fails, the unbound RNA can reinvade the DNA duplex to form an R-loop.

Once these R-loops form, cells posses several mechanisms to remove them. The RNases H can degrade the RNA in a DNA:RNA hybrid, and are discussed in detail in the next section. In addition to the RNases H, cells have many RNA-DNA helicases, which could potentially remove hybrids, including the Pif1 family of helicases, DHX9 helicase, and Senataxin helicase. Senataxin helicase (Sen1 in yeast) has been implicated in removing hybrids. Loss of Sen1 in yeast leads to an accumulation of hybrids and induces TAR (Mischo et al. 2011). In human cells, Senataxin and the related helicase Aquarius also resolve R-loops to prevent DNA damage (Skourti-Stathaki et al. 2011; Sollier et al. 2014).

The RNases H prevent the accumulation of DNA:RNA hybrids in the genome

Organisms from bacteria through humans have two RNases H that can degrade the RNA in a DNA:RNA hybrid. Cells lacking the RNases H have elevated levels of R-loops genome-wide and increased genome instability (Wahba et al. 2011; Chan et al. 2014). There are two classes of RNases H; the first is termed RNase H1 in eukaryotes (RNase HI in prokaryotes) and the second is RNase H2 (RNase HII in prokaryotes) (Cerritelli & Crouch 2009). The two RNases H appear to have overlapping activities. How they differ in relation to preventing R-loop mediated damage has been is an open question and a major motivator of my doctoral research.

RNase H1 degrades RNA in R-loops

In both prokaryotes and eukaryotes, RNase H1 is a monomeric enzyme which is able to degrade an RNA molecule involved in a DNA:RNA hybrid. Typical eukaryotic RNase H1 enzymes contain a hybrid binding domain (HBD) and a catalytic RNase H

domain (Cerritelli & Crouch 1998). The RNase H1 proteins of human and mouse also contains a mitochondrial targeting sequence (MTS), however, the MTS is absent from the *S. cerevisiae* protein (Cerritelli & Crouch 2009). Although most prokaryotic RNase HI proteins lack a HBD, the structure of the RNase H domain is highly conserved between *E. coli* and humans (Yang et al. 1990; Kanaya 1998; Morikawa & Katayanagi 1998), highlighting the fundamental importance of RNase H enzymes to basic cellular functions. This RNase H domain catalyzes the degradation of 4 or more RNA bases in a DNA:RNA duplex. Its processive degradation of RNA is dependent on the HBD (Nowotny et al. 2008). Concordantly, the prokaryotic enzymes lack the HBD and cleave the substrate distributively (Gaidamakov et al. 2005).

Eukaryotic RNase H2 degrades R-loops and single ribonucleotides imbedded in DNA

While RNase H1 is a monomeric enzyme, the eukaryotic RNase H2 is a three-subunit complex. This complex is comprised of one enzymatic subunit (encoded by *RNH201* in *S. cerevisiae*) and two regulatory subunits (encoded by *RNH202* and *RNH203* in *S. cerevisiae*). All three subunits are necessary for activity in yeast, although the roles of *RNH202* and *RNH203* remain unclear. The human RNase H2 complex also contains three subunits, and while the catalytically active subunit is well conserved with yeast *RNH201* and is similar to type two prokaryotic enzymes, there is little to no conservation with the two regulatory subunits of yeast (Chon et al. 2009). The human regulatory subunit RNASEH2B interacts with PCNA, which likely coordinates its DNA repair functions (Chon et al. 2009; Sparks et al. 2012).

RNase H2 also differs from RNase H1 in its activities. Like RNase H1 it can processively cleave long runs of ribonucleotides in a DNA:RNA hybrid, which I will refer to as R-loop removal. Additionally, it can cleave a single ribonucleotide imbedded into duplex DNA (Jeong et al. 2004). This additional activity beyond RNase H1 might imply that the major role of RNase H2 is specialized to the removal of ribonucleotides that are mistakenly incorporated into DNA during replication. However, several different studies, some of them presented in this dissertation, argue that RNase H2 is the major R-loop remover in eukaryotic cells, and that this activity protects against chromosome instability. In biochemical assays, RNase H2 performs the major R-loop removing RNase H activity. Crude extracts of yeast cells deficient in RNase H2 (*rnh201*Δ*)* exhibit an approximately 70% decrease in RNase H activity (Arudchandran et al. 2000; Frank et al. 1998). Mouse embryo extracts lacking RNase H2 exhibit a 90% decrease in RNase H activity (Reijns et al. 2012).

During DNA replication, ribonucleotides can be mistakenly incorporated into DNA. In *S. cerevisiae* an estimated 10,000 ribonucleotides are incorporated in each round of replication (Nick McElhinny et al. 2010). In order to maintain the integrity of the genome, these ribonucleotides must be recognized, excised and replaced. The pathway by which this is accomplished is termed the ribonucleotide excision repair (RER) pathway. RER is initiated by RNase H2 when it cleaves the 5' end of the misincorporated ribonucleotide. Strand displacement synthesis is then carried out by DNA polymerase δ, followed by flap removal by Fen1 or Dna2, and ligation by DNA ligase 1 (Sparks et al. 2012).

Defects and Diseases associated with RNase H mutants

While the RNases H are present and conserved through prokaryotes and eukaryotes, neither are essential in bacteria or budding yeast (Rydberg & Game 2002; Arudchandran et al. 2000). However, null mutations in either enzyme are lethal in multicellular eukaryotes. RNase H1 mutants of *Drosophila melanogaster* and mouse are embryonic lethal (Filippov et al. 2001; Cerritelli et al. 2003). The embryonic lethality in mice was linked to defects in mitochondrial replication. Human B-lymphocytes depleted of RNase H1 show increased chromosome breaks (Helmrich et al. 2011). Recently, a point mutation in RNase H1 has been associated with an adult-onset neuromuscular disease (Reyes et al. 2015). In cells carrying the mutant form of RNase H1, a mitochondrial R-loop is in low abundance, leading to aggregation of mitochondrial DNA, implying that a specific, regulated R-loop is necessary for mitochondrial replication (Akman et al. 2016).

Similar to RNase H1, RNase H2 mutants of mouse are embryonic lethal (Reijns et al. 2012). The embryos accumulate large amounts of ribonucleotides into their genomic DNA and mount a p53-dependent DNA damage response. In humans, mutations in all three RNase H2 subunits lead to a neuroinflammatory autoimmune disease called Aicardi-Goutières Syndrome (AGS). Twenty-nine different point mutations in the three subunits have been identified in AGS patients (Figiel et al. 2011). Many of these mutations have been shown to decrease RNase H2 enzymatic activity, but the precise mechanism for the autoimmune response remains unclear. One hypothesis is that persistent R-loops in the genome mimic viral infections and stimulate an immune response, while another hypothesis says that the DNA damage and damage response caused by misincorporated ribonucleotides triggers an innate immune response (reviewed in (Feng & Cao 2016)).

S. cerevisiae has proven to be a very useful model organism to study the eukaryotic RNases H. While the RNases H are dispensable for viability, mutants deficient for RNase H1 and RNase H2 experience DNA damage and genome instability. RNase H single and double mutants are sensitive to the DNA damaging agents hydroxyurea and ethyl methanesulphonate (Arudchandran et al. 2000). RNase H double mutants accumulate high levels of genomic R-loops (Wahba et al. 2011; Chan et al. 2014). Interestingly, unlike other R-loop accumulating transcription mutants, RNase H mutants do not increase TAR (Huertas & Aguilera 2003). However, RNase H mutants exhibit elevated chromosome instability of both artificial and endogenous yeast chromosomes (Wahba et al. 2011; O'Connell et al. 2015).

As described above, RNase H2 can remove ribonucleotides that are misincorporated into DNA. Yeast mutants lacking RNase H2 accumulate ribonucleotides in their genomic DNA. These lesions appear to be processed by pathways that are mutagenic and induce chromosome instability (Kim et al. 2011; Conover et al. 2015; O'Connell et al. 2015). These mutagenic outcomes are dependent on topoisomerase I, which can cleave the at the ribonucleotide and leave a gap to be filled in by an error-free or error-prone polymerase, or the backbone can be religated without synthesis, resulting in 2-5 base pair deletions (Kim et al. 2011; Sparks & Burgers 2015; Cho et al. 2013). In addition to these small 2-5 bp deletions, RNase H2 deficient cells also exhibit increased chromosome instability. Interestingly, the chromosome instability phenotype of RNase

H2 mutants is dependent on topoisomerase I; the double mutant has near wild-type levels of instability (Conover et al. 2015). This instability was observed to increase when combined with a DNA polymerase mutant that incorporates more ribonucleotides into DNA (Conover et al. 2015). However there are conflicting reports as to whether the chromosome instability phenotype of an RNase H2 mutant combined with a DNA polymerase mutant that incorporates fewer ribonucleotides increases or decreases (Conover et al. 2015; O'Connell et al. 2015). In chapter 2 of this dissertation I present results to clarify these seemingly contradictory observations.

Sequence contribution to hybrid formation

Sequences prone to R-loop formation

R-loop formation is also influenced by the properties of the DNA and/or RNA sequence involved in the hybrid. Initial studies of the THO complex mutant impaired for transcription *hpr1*Δ in yeast introduced the idea that the sequence being transcribed affects TAR. In the *hpr1*Δ mutant, transcription of the E. coli derived LacZ gene caused a hyper-recombination phenotype. However, transcription of a yeast derived *PHO5* gene did not induce recombination (Chávez & Aguilera 1997). Further study revealed that Hpr1 was important for preventing TAR from G+C rich or long transcripts (Chávez et al. 2001). More recent work in human cell lines has shown that transcription of the longest human genes is associated with RNA:DNA hybrid mediated fragile site instability (Helmrich et al. 2011), again implying that transcript length is an important factor. These long human genes require longer than one cell division to complete transcription.

G-rich DNA is a promoter of R-loop formation. *In vitro* analysis of R-loop formation has identified sequences that promote both their initiation and elongation. G-clusters (GGGG) on the nontranscribed strand are strong promoters of R-loop initiation, and general G density downstream promotes R-loop elongation (Roy et al. 2008; Roy & Lieber 2009). R-loops also form at a variety of G-rich repeats *in vivo*. In *E. coli* and human cells, R-loops form at CTG repeats and induce their instability (Lin et al. 2010). GAA repeats form hybrids *in vitro* and *in vivo* in *E. coli* (Grabczyk et al. 2007). *In vitro* transcription of G-containing CTG, CAG, CGG, CCG, and GAA repeats also yield R-loops (Reddy et al. 2011), and in human cell extracts, CTG and GGGCC repeats form R-loops (Reddy et al. 2014).

R-loop formation in cis and trans

Certain sequences are hybrid prone, and it appears that these sequences can form hybrids in both *cis* and *trans.* That is, an RNA molecule can form a hybrid at the site of active transcription or at another homologous sequence in the genome that is not being transcribed. Evidence for co-transcriptional R-loop formation came from the early studies of TAR. In *hpr1*Δ transcriptional elongation mutants, TAR and persistent hybrid formation is dependent on the integrity of the RNA. If the RNA is able to self-cleave by encoding a hammerhead ribozyme, there is no longer recombination or hybrid persistence (Huertas & Aguilera 2003). Additionally, hybrids can form in *trans*, in a manner

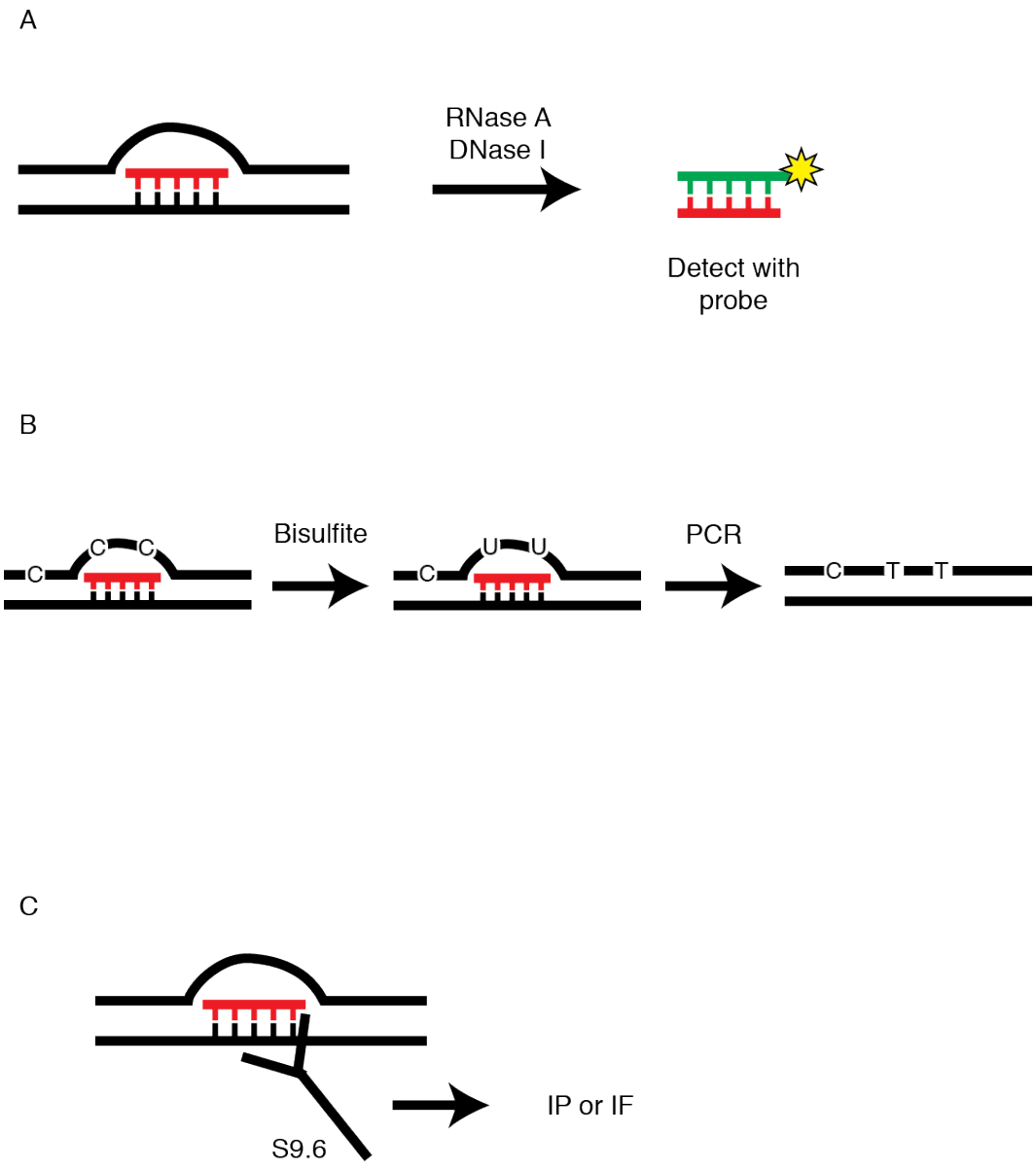


Figure 1.3
Techniques for the detection of R-loops

Diagrams of methods used to detect R-loops.
(A) Isolation of RNA involved in R-loops by sequential nuclease digestion with RNase A and DNase I followed by detection with a labeled DNA probe.
(B) Bisulfite sensitivity assay. Bisulfite converts C in single stranded, but not double stranded DNA into U. These changes can be detected after PCR amplification and sequencing in order to infer the location of ssDNA involved in R-loops.
(C) Detection of R-loops by S9.6 antibody against DNA:RNA hybrids. The antibody has been used to immunoprecipitate (IP) or for detection by immunofluorescence (IF).

dependent on the homologous recombination machinery protein Rad51 (Wahba et al. 2013). In these experiments, an RNA transcribed from a locus on chromosome III was able to stimulate hybrid formation of a homologous sequence on an artificial chromosome and induce the instability of that chromosome. These effects could be reduced by either the overexpression of RNase H or deletion of Rad51. Together, these studies of transcription and hybrid formation in yeast show that hybrid-prone sequences can form R-loops in both *cis* and *trans.*

Techniques for the detection of R-loops

Many studies have used either the deletion or overexpression of the RNases H as a genetic test for the presence of DNA:RNA hybrids. In addition to genetics there are several biochemical techniques used to detect R-loops *in vivo.*

Nuclease digestion

R-loops can be detected by differential digestion with specific nucleases (Figure 1.3A). To detect hybrids at the model hybrid-forming TAR locus, total nucleic acids were extracted from yeast (Huertas & Aguilera 2003). These nucleic acids were then treated with combinations of RNase A, DNase I and RNase H. RNase A degrades single-stranded RNA when incubated at higher salt concentrations and DNase I degrades both single- and double-stranded DNA. As described in detail above, RNase H degrades the RNA in a DNA:RNA hybrid. RNA in DNA:RNA hybrids was detected by digestion of total nucleic acids with RNase A and DNase I, followed by spotting on a membrane and probing with a DNA probe. These sequences were confirmed to be involved in a DNA:RNA hybrid because they were sensitive to RNase H digestion.

Bisulfite sequencing

R-loops formed *in vivo* can also be detected by bisulfite modification followed by sequencing (Figure 1.3B). Single-stranded DNA treated with bisulfite results in the conversion of deoxycytosine to deoxyuracil (C to U) (Gough et al. 1986). Upon amplification by PCR, DNA polymerase with place an A opposite the U, ultimately resulting in mutation of the original C to T (Yu et al. 2003). Cytosine bases in double-stranded DNA are unaffected by DNA treatment. The bisulfite treated and PCR amplified molecules are then sequenced to determine the location of induced mutations. Using this technique, the non-template, displaced, single-stranded DNA in an R-loop can be detected.

Antibody-based detection

The most popular techniques for R-loop detection make use of the S9.6 mouse monoclonal antibody which specifically recognizes DNA:RNA hybrids (Boguslawski et al. 1986; Hu et al. 2006) (Figure 1.3B). This antibody has been used to detect hybrids by probing total nucleic acids spotted on a membrane (Wahba et al. 2013), immunofluorescence of nuclei on spread on glass slides (Wahba et al. 2011; Chan et al.

2014), and for immunoprecipitation of R-loops followed by qPCR or sequencing (Hage et al. 2010; Ginno et al. 2012; Hage et al. 2014). While there has been some concern about cross-reaction of the antibody with double-stranded RNA (Zhang et al. 2015), the antibody has been shown to be specific to DNA:RNA hybrids in all the above contexts by control experiments abolishing the antibody signal upon *in vivo* overexpression of RNase H and *in vitro* RNase H treatment (Wahba et al. 2011; Wahba et al. 2016; Chan et al. 2014).

Mapping R-loops in the genome

Approaches used for genome-wide mapping of hybrids

R-loops have been mapped genome-wide in *E. coli,* budding yeast and human cells. Most of these studies relied on the immunoprecipitation of DNA:RNA hybrids by the S9.6 antibody. The methods have differed in pre-treatment and fragmentation of nucleic acids, microarray- or sequencing-based analysis, and sequencing library preparation.

The only genome-wide study of R-loops that did not utilize the S9.6 antibody was done in *E. coli* using the bisulfite sensitivity assay (described above) followed by deep sequencing. This assay was performed in wild-type cells, *nusG* mutants defective in Rho-dependent transcription termination, and in wild-type cells overexpressing RNase HI (Leela et al. 2013). The study found that the locations of R-loops were similar in wild-type and *nusG* mutants strains, but the frequency of R-loop formation was greater in *nusG* mutants. They concluded that R-loops naturally form in many loci in the genome, and that this is further exacerbated by defects in transcription termination. This conclusion was further supported by the fact that fewer R-loops form in wild-type cells overexpressing RNase HI.

R-loops were mapped in human pluripotent Ntera2 cells (Ginno et al. 2012) using two different techniques to precipitate hybrids: precipitation of chromatin using a catalytically dead RNase H1 that binds to, but does not degrade, DNA:RNA hybrids, termed DRIVE (DNA:RNA *in vitro* enrichment), and immunoprecipitation by the S9.6 antibody, termed DRIP (DNA:RNA immunoprecipitation). The DRIVE technique identified 1,224 peaks and DRIP identified 20,862 peaks. The overlap between these peaks was not very complete, and future studies from the same authors have relied solely on DRIP (Ginno et al. 2013). For the DRIP protocol, genomic DNA was digested with a cocktail of restriction enzymes. This was done to gently fragment the DNA for immunoprecipitation and sequencing library preparation, while preserving the R-loops. The specificity of the technique was shown by pre-treating the genomic DNA with RNase H *in vitro* and seeing the abolishment of DRIP signal.

Following this initial application of DRIP to mapping R-loops in human cells, three different studies have mapped R-loops genome-wide in yeast (Chan et al. 2014; Hage et al. 2014; Wahba et al. 2016). The third study was performed in the Koshland Lab and included my work, which will be described in more detail in chapter 3 of this dissertation. The initial report by Chan et al. termed their technique DRIP-Chip. In this study genomic DNA was sheared by sonication, R-loops were immunoprecipitated by the S9.6 antibody, PCR amplified and analyzed by microarray hybridization. In the second

study by El Hage et al., termed ChIP-seq, a similar initial shearing and precipitation protocol was performed, followed by deep sequencing. While both of these studies identified regions of hybrid formation, their results differed from each other and from other previous observations in the field. The DRIP-Chip study found that a large proportion of the genome was hybrid forming, around one-third. In contrast, the ChIP-seq experiment found that only a few loci were hybrid-forming. Both studies saw no difference in location or levels of hybrid formation between wild-type and RNase H deficient cells. This was contradictory to previous cytological data in the field showing that hybrids accumulate to much higher levels in RNase H mutants (Wahba et al. 2011; Chan et al. 2014). Neither of these studies had a large signal difference between the peaks and background loci. We therefore hypothesized that their poor signal was due to the harsh shearing conditions of sonication.

For our study, we optimized a protocol to preserve hybrids and termed it S1-DRIP-seq (Wahba et al. 2016). We determined that the process of sonication was destroying hybrids, and hypothesized that this was due to melting of the DNA:RNA hybrid by energy transferred during sonication, and replacement with reannealed DNA:DNA pairing. To prevent DNA pairing, we first treated extracted genomic DNA with S1 nuclease to degrade the single-stranded DNA molecule displaced in the R-loop. Following S1 nuclease treatment, most of the genomic hybrids were maintained through sonication. We then immunoprecipitated the hybrids and prepared sequences libraries. This refinement of the technique allowed for high resolution mapping with a high signal to background ratio.

Recently there have been two additional studies refining the DRIP technique in human cell lines. One study employed sonication to achieve higher resolution as well as performing strand-specific library preparation in a technique they term RDIP-seq (RNA:DNA immunoprecipitation) (Nadel et al. 2015). This report did not comment on the effect of sonication on hybrid retention, but did suffer from lower signal to background ratios. The most recent report termed their technique DRIPc-seq (DNA:RNA immunoprecipitation followed by cDNA conversion) (Sanz et al. 2016) and also sought to obtain higher resolution and strand-specific information. To accomplish this, they still gently fragmented the DNA using a cocktail of restriction enzymes, and immunoprecipitated the R-loops. They next removed the DNA from the DNA:RNA hybrids by DNase I digest and then reverse transcribed the RNA into cDNA to be used in sequencing library preparation. They applied this technique to a panel of human cell lines as well as two mouse cell lines.

Hybrid-prone loci in human cells

These genome-wide studies have elucidated genomic features that are prone to hybrid formation. A major finding from the studies in human cell lines is that R-loops are a feature in some promoter elements. Specifically they are common in unmethylated stretches of DNA in promoters called CpG islands (Ginno et al. 2012). These CpG islands have a high GC skew; that is the non-template strand is G-rich. Further study of these GC skewed CpG islands revealed four subclasses of promoters that were more or less prone to hypomethylation (Ginno et al. 2013). Only those with the highest GC density, CpG composition, and GC skew were likely to be hypomethylated and have R-

loop formation. The R-loop forming promoters appear to have a more open chromatin conformation, as they are highly accessible to DNase I and MNase (Sanz et al. 2016). Using data sets of histone modifications and transcription associated proteins, Sanz et al. showed that R-loop positive promoters show signatures of open and active promoters. Together, these studies suggest a role of R-loops in influencing epigenetic states.

R-loops were additionally found at the 3' end of genes, where they were only slightly correlated with GC skew (Ginno et al. 2013; Sanz et al. 2016). The R-loop was broad and peaked just prior to the polyadenlyation site. Indeed, 3' R-loops were found only in genes that undergo poly-A dependent termination. These R-loops appear to be involved in transcription termination as they influence RNA polymerase II positioning and enrich the transcription termination factor Paf1.

Hybrids have also been found in human cells in previously studied hybrid-forming loci from yeast. R-loops were detected by RDIP-seq in the ribosomal DNA (rDNA) locus (Nadel et al. 2015). RDIP signal was detected at intergenic spacer regions within the rDNA, indicating that these hybrids likely form soon after transcription and before rRNA maturation. Hybrids were also detected at repetitive sequences, specifically at satellite repeats, simple repeats and low complexity repeats. Interestingly, they were not detected at repetitive transposable elements as they are in yeast, as described below.

Hybrid-prone loci in yeast cells

Even before the genome-wide mapping efforts, the rDNA had been known to be a locus of strong hybrid formation (Hage et al. 2010). Subsequently, all three yeast mapping studies identified the rDNA as a locus of hybrid formation (Chan et al. 2014; Hage et al. 2014; Wahba et al. 2016). In both the ChIP-seq (Hage et al. 2014) and S1-DRIP-seq (Wahba et al. 2016) experiments, the rDNA represented a large proportion of reads in the immunoprecipitated (IP) samples. In the ChIP-seq experiment, while 11% of reads mapped to the rDNA in the total input, ~30-40% of reads from the IP samples were rDNA, dependent on genotype. In our S1-DRIP-seq, ~10% of input and 46% of IP reads mapped to the rDNA.

A major class of hybrid-forming sequences identified by the genome-wide mapping studies are transposable Ty elements. While Ty elements were identified as loci of strong hybrid formation in all three studies, there was some controversy over whether the R-loops being detected are present at the Ty genomic locus or in the reverse transcribed cDNA particles that are part of the Ty element's life cycle, as proposed by El Hage et al. (Hage et al. 2014). However, we extensively studied the nature of hybrid formation at Ty elements and found that Ty elements do form hybrids at their genomic loci, as will be discussed in more detail in chapter 3 (Wahba et al. 2016).

While the rDNA and Ty loci are strong hybrid-forming loci, about a quarter of hybrids mapped to unique sequences in the genome (Wahba et al. 2016). One factor that strongly influences hybrid formation in genes is their transcription level. The first indication of this came from the DRIP-Chip study (Chan et al. 2014) which found that hybrid forming genes tended to be highly transcribed and higher in GC content. However due to the limited signal and resolution of this study they were unable to determine if this correlation was confounded by the fact that highly transcribed genes are GC rich. The ChIP-seq study also found that genes with high levels of transcription and GC richness

were more hybrid prone. Finally, our S1-DRIP-seq study found both a near-perfect correlation and a sufficiency to high levels of transcription on hybrid formation. We did not see a significant difference in GC content between expression-matched hybrid and non-hybrid genes, suggesting that previously reported correlations might have been due solely to expression level.

Hybrids additionally form at several classes of noncoding RNAs: tRNAs, snRNAs , snoRNAs and telomeric TERRA RNAs. Hybrids at tRNA loci appear to be especially susceptible to the RNases H, and in their absence hybrid levels are greatly increased (Chan et al. 2014; Hage et al. 2014). El Hage et al. observed that the increased hybrid formation in RNase H deficient cells at tRNA loci altered the dynamics of tRNA maturation, indicating an important role of the RNases H in preventing issues in tRNA synthesis.

A few other sequences were found to be hybrid prone in the individual mapping studies. Chan et al. found that antisense transcripts were associated with hybrid formation. El Hage et al. found that intron-containing genes were prone to hybrid formation in their second exon, especially so in ribosomal protein genes. Our S1-DRIP-seq study additionally identified long homopolymeric dA:dT tracts as highly predictive of hybrid formation.

Lastly, high levels of hybrid formation were detected in the mitochondrial genome. These hybrids were especially elevated in RNase H deficient cells. El Hage et al. showed that this elevation was due to the loss of RNase H1 but not RNase H2. This is concordant with the mitochondrial defects observed in mammalian cells with RNase H1 defects.

Conversion of R-loops into DNA damage

Damage induced by AID

One way that hybrids have been shown to be converted into DNA damage is by the action of nucleases that specifically recognize R-loops. R-loops form at the switch (S) region of the Ig locus in B cells as part of B cell maturation, as will be described in detail below. The R-loops are recognized by the endonuclease activation-induced cytidine deaminase (AID). AID promotes the conversion of dC to dU, which activates the base excision repair process in which uracil DNA glycosylase excises the uracil. This results in an abasic site and a single-strand DNA break (Basu et al. 2011). However, AID is not expressed in most cells types and there is no homolog in yeast, so it is unlikely that this is a major mechanism by which R-loops are converted into damage. AID is able to induce TAM and TAR at R-loop-forming loci when ectopically expressed in yeast (Gómez-González & Aguilera 2007).

Damage induced by the nucleotide excision repair machinery

The nucleotide excision repair (NER) pathway normally recognizes and repairs bulky lesions in DNA by excising an oligonucleotide including the lesion and filling in the gap with repair synthesis. However, studies in human and yeast cells have implicated NER proteins in instigating R-loop-associated DNA damage. In mutants with high loads

of R-loops, depletion of the NER factors XPG or XPF (or deletion of the yeast homolog *RAD2*) leads to the accumulation of R-loops and decreases DNA damage (Sollier et al. 2014). These observations lead to a model by which XPG and/or XPF, which are flap endonucleases, recognize the R-loop as a flap-like bulky lesion and make single stranded DNA nicks. It remains unclear how this is further processed into double strand breaks and DNA damage.

Collisions between R-loops and DNA replication machinery

One prominent model for the conversion of R-loops into DNA damage involves collisions between the DNA replication fork and R-loops and/or stalled RNA polymerases. The TAR phenotypes of transcription elongation mutant *hrp1*Δ yeast cells were shown to be stimulated by transcription during S-phase, but not during G2 (Wellinger et al. 2006). Additionally *hpr1*Δ mutants show synthetic lethality with mutants defective in the S-phase checkpoint (Gómez-González et al. 2009). In *E. coli* cells, R-loop forming regions impair replication fork progression and cells must be replication competent in order to undergo R-loop-mediated genomic instability (Gan et al. 2011). Similarly, human cells depleted for splicing factors, which have been shown to form R-loops, induce the DNA double strand break marker γ-H2AX primarily in cells in S-phase (Gan et al. 2011). Many studies in yeast, *C. elegans*, and human cells have shown that R-loop forming mutants experience replication stress and DNA damage (Santos-Pereira et al. 2013; Castellano-Pozo et al. 2012; Helmrich et al. 2011). However, the mechanism by which replication fork impairment is converted into DNA damage and genome instability has yet to be determined.

Deleterious R-loop damage processing by the break induced replication pathway

While R-loops may be processed into DNA damage by one or more of the mechanisms explored above, another requirement of genome instability is the improper repair of the DNA damage resulting in mutation or loss of genetic information. Recent work in the Koshland lab has explored the repair pathways by which R-loop-induced damage is repaired (Amon and Koshland, in preparation). They found that depletion of topoisomerase I in *rnh1*Δ *rnh201*Δ cells induces high levels of R-loops and DNA damage, and results in lethality. Interestingly, this DNA damage accumulates after S-phase during G2. They show that the lethality is dependent on the break induced replication pathway. These results suggest a model in which R-loops both induce DNA damage and inhibit repair.

Positive roles of scheduled R-loops

While the accumulation of R-loops can lead to deleterious genome instability, R-loops have also been shown to play roles in many physiological processes. How these scheduled R-loops are prevented from causing DNA damage is still an open question.

Class Switch Recombination

The first documented case of R-loops playing a physiological role was in class switch recombination (CSR) at the immunoglobulin locus during B cell maturation. CSR is a programmed rearrangement of the locus encoding the immunoglobulin constant region of the heavy chain in an antibody. This process involves the enzyme activated cytidine deaminase (AID) which targets cytosines on single-stranded DNA and coverts them to uracils. RNA polymerase II transcribes at the switch (S) regions resulting in R-loop formation (Yu et al. 2003). This R-loop provides a single strand of DNA as a target for AID. How this action by AID results in recombination is unclear, but may involve NHEJ.

Maintenance of Epigenetic marks

As described above, studies mapping R-loops genome-wide in human cells found that hybrids are found at CpG island promoters, which characteristically lack DNA methylation (Ginno et al. 2012). In that study they also showed that R-loop formation was sufficient to protect DNA from methylation by the DNA methyltransferase DMNT3B1. R-loops regions were also enriched for the promoter-associated histone marks like H3K4 di- and tri-methylation and H3K36 tri-methylation (Sanz et al. 2016). These correlative studies suggest that R-loops may affect epigenetic marks and create open chromatin at promoters.

In contrast to the open promoter-like states observed in human cells, R-loops have also been associated with condensed heterochromatin. In fission yeast, the formation of heterochromatin around the centromere is mediated by a DNA:RNA hybrid formed by a noncoding RNA (Nakama et al. 2012). In budding yeast, the R-loop-accumulating mutant *hpr1*Δ has a modest increase in accumulation of the histone mark H3S10P, in an R-loop-dependent manner. H3S10P is considered to be a mark of chromatin condensation. This accumulation of H3S10P was also observed in *C. elegans* and human cells (Castellano-Pozo et al. 2013). The discrepancy in types of chromatin marks correlated to R-loop formation and the mechanism by which R-loops may mediate these chromatin changes remains unclear.

Transcription termination

R-loops are also involved in the regulation of transcription termination. In both the yeast and human genomes, R-loops are found at the 3' end of genes (Wahba et al. 2016; Ginno et al. 2013). 3' R-loops have been associated with poly(A) dependent transcription termination in mammalian cells. R-loops facilitate RNA Pol II pausing and Senataxin is required to resolve these R-loops for efficient transcription termination. In a more convoluted pathway, R-loops also can form at the 3' end of a gene and encourage antisense transcription at the locus. This then results in a double-stranded RNA that is processed by the RNAi machinery and promotes the formation of repressive heterochromatic sites with H3K9me2 and HP1γ, which enhance RNA Pol II pausing before termination (Skourti-Stathaki et al. 2014).

Antisense noncoding RNAs

In addition to a role in heterochromatin formation and transcription termination, R-loop forming antisense noncoding RNAs have been shown to regulate genes. In *Arabidopsis thaliana* the long noncoding RNA transcript *COOLAIR* is antisense to the floral repressor gene *FLC* and is involved in its silencing during vernalization. After prolonged periods of cold an R-loop forms over the promoter of *COOLAIR* repressing it, allowing for activation of *FLC* (Sun et al. 2013). In contrast, in human cells, transcription of the vimentin (VIM) gene is promoted by the formation of an R-loop formed by an antisense transcript (Boque-Sastre et al. 2015). In yeast an R-loop of a long noncoding RNA has been shown to increase the rate of induction of the galactose utilization genes (Cloutier et al. 2016). Further investigation of the ways that R-loops formed by antisense transcripts modulate transcription will be of interest.

Aims of research

In this dissertation I present experiments further exploring the roles of DNA:RNA hybrids in genome instability. In chapter 2, I investigate the differential roles that the RNases H play in preventing hybrid-mediated chromosome instability. In chapter 3, I investigate the roles of chromosomal context on hybrid formation. Finally in the appendix I present some preliminary explorations of the contribution of specific sequences to hybrid formation and chromosome instability.

References

Aguilera, A., 2002. The connection between transcription and genomic instability. *The EMBO Journal*, 21(3), pp.195–201.

Akman, G. et al., 2016. Pathological ribonuclease H1 causes R-loop depletion and aberrant DNA segregation in mitochondria. *Proceedings of the National Academy of Sciences of the United States of America*, p.201600537.

Arudchandran, A. et al., 2000. The absence of ribonuclease H1 or H2 alters the sensitivity of Saccharomyces cerevisiae to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair. *Genes to cells : devoted to molecular & cellular mechanisms*, 5(10), pp.789–802.

Basu, U. et al., 2011. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell*, 144(3), pp.353–363.

Boguslawski, S.J. et al., 1986. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *Journal of immunological methods*, 89(1), pp.123–130.

Boque-Sastre, R. et al., 2015. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(18), pp.5785–5790.

Brock, R.D., 1971. Differential mutation of the beta-galactosidase gene of Escherichia coli. *Mutation research*, 11(2), pp.181–186.

Castellano-Pozo, M. et al., 2013. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Molecular Cell*, 52(4), pp.583–590.

Castellano-Pozo, M., García-Muse, T. & Aguilera, A., 2012. R-loops cause replication impairment and genome instability during meiosis. *EMBO reports*, 13(10), pp.923–929.

Cerritelli, S.M. & Crouch, R.J., 1998. Cloning, expression, and mapping of ribonucleases H of human and mouse related to bacterial RNase HI. *Genomics*, 53(3), pp.300–307.

Cerritelli, S.M. & Crouch, R.J., 2009. Ribonuclease H: the enzymes in eukaryotes. *The FEBS journal*, 276(6), pp.1494–1505.

Cerritelli, S.M. et al., 2003. Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Molecular Cell*, 11(3), pp.807–815.

Chan, Y.A. et al., 2014. Genome-Wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with DRIP-Chip M. Snyder, ed. *PLoS Genetics*, 10(4), p.e1004288.

Chávez, S. & Aguilera, A., 1997. The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes & development*, 11(24), pp.3459–3470.

Chávez, S. et al., 2000. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. *The EMBO Journal*, 19(21), pp.5824–5834.

Chávez, S. et al., 2001. Hpr1 is preferentially required for transcription of either long or G+C-rich DNA sequences in Saccharomyces cerevisiae. *Molecular and cellular biology*, 21(20), pp.7054–7064.

Cho, J.-E. et al., 2013. Two distinct mechanisms of Topoisomerase 1-dependent mutagenesis in yeast. *DNA repair*, 12(3), pp.205–211.

Chon, H. et al., 2009. Contributions of the two accessory subunits, RNASEH2B and RNASEH2C, to the activity and properties of the human RNase H2 complex. 37(1), pp.96–110. Available at: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkn913.

Cloutier, S.C. et al., 2016. Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation. *Molecular Cell*, 61(3), pp.393–404.

Conover, H.N. et al., 2015. Stimulation of Chromosomal Rearrangements by Ribonucleotides. *Genetics*, 201(3), pp.951–961.

Datta, A. & Jinks-Robertson, S., 1995. Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science*, 268(5217), pp.1616–1619.

Domínguez-Sánchez, M.S. et al., 2011. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. C. E. Pearson, ed. *PLoS Genetics*, 7(12), p.e1002386.

Drolet, M. et al., 1995. Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proceedings of the National Academy of Sciences of the United States of America*, 92(8), pp.3526–3530.

Drolet, M., Bi, X. & Liu, L.F., 1994. Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *The Journal of biological chemistry*, 269(3), pp.2068–2074.

Dürrenberger, M. et al., 1988. Intracellular location of the histonelike protein HU in Escherichia coli. *Journal of bacteriology*, 170(10), pp.4757–4768.

Feng, S. & Cao, Z., 2016. Is the role of human RNase H2 restricted to its enzyme activity? *Progress in biophysics and molecular biology*, 121(1), pp.66–73.

Figiel, M. et al., 2011. The structural and biochemical characterization of human RNase H2 complex reveals the molecular basis for substrate recognition and Aicardi-Goutières syndrome defects. *The Journal of biological chemistry*, 286(12), pp.10540–10550.

Filippov, V., Filippov, M. & Gill, S.S., 2001. Drosophila RNase H1 is essential for development but not for proliferation. *Molecular genetics and genomics : MGG*, 265(5), pp.771–777.

Frank, P., Braunshofer-Reiter, C. & Wintersberger, U., 1998. Yeast RNase H(35) is the counterpart of the mammalian RNase HI, and is evolutionarily related to prokaryotic RNase HII. *FEBS letters*, 421(1), pp.23–26.

Gaidamakov, S.A. et al., 2005. Eukaryotic RNases H1 act processively by interactions through the duplex RNA-binding domain. *Nucleic acids research*, 33(7), pp.2166–2175.

Gan, W. et al., 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes & development*, 25(19), pp.2041–2056.

Ginno, P.A. et al., 2013. GC skew at the 5“ and 3” ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome research*, 23(10), pp.1590–1600.

Ginno, P.A. et al., 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular Cell*, 45(6), pp.814–825.

Gough, G.W., Sullivan, K.M. & Lilley, D.M., 1986. The structure of cruciforms in supercoiled DNA: probing the single-stranded character of nucleotide bases with bisulphite. *The EMBO Journal*, 5(1), pp.191–196.

Gómez-González, B. & Aguilera, A., 2007. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proceedings of the National Academy of Sciences of the United States of America*, 104(20), pp.8409–8414.

Gómez-González, B. et al., 2011. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *The EMBO Journal*, 30(15), pp.3106–3119.

Gómez-González, B., Felipe-Abrio, I. & Aguilera, A., 2009. The S-phase checkpoint is required to respond to R-loops accumulated in THO mutants. *Molecular and cellular biology*, 29(19), pp.5203–5213.

Grabczyk, E., Mancuso, M. & Sammarco, M.C., 2007. A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic acids research*, 35(16), pp.5351–5359.

Hage, El, A. et al., 2014. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. D. Engelke, ed. *PLoS Genetics*, 10(10), p.e1004716.

Hage, El, A. et al., 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes & development*, 24(14), pp.1546–1558.

Helmrich, A., Ballarino, M. & Tora, L., 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular Cell*, 44(6), pp.966–977.

Herman, R.K. & Dworkin, N.B., 1971. Effect of gene induction on the rate of mutagenesis by ICR-191 in Escherichia coli. *Journal of bacteriology*, 106(2), pp.543–550.

Hu, Z. et al., 2006. An antibody-based microarray assay for small RNA detection. *Nucleic acids research*, 34(7), pp.e52–e52.

Huertas, P. & Aguilera, A., 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular Cell*, 12(3), pp.711–721.

Ikeda, H. & Matsumoto, T., 1979. Transcription promotes recA-independent recombination mediated by DNA-dependent RNA polymerase in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 76(9), pp.4571–4575.

Jeong, H.-S. et al., 2004. RNase H2 of Saccharomyces cerevisiae is a complex of three proteins. *Nucleic acids research*, 32(2), pp.407–414.

Kanaya, S., 1998. *Enzymatic activity and protein stability of E. coli ribonuclease HI*, Ribonucleases H.

Kim, N. et al., 2011. Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science*, 332(6037), pp.1561–1564.

Leela, J.K. et al., 2013. Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 110(1), pp.258–263.

Li, X. & Manley, J.L., 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell*, 122(3), pp.365–378.

Lin, Y. et al., 2010. R loops stimulate genetic instability of CTG.CAG repeats. *Proceedings of the National Academy of Sciences of the United States of America*, 107(2), pp.692–697.

Liu, L.F. & Wang, J.C., 1987. Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 84(20), pp.7024–7027.

Mischo, H.E. et al., 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Molecular Cell*, 41(1), pp.21–32.

Morikawa, K. & Katayanagi, K., 1998. *Crystal structure of RNase H from prokaryotes*, Ribonucleases H.

Nadel, J. et al., 2015. RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics & chromatin*, 8(1), p.46.

Nakama, M. et al., 2012. DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes to cells : devoted to molecular & cellular mechanisms*, 17(3), pp.218–233.

Nick McElhinny, S.A. et al., 2010. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11), pp.4949–4954.

Nowotny, M. et al., 2008. Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD. *The EMBO Journal*, 27(7), pp.1172–1181.

O'Connell, K., Jinks-Robertson, S. & Petes, T.D., 2015. Elevated Genome-Wide Instability in Yeast Mutants Lacking RNase H Activity. *Genetics*, 201(3), pp.963–975.

Paulsen, R.D. et al., 2009. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Molecular Cell*, 35(2), pp.228–239.

Prado, F., Piruat, J.I. & Aguilera, A., 1997. Recombination between DNA repeats in yeast hpr1delta cells is linked to transcription elongation. *The EMBO Journal*, 16(10), pp.2826–2835.

Reddy, K. et al., 2011. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic acids research*, 39(5), pp.1749–1762.

Reddy, K. et al., 2014. Processing of double-R-loops in (CAG)·(CTG) and C9orf72 (GGGGCC)·(GGCCCC) repeats causes instability. *Nucleic acids research*, 42(16), pp.10473–10487.

Reijns, M.A.M. et al., 2012. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell*, 149(5), pp.1008–1022.

Reyes, A. et al., 2015. RNASEH1 Mutations Impair mtDNA Replication and Cause Adult-Onset Mitochondrial Encephalomyopathy. *American journal of human genetics*, 97(1), pp.186–193.

Roy, D. & Lieber, M.R., 2009. G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Molecular and cellular biology*, 29(11), pp.3124–3133.

Roy, D., Yu, K. & Lieber, M.R., 2008. Mechanism of R-loop formation at immunoglobulin class switch sequences. *Molecular and cellular biology*, 28(1), pp.50–60.

Rydberg, B. & Game, J., 2002. Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26), pp.16654–16659.

Santos-Pereira, J.M. et al., 2013. The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes & development*, 27(22), pp.2445–2458.

Sanz, L.A. et al., 2016. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Molecular Cell*, 63(1), pp.167–178.

Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N.J., 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*, 516(7531), pp.436–439.

Skourti-Stathaki, K., Proudfoot, N.J. & Gromak, N., 2011. Human senataxin resolves

RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular Cell*, 42(6), pp.794–805.

Sollier, J. et al., 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular Cell*, 56(6), pp.777–785.

Sparks, J.L. & Burgers, P.M., 2015. Error-free and mutagenic processing of topoisomerase 1-provoked damage at genomic ribonucleotides. *The EMBO Journal*, 34(9), pp.1259–1269.

Sparks, J.L. et al., 2012. RNase H2-initiated ribonucleotide excision repair. *Molecular Cell*, 47(6), pp.980–986.

Stirling, P.C. et al., 2012. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes & development*, 26(2), pp.163–175.

Sun, Q. et al., 2013. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science*, 340(6132), pp.619–621.

Wahba, L. et al., 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Molecular Cell*, 44(6), pp.978–988.

Wahba, L. et al., 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes & development*, 30(11), pp.1327–1338.

Wahba, L., Gore, S.K. & Koshland, D., 2013. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *eLife*, 2, p.e00505.

Wellinger, R.E., Prado, F. & Aguilera, A., 2006. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Molecular and cellular biology*, 26(8), pp.3327–3334.

Yang, W. et al., 1990. Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science*, 249(4975), pp.1398–1405.

Yu, K. et al., 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature immunology*, 4(5), pp.442–451.

Zhang, Z.Z. et al., 2015. Complexities due to single-stranded RNA during antibody detection of genomic rna:dna hybrids. *BMC research notes*, 8(1), p.127.

Chapter 2: Differential Roles of the RNases H in preventing chromosome instability

Abstract

DNA:RNA hybrids can lead to DNA damage and genome instability. This damage can be prevented by degradation of the RNA in the hybrid by two evolutionarily conserved enzymes, RNase H1 and RNase H2. Indeed, RNase H deficient cells have increased chromosomal rearrangements. However the quantitative and spatial contributions of the individual enzymes to hybrid removal has been unclear. Additionally, RNase H2 can remove single ribonucleotides misincorporated into DNA during replication. The relative contribution of DNA:RNA hybrids and misincorporated ribonucleotides to chromosome instability was also uncertain. To address these issues, we studied the rate and location of loss of heterozygosity events on chromosome III in *Saccharomyces cerevisiae* that were defective for RNase H1, H2 or both. We showed that RNase H2 plays the major role in preventing chromosome III instability through its hybrid-removal activity. Furthermore, RNase H2 acts pervasively at many hybrids along the chromosome. In contrast, RNase H1 acts to prevent LOH within a small region of chromosome III, and this instability is dependent upon two hybrid prone regions. This restriction of RNase H1 activity to a subset of hybrids is not due to its constrained localization as we found it at hybrids genome wide. This result suggests that the genome protection activity of RNase H1 is regulated at a step after hybrid recognition. The global function of RNase H2 and the region specific function of RNase H1 provide insight on why these enzymes with overlapping hybrid-removal activities have been conserved throughout evolution.

Background

Preventing chromosome instability is an essential process for maintaining genetic information. A source of chromosome instability is the accumulation of R-loops, which form when an RNA molecule hybridizes with a portion of genomic DNA, creating a DNA:RNA hybrid and a displaced single-stranded DNA (Reviewed in (Santos-Pereira & Aguilera 2015)). One mechanism to prevent hybrid-mediated damage involves RNase H1 and RNase H2, two endogenous enzymes conserved from bacteria to humans that can degrade the RNA in R-loops (Reviewed in (Cerritelli & Crouch 2009)). RNase H2 also functions in the removal of single ribonucleotides that are inappropriately incorporated into DNA by DNA polymerases during replication. Why RNase H1 and H2, which appear to have overlapping functions, remain highly conserved across many branches of life has been an outstanding question. Two areas of inquiry that will help address this conundrum are 1) does one of the RNases H carry the major burden of preventing spontaneous R-loop mediated chromosome instability, and if so which, and 2) do the RNases H protect the same or different regions of the genome from R-loop mediated damage?

Whether the two RNases H differentially contribute to protecting against hybrid-mediated genome instability has been controversial. Studies of the inactivation of RNase

H1 in yeast have shown little effect on chromosome instability (Wahba et al. 2011; O'Connell et al. 2015) while inactivation of RNase H2 has been shown to increase chromosome instability (O'Connell et al. 2015; Conover et al. 2015).However, results from the Conover study suggested that the elevated instability in the RNase H2 deficient cells was due to elevated misincorporation of ribonucleotides rather than the failure to remove hybrids. In contrast, results from the O'Connell study suggested the opposite; elevated instability was due to hybrids rather than misincorporated ribonucleotides. The failure in these studies to reveal a prominent role of RNase H1 in protecting against hybrid-mediated chromosome instability was also surprising. Previous studies had shown that RNase H1, when constitutively overexpressed, can suppress genome-wide hybrid formation and hybrid-mediated genome instability induced by mutations in the RNA biogenesis machinery (Wahba et al. 2011; Huertas & Aguilera 2003; Stirling et al. 2012). These results suggested that RNase H1 had the ability to remove many hybrids within the cell, but only under artificial conditions of constitutive overexpression, not addressing the roles that RNases H play in physiological conditions.

Whether the two RNases H protect different regions of the genome from hybrid-mediated instability also remains unanswered. A recent study mapped the mitotic recombination events genome-wide in cells deficient for both RNase H1 and H2 (O'Connell et al. 2015). These mitotic recombination events presumably marked the sites of repair from damage induced by RNase H deficiency. No correlation was observed between the positions of these mitotic events and the positions of R-loop prone regions that had been defined by a genome-wide tiling array map (Chan et al. 2014). This failure in correlation could be attributed both to the low number of events due to the genome-wide nature of study and the low resolution of the R-loop map.

In this study we have brought a number of new strategies and tools to address these questions. First we used diploid strains of *S. cerevisiae* deficient for RNase H1 (*rnh1*Δ), RNase H2 (*rnh201*Δ), or both (*rnh1*Δ *rnh201*Δ) with different markers on the two copies of chromosome III that allowed quantitative analyses of Loss of Heterozygosity (LOH) events. These strains also allowed the mapping of the junctions of LOH as a means to localize sites of damage to potential hybrid prone regions. We also exploited a new high resolution map of hybrid prone regions throughout the genome (Wahba et al. 2016) to identify hybrids that are potentially causative for LOH. With these tools, we showed that RNase H2 is the predominant effector, through its hybrid removal activity, for preventing hybrid-induced instability at many distinct sites on chromosome III. Moreover, the existence of distinct sites suggests that many hybrids are capable of inducing damage. In contrast, we showed that RNase H1 acts preferentially to prevent hybrid-mediated instability within a single mapping interval. This instability is correlated specifically with two hybrid prone sequences within the region. Further analyses of RNase H1 localization suggests that this enzyme binds to hybrids within this interval as well as to most, if not all, hybrids in the genome. This result suggests that its specificity for hybrid removal occurs at a step after hybrid recognition. Thus the hybrid removal activities of RNases H1 and H2 have distinct quantitative and spatial functions in vivo.

Results

RNase H deficiency increases Loss of Heterozygosity

To investigate the links between the RNases H, DNA:RNA hybrids and chromosome instability, we designed an assay to measure loss of heterozygosity (LOH) in wild-type and RNase H-deficient diploid yeast. We inserted genetic markers on one homolog of chromosome III in a diploid (Figure 2.1A). We marked the right arm with the *URA3* gene inserted at the *BUD5* locus and the left arm with *LEU2*. These genetic markers allowed us to monitor loss of genetic information from both arms of the chromosome. To assay for LOH, we grew single colonies on media lacking uracil to select for cells that maintained the *URA3* marker. This selection eliminated LOH jackpots due to events that happened early in the colony formation. We then plated the cells from individual colonies onto 5-FOA plates to select for cells that lost the *URA3* marker, and therefore had undergone an LOH event during the last few cell divisions prior to harvesting the colonies (Figure 2.1B). The fraction of 5-FOA resistant colonies compared to the total number of viable cells provided an approximation of the rate of LOH (Materials and Methods). The 5-FOA resistant colonies were then replica plated to media lacking leucine to differentiate between LOH events that retained the *LEU2* and therefore had undergone partial LOH (truncation or mitotic recombination) from those that lost *LEU2* and therefore had undergone chromosome loss.

Using this assay, we investigated the chromosome instability phenotypes of wild-type and RNase H mutants. Deficiency of both RNases H (*rnh1*Δ/Δ *rh201*Δ/Δ) elevated chromosome instability 14.9 fold over wild-type levels (Figure 2.1C). These events in both wild-type and RNase H deficient cells were about half terminal LOH and half chromosome loss (Figure 2.2). These results corroborate previous reports that deletion of both of the RNases H elevates the levels of chromosome instability (Wahba et al. 2011; O'Connell et al. 2015). We then generated RNase H single mutants by taking the RNase H double mutant and adding back a single copy of either *RNH1* or *RNH201* to its endogenous locus in order to confirm linkage between the observed phenotypes and the presence or absence of a specific RNase H. A strain deficient in RNase H2 but harboring a wild-type copy of *RNH1* (*RNH1/*Δ *rnh201*Δ/Δ) showed instability levels only slightly lower than the double mutant (Figure 2.1C). In contrast, a strain deficient in RNase H1 but harboring one wild-type copy of *RNH201* (*rnh1*Δ/Δ *RNH201/*Δ) showed wild-type instability levels (Figure 2.1C). Therefore, the majority of the chromosome instability phenotype in RNase H deficient cells is due to the lack of RNase H2.

RNase H2 can perform two enzymatic activities: removal of R-loops and removal of single ribonucleotides misincorporated into DNA. The relative contribution of these activities to the maintenance of chromosome stability has remained unclear. To address this controversy, we utilized a mutant of Rnh201 that lacks the activity of removing single ribonucleotides, but retains the activity of degrading R-loops (Chon et al. 2013). This separation of function allele (*RNH201-P45D,Y219A*) was shown to have undetectable levels of single ribonucleotide removal activity in multiple in vitro and in vivo assays. Here, we refer to this allele as *RNH201-hr* (hybrid removal) because it retains hybrid removal activity. Strains harboring a copy of *RNH201-hr* at the endogenous *RNH201* locus (*rnh1*Δ/Δ *RNH201-hr/*Δ) reduced chromosome III instability

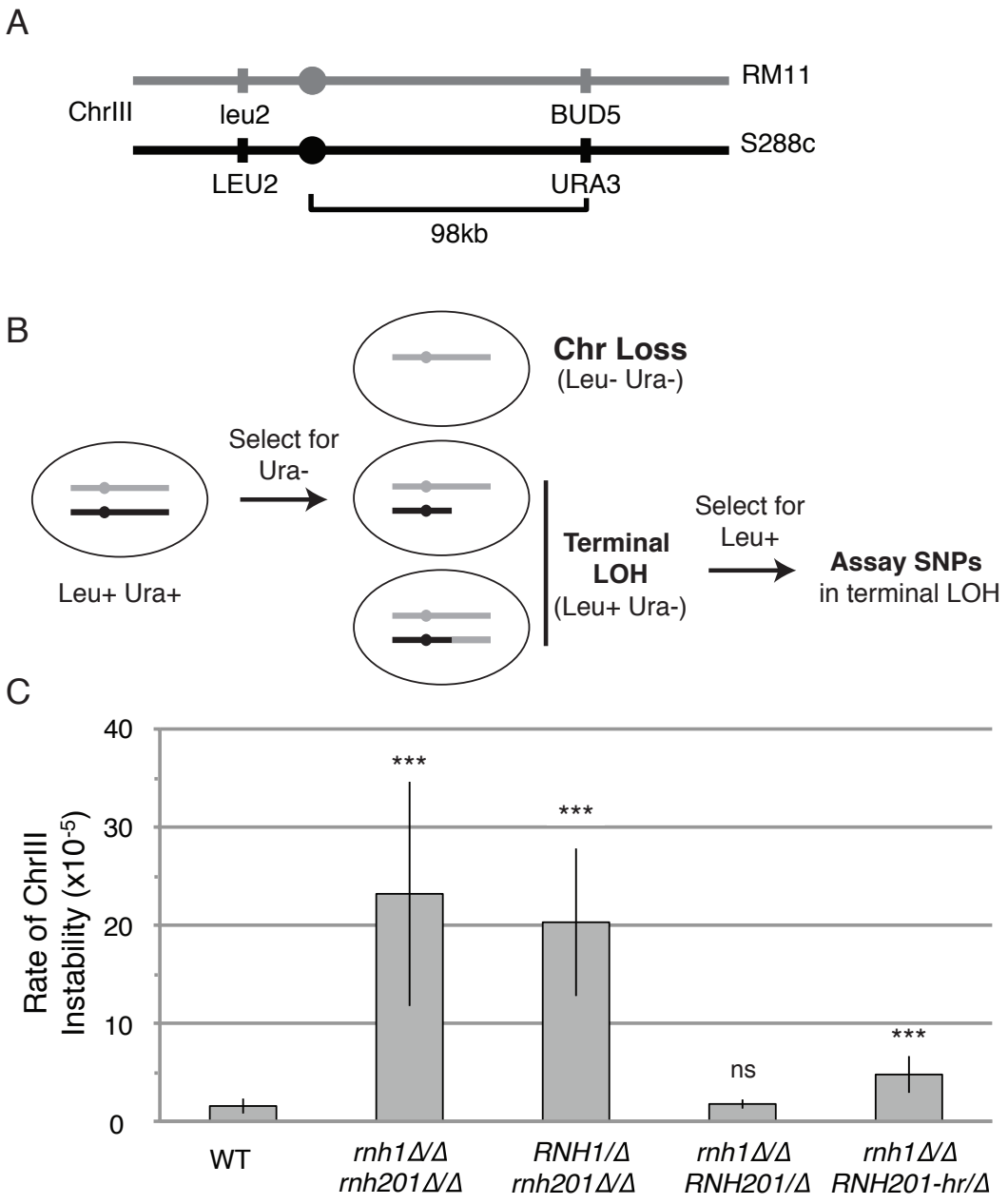


Figure 2.1
Assay for Loss of Heterozygosity (LOH) of Chromosome III in RNase H mutants

(A) Diagram of chromosome III genetic markers relevant for LOH assay.
(B) Schematic of LOH assay workflow and possible outcomes for chr III. Marked S288c derived chr is diagramed in black and RM11 derived chr is diagrammed in grey. Diploid cells (Leu+ Ura+) are propagated on media lacking uracil. Individual colonies are then plated onto media containing 5-FOA, selecting for loss of the *URA3* marker. Resultant colonies may have complete chromosome loss or a terminal LOH, shown here as a de novo telomere addition and a recombination repair event. Colonies are then replica plated to media lacking leucine to select for Leu+ colonies. The SNPs of resultant colonies are then assayed for heterozygosity by Sanger sequencing.

(C) Rates of LOH of wild-type and RNase H mutants. Mean of n >40 parent colonies are shown with error bars representing +/- 1 standard deviation. Statistical analysis comparing mutants to wild-type using unpaired t-test, *** is p<0.001, ns is not significant at p<0.1.

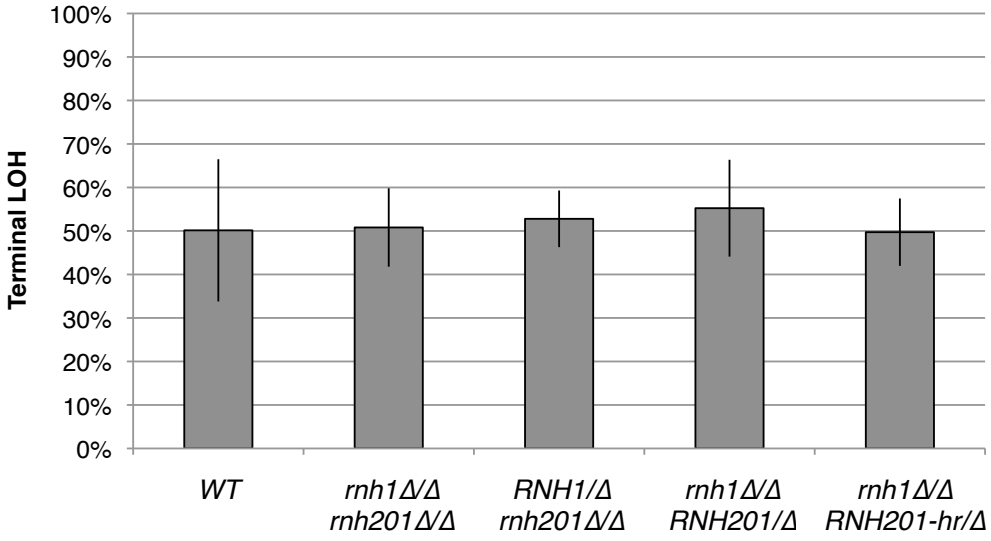


Figure 2.2
Percent Terminal LOH
The percent of colonies that underwent LOH that had terminal LOH events (Leu+, Ura-) as opposed to whole chromosome loss (Leu-, Ura-) in wild-type and RNase H mutants.

to just 3 fold over wild-type levels (Figure 2.1C). This suppression indicates that hybrids are the major cause of the LOH events and that the hybrid removal activity of RNase H2 is the main protector against hybrid mediated chromosome instability.

Mapping LOH events in RNase H deficient strains

We previously mapped hybrid-prone regions genome-wide in wild-type and RNase H mutants of haploid S288c yeast using a high resolution technique termed S1-DRIP-seq (S1 nuclease DNA:RNA immunoprecipitation with deep sequencing) (Wahba et al. 2016) and identified hybrid prone regions within this 98kb region of the right arm of chromosome III between the centromere and our *URA3* marker (Figure 2.3A). These hybrid prone regions were candidates for causing the chromosome III LOH induced by RNase H deficiency. To assess the potential causality of these hybrid regions to those

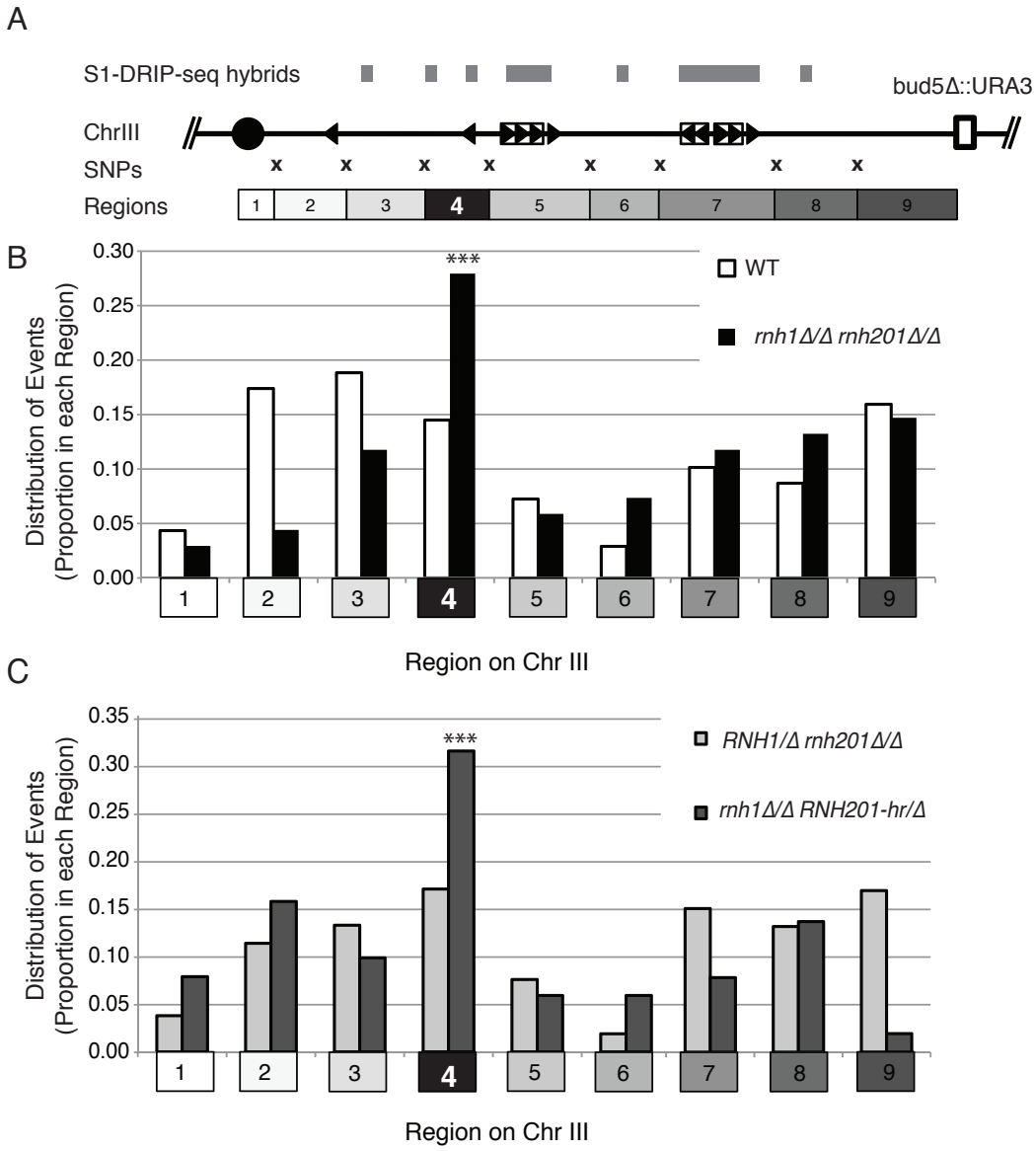
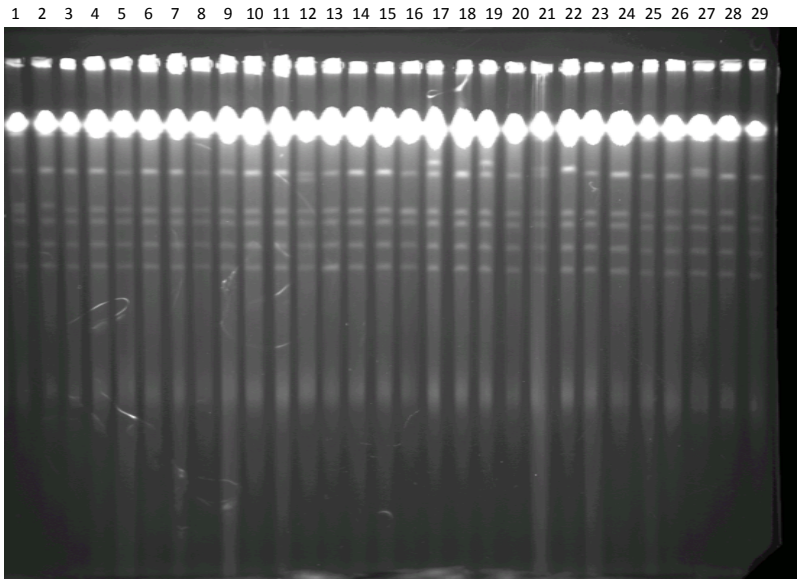
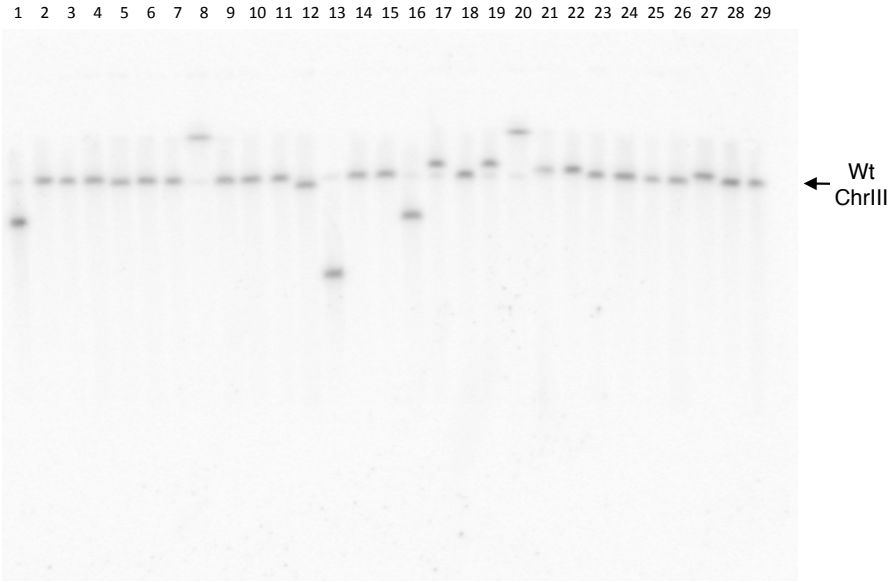


Figure 2.3
Distribution of LOH junctions

(A) Diagram of region of the right arm of chromosome III assayed for LOH junctions. The top row shows the locations of hybrid-prone sequences mapped in Wahba 2016. The second line shows chr III with the centromere diagramed as a circle, Ty elements are diagrammed as boxed triangles, solo delta elements as triangles, and the location of the *URA3* marker inserted at the *BUD5* locus as a square. The third row shows the locations of the SNPs (marked as "X") assayed for heterozygosity. The fourth row shows the 9 regions in which LOH junctions may occur.
(B and C) Locations of LOH boundaries. Boundaries in wild-type (n=69) and RNase H double mutant cells (n=68) (B) and RNase H2 (n=53) and H1 (n=51) deficient cells (C) were mapped. The proportion of LOH boundaries occurring in each of the nine regions is plotted. *** indicates $p<0.001$ using χ^2 test.



EtBr stained Pulsed Field Gel



Southern (LEU2 probe)

Figure 2.4
Sizes of Chr III in colonies that have undergone terminal LOH
The top panel shows an ethidium bromide stained pulsed field electrophoresis gel. The bottom panel shows a Southern blot of that gel with a probe against LEU2, which probes for the S288c derived Chr III homolog. Chromosomes are shown for wt (lanes 1-2, 27-28), *rnh1*Δ/Δ *rnh201*Δ/Δ (lanes 3-12, 25-26), *rnh1*Δ/Δ *RNH201*/Δ (lanes 13-17, 21-24), and *RNH1*/Δ *rnh201*Δ/Δ (lanes 17-20). The parental size of Chr III in a non-rearranged cell is shown in the rightmost lane 29.

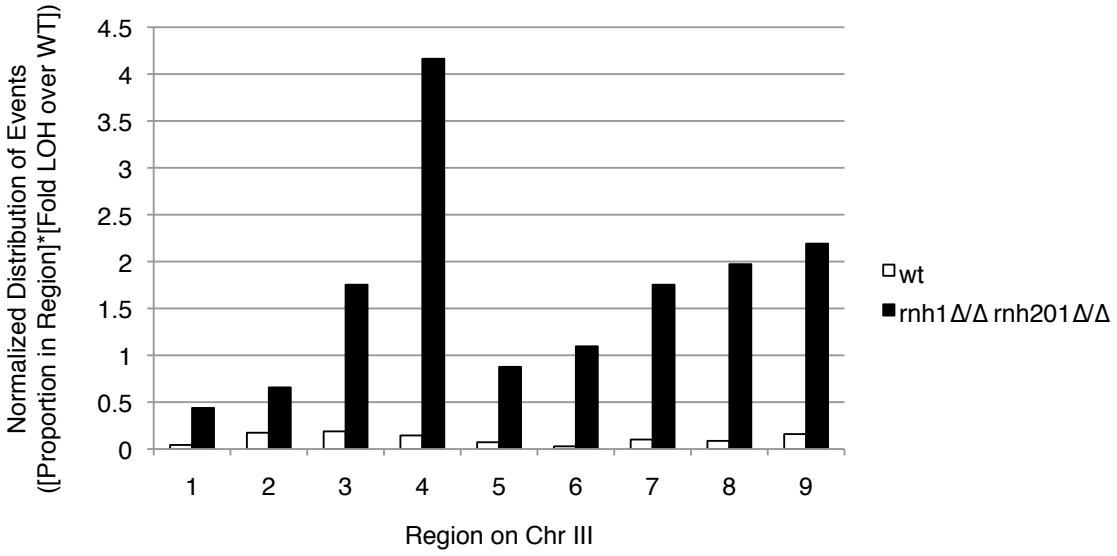


Figure 2.5
Normalized Distribution of LOH boundaries
The distribution of LOH boundary events in each SNP-defined interval was normalized to the fold rate of LOH in RNase H deficient cells over wild-type cells.

LOH events, we asked three related questions. Did the pattern of LOH in this region suggest that multiple hybrids were capable of inducing LOH events? Did the pattern of LOH suggest that some hybrids were more likely to induce LOH events? Did the two RNases H play different roles in protecting against hybrid induced LOH?

In order to map the LOH events on the right arm of chromosome III, we constructed diploids deriving one set of parental chromosomes from the haploid S288c strain used to map hybrids and the second set of parental chromosomes from the haploid RM11. The RM11 strain differed from the S288c by about 46,000 SNPs (Qi et al. 2009). We first identified progeny that had undergone terminal LOH on the right arm of chromosome III by their growth phenotype (Ura- Leu+; Figure 2.1B). We then mapped the LOH in these progeny by monitoring the heterozygosity of the SNPs along the chromosome arm at approximately 10kb intervals (Figure 2.3A). The junctions between the regions of retention of heterozygosity and the loss of heterozygosity marked the sites of resolution of damage induced by RNase H deficiency, either sites of crossover, initiation of BIR, or de novo telomere addition. Indeed, when we examined the size of the chromosomes from colonies with terminal LOH events, many had a wild-type karyotype consistent with LOH by a mechanism of homologous recombination, but some had a chromosome III of a smaller or larger size, indicating the occurrence of de novo telomere addition and other complex chromosome rearrangements (Figure 2.4).

The junctions of spontaneous LOH events in the wild-type strain were uniformly distributed along the 98kb chromosomal segment, with no region showing any more or

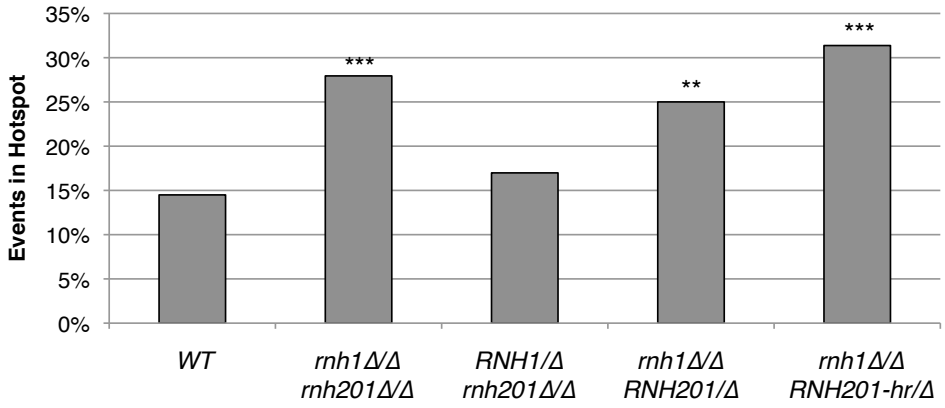


Figure 2.6
LOH boundary events in region 4 hotspot
Percent of LOH boundary events in the region 4 hotspot in wild-type and RNase H mutants. RNase H1 deficient cells with either wild-type *RNH201* or *RNH201-hr* display an instability hotspot. *** indicates $p<0.001$ and ** indicates $p<0.01$ using χ^2 test.

---

less events than would be predicted by the interval length (χ^2 test) (Figure 2.3B, white bars). In contrast, the junctions of LOH events mapped in the RNase H deficient strain (*rnh1*Δ/Δ *rh201*Δ/Δ) were not uniformly distributed. There was an over representation of junctions in region 4, at a higher proportion than would be expected by the length of the interval ($p<0.001$, χ^2 test) (Figure 2.3B, black bars). However, this hotspot represented only 25% of the total LOH events, indicating that most junctions of LOH occurred in other intervals in *rnh1*Δ *rnh201*Δ cells. The overall rate of LOH in the RNase H deficient cells was 14.9 fold higher than wild-type cells, so LOH events were elevated in all of the intervals, and even more so in interval 4. The distribution of events normalized to the overall rate of LOH is shown in Figure 2.5.

To understand the contributions of the individual loss of Rnh1 and Rnh201 to the pattern of LOH on chromosome III, we examined the pattern of LOH in *rnh201*Δ and *rnh1*Δ single mutants. Unlike the RNase H double mutants, strains deficient in RNase H2 (*RNH1/*Δ *rnh201*Δ/Δ) showed a more uniform distribution of events, similar to wild-type (Figure 2.3C, light gray bars). Given that the rate of LOH in these cells was 13 fold higher than wild-type levels, this result indicates that loss of Rnh201 led to the induction of damage in most if not all regions on chromosome III. Since hybrid-forming regions were dispersed along the chromosome, these results are consistent with multiple hybrid prone regions inducing the damage that led to LOH, and that RNase H2 suppressed this damage by removing these hybrids.

In contrast, cells lacking *RNH1*, but expressing *RNH201-hr* (*rnh1*Δ/Δ *RNH201-hr/*Δ) had a hotspot of LOH junctions mapping to region 4, similar to the RNase H double

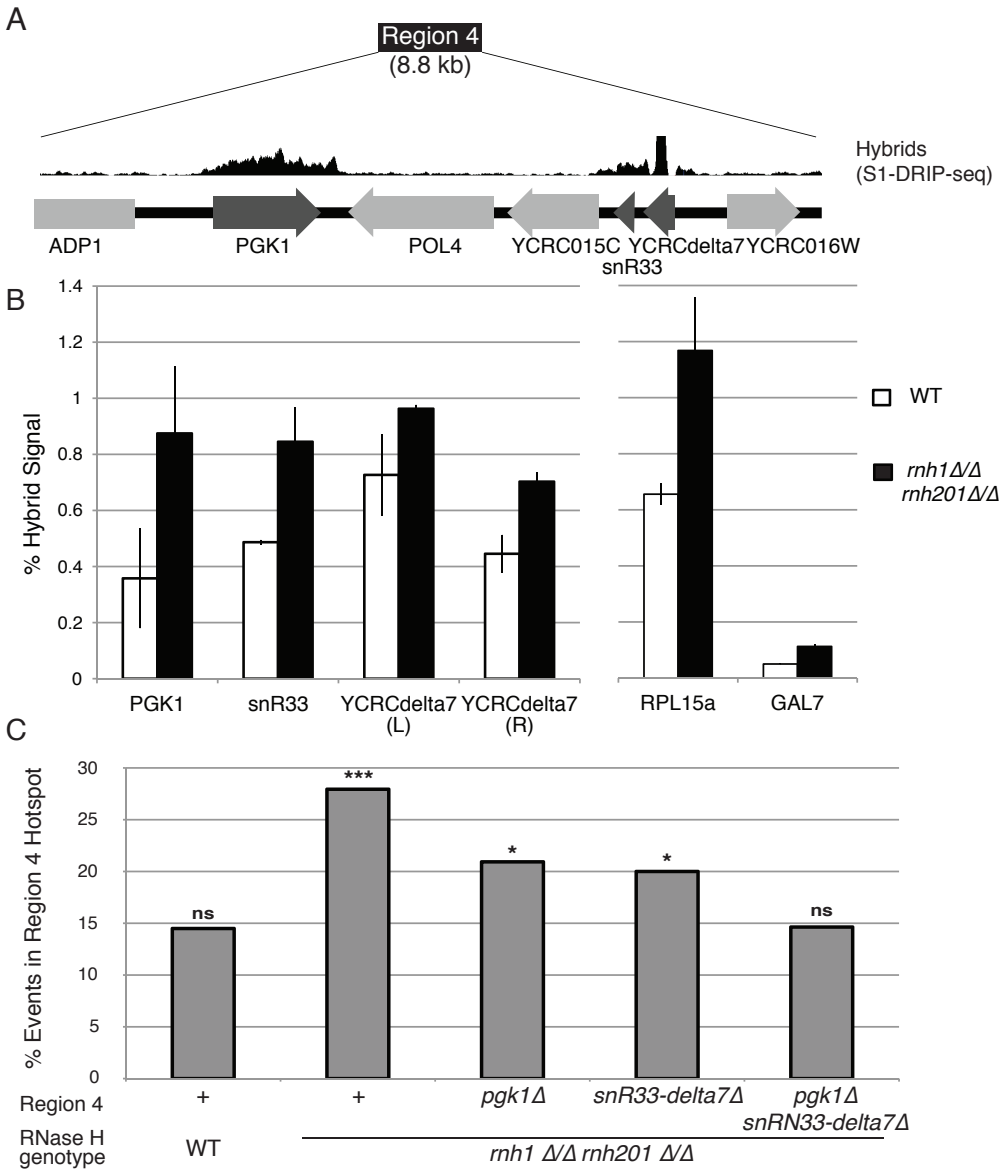


Figure 2.7
Characterization of hybrid-prone sequences in Region 4 hotspot

(A) Diagram of region 4 hotspot. Hybrids by S1-DRIP-seq reads (Wahba et al. 2016) are shown above sequence features from the Saccharomyces Genome Database.
(B) Hybrid signal by DRIP-qPCR. Hybrid signal (as a percentage of input) at hybrid-prone sequences in the hotspot are shown (PGK1, snR33, unique sequence just left of YCRCdelta7 and unique sequence just right of YCRCdelta7) well as another known hybrid-prone sequence (RPL15a) and non-hybrid-prone sequence (GAL7). Error bars represent +/- 1 standard deviation.
(C) Percent of LOH boundary events in the region 4 hotspot in wild-type and RNase H double mutants with deletions of hybrid-forming sequences. *** indicates $p<0.001$, * indicates $p<0.01$, ns indicates not significant at $p<0.1$ using χ^2 test.

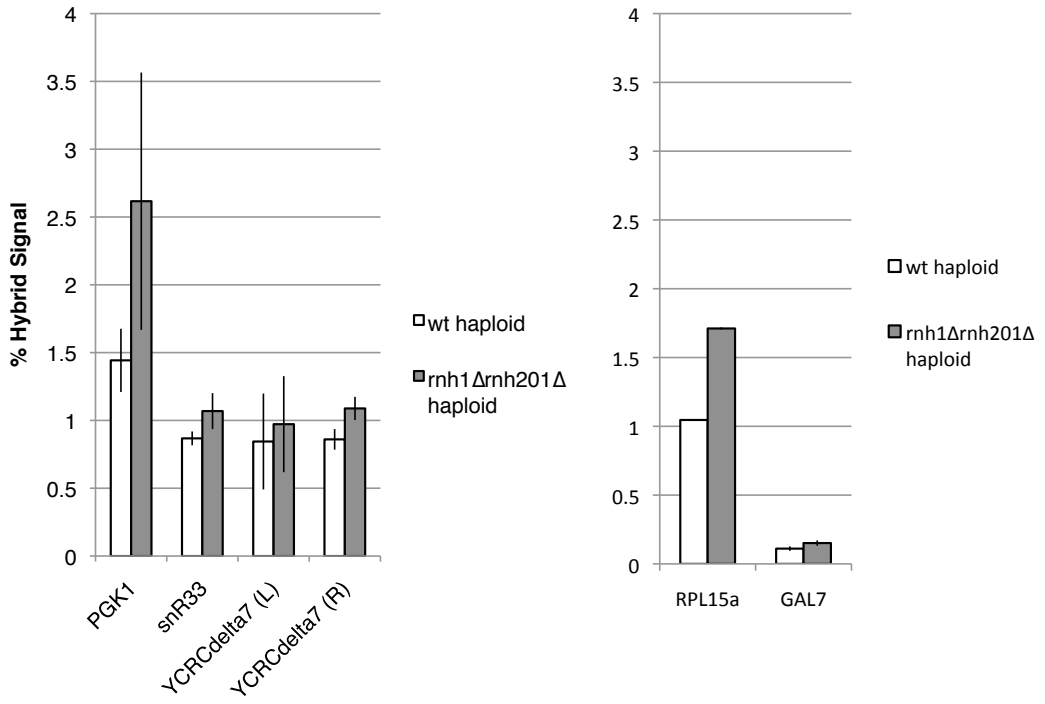


Figure 2.8
Hybrid signal in haploids
Hybrid signal by DRIP-qPCR in S288c haploids. Hybrid signal at hybrid-prone sequences in the hotspot are shown as well as another known hybrid-prone sequence (RPL15a) and non-hybrid-prone sequence (GAL7). Error bars represent +/- 1 standard deviation.

---

mutant (Figure 2.3C, dark gray bars). Cells lacking *RNH1* but expressing a wild-type copy of *RNH201* (*rnh1*Δ/Δ *RNH201*/Δ) also had a hotspot in region 4 (Figure 2.6). Therefore, the hotspot of junctions in region 4 was dependent on RNase H1 but not H2. These results demonstrate a specific role of RNase H1 in targeting chromosome instability at a particular region of the chromosome.

Two hybrid-forming regions contribute to localized instability

Region 4 contains multiple hybrid-prone regions, suggesting that the RNase H1 dependent chromosome instability in this region may be due to hybrid formation. Two clusters of hybrid formation were identified in region 4 in haploid wild-type and RNase H deficient cells by S1-DRIP-seq (Wahba et al. 2016): *PGK1*, and a region containing *snR33* and *YCRCdelta7* (Figure 2.7A). These hybrid-prone loci can be categorized into known hybrid-prone families as a highly transcribed gene, a snoRNA, and a repetitive

solo delta element, respectively (Wahba et al. 2016).To confirm that these hybrid-prone regions form hybrids in the diploid cells used in this study, we performed DRIP followed by qPCR at the loci of interest (Figure 2.7B). We found that all three regions formed hybrids in the diploid strains and, like all the hybrid regions identified by our S1-DRIP-seq study, these regions were hybrid forming in both wild-type and *rnh1*Δ/Δ *rnh201*Δ/Δ strains. Additionally, hybrids at these sites, like most hybrid regions genome-wide, formed at higher levels in RNase H deficient cells than in wild-type. We verified that these loci also form hybrids on the genetically marked chromosome III homolog by performing DRIP on the haploid S288c parents of the diploids assayed in this study (Figure 2.8).

We next generated deletions within the hotspot interval to determine whether the hybrid-prone sequences were necessary for the instability at this hotspot. After deleting the hybrid prone sequences from the homolog bearing the genetic markers, we then assayed the location of the LOH junctions in an *rnh1*Δ/Δ *rnh201*Δ/Δ strain. Deletion of either *PGK1* or a region encompassing *snR33* and *YCRCdelta7,* replacing these loci with a similarly sized *HIS3* marker, led to a significant depression in LOH events mapping to the hotspot (Figure 2.7C). While there was a smaller proportion of LOH events at the hotspot upon deletion of these sequences, there were still more events occurring at the hotspot than would be expected based on the size of the interval ($p<0.05$, $\chi 2$ test). Deletion of both *PGK1* (with *HIS3*) and the region encompassing *snR33* and *YCRCdelta7* (with *TRP1*) led to the elimination of the LOH hotspot. In these strains, the proportion of events occurring in the hotspot was the same as in the wild-type strain, at the level expected based on the interval length. In total, these experiments showed that the hotspot for LOH events resulted from the contribution of multiple hybrid forming sequences within the region. The proximity of these causative hybrid prone regions to the LOH events in the hotspot interval suggests that hybrid-induced damage and its repair occur proximal to the causative hybrids.

The RNases H localize to hybrid-forming regions

Our LOH assay showed that RNase H2 plays the major protective role against chromosome instability. The mapping experiments of LOH events in RNase H2 deficient cells suggested that RNase H2 protected against hybrid induced damage at multiple distinct intervals. To test whether this broad mode of action was reflected at the level of localization, we performed chromatin immunoprecipitation (ChIP) studies of Rnh201. Rnh201 was enriched at the *PGK1* locus, but was not significantly enriched at the other hotspot hybrids (Figure 2.9A and C). It was only weakly enriched at most other tested hybrid-prone regions, with stronger enrichment detected at *LSR1* and *RPL15a* (Figure 2.9B). This weak enrichment over background could be due shortcomings of the ChIP procedures. Alternatively, it might reflect an intrinsic property of Rnh201. RNase H2 removes single ribonucleotides, which are presumably randomly incorporated in the genome, which could lead to a more diffuse localization of Rnh201. Another possibility is that RNase H2 is only transiently associated with most of the hybrid regions at which it acts. Regardless, we were unable to confidently identify the localization of RNase H2.

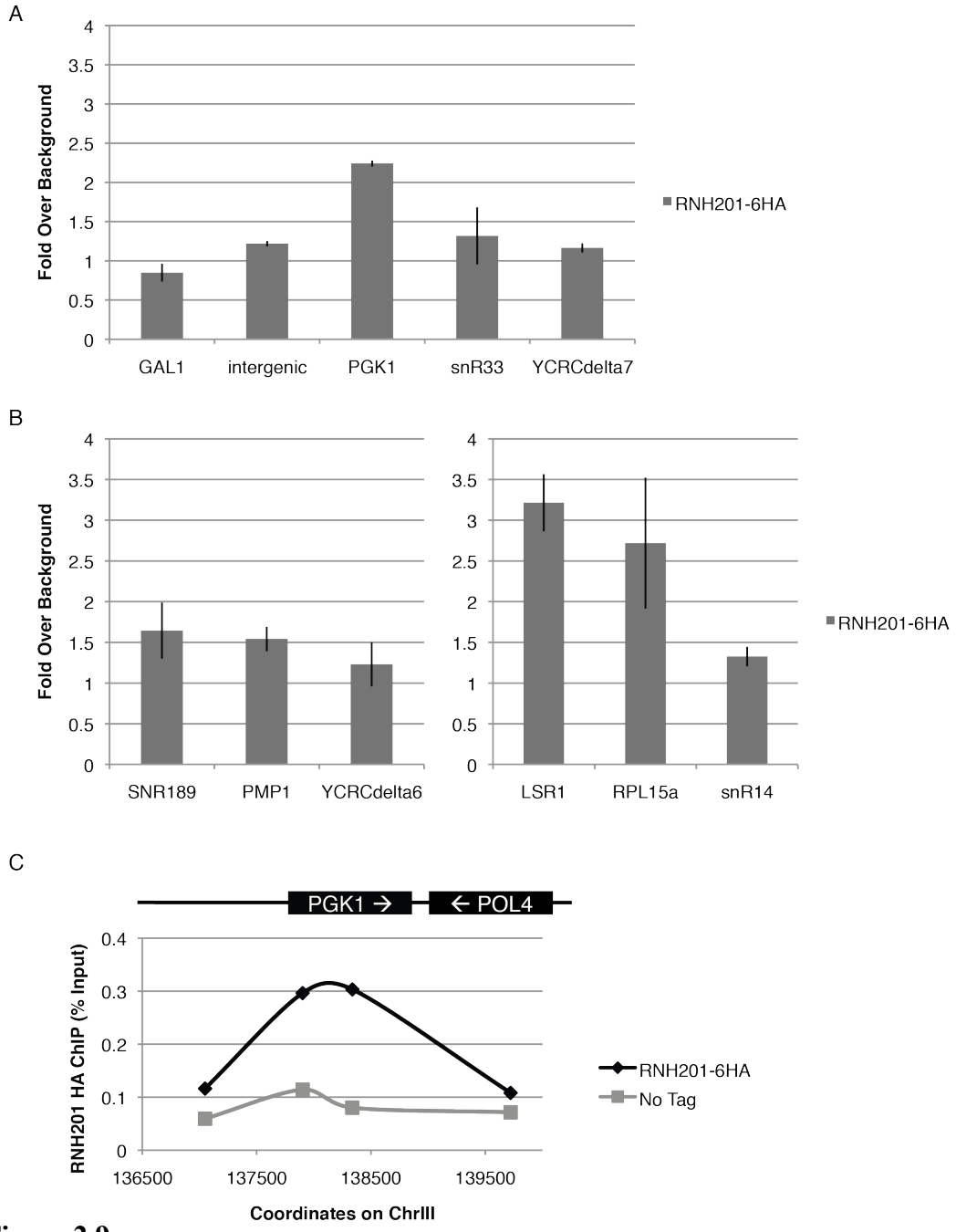


Figure 2.9

Localization of RNase H2

(A and B) Enrichment of Rnh201 with 6xHA tag at the region 4 hotspot (A) and other hybrid-prone loci (B). Fold enrichment over non-hybrid background loci is shown. Two of these background loci (GAL7 and an intergenic sequence upstream of RGS2) are shown. Error bars represent +/- 1 standard deviation.

(C) Percent enrichment of HA in HA-tagged Rnh201 and untagged cells at primers in the hybrid-prone PGK1 locus and up and downstream, non-hybrid-prone loci.

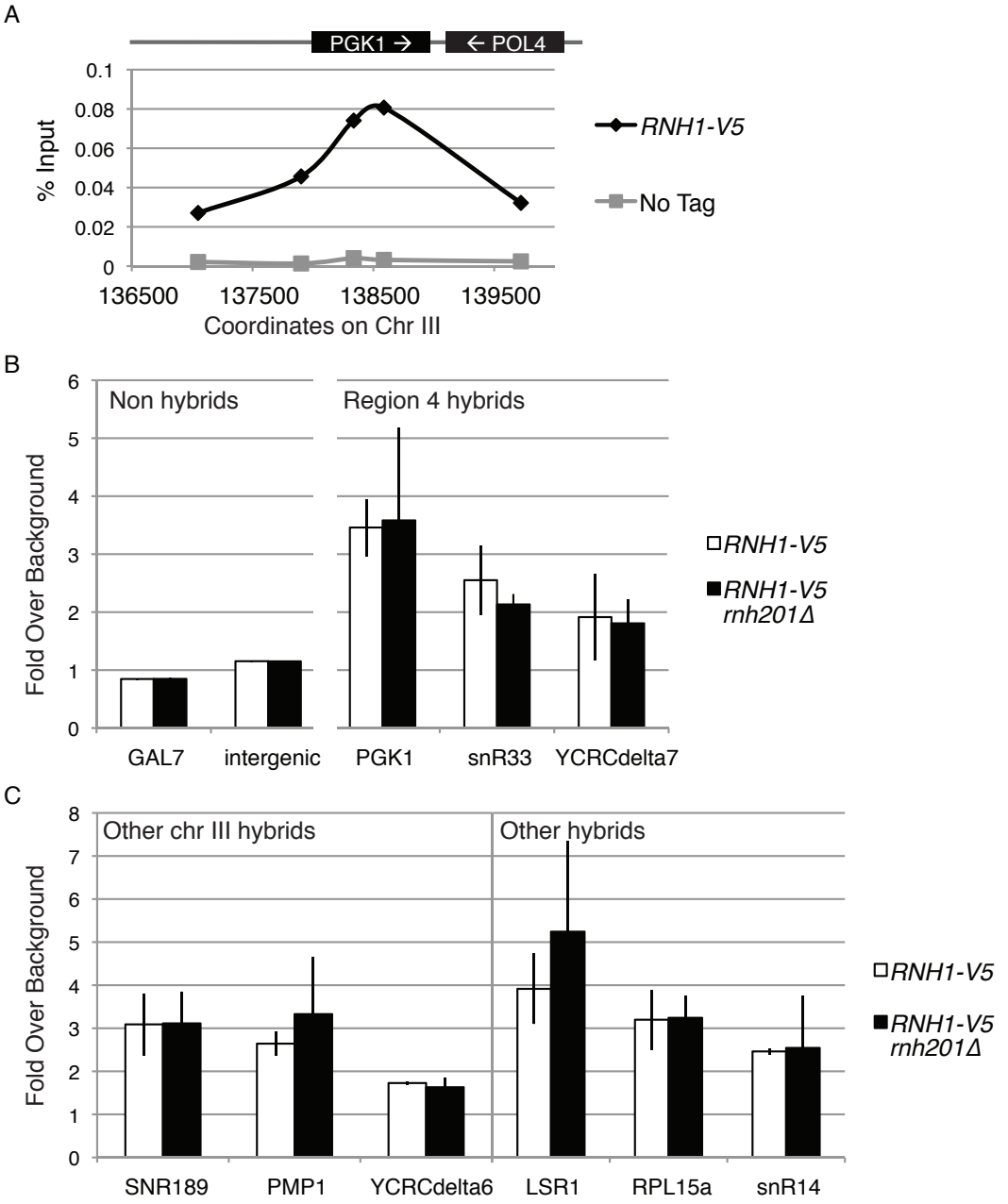


Figure 2.10
RNase H1 localization by ChIP
Percent enrichment of V5 in V5-tagged Rnh1 and untagged cells at primers in the hybrid-prone PGK1 locus and up and downstream, non-hybrid-prone loci.
(B and C) Enrichment of Rnh1 at the region 4 hotspot (B) and other hybrid-prone loci (C) in wild-type and rnh201Δ cells. Fold enrichment over non-hybrid background loci is shown. Two of these background loci (GAL7 and an intergenic sequence upstream of RGS2) are shown in (B). Error bars represent +/- 1 standard deviation.

The LOH SNP mapping assay revealed that RNase H1 protected against LOH within a specific interval while having a limited role for protection in adjacent intervals. We asked whether this interval-specific function of RNase H1 reflected its preferred localization to this region. To address this question we performed ChIP studies of Rnh1 localization using an internal 3xV5 tag after P85 (RNH1-V5). The location of the tag was chosen in a less evolutionarily conserved region of the protein to minimally disrupt protein function (Materials and Methods).

We were able to detect Rnh1 localization specifically to hybrid forming regions. As shown in Figure 2.10A, Rnh1 localized to the hybrid-forming open reading frame of *PGK1* and not to upstream or downstream non-hybrid-forming sequences. Our ChIP studies had very little non-specific signal, with a strain containing no V5 tag showing very low signal (Figure 2.10A, grey line). We found that Rnh1 localized to all the hybrid-forming loci in the instability hotspot. Rnh1 ChIP signal was enriched 3.5 fold over background at the *PGK1* locus, 2.6 fold at snR33 and 1.9 fold at *YCRCdelta7* (Figure 2.10B). Rnh1 also localized to other hybrid-forming loci along the right arm of chromosome III: 3.1 fold at *SNR189*, 2.6 fold at *PMP1* and 1.7 fold at *YCRCdelta6* (Figure 2.10C). In fact, Rnh1 localized to all tested hybrid-forming loci. Other strong hybrid-forming loci *LSR1*, *RPL15a* and *snR14* all showed enrichment of Rnh1 (Figure 2.10C). We additionally performed the ChIP using Rnh1 tagged with a 3xHA tag. This ChIP showed a similar pattern of localization as the 3xV5 tagged protein (Figure 2.11 A and B).

As previously discussed, RNase H1 and RNase H2 have overlapping enzymatic functions. To determine if RNase H2 has an effect on RNase H1 localization, we deleted *RNH201* and performed ChIP of Rnh1. We found that Rnh1 localization was similar at all tested loci in the wild-type and *rnh201*Δ background (Figure 2.10 B and C), indicating that Rnh201 does not affect Rnh1 localization.

A previous study in human cells used a catalytically dead RNase H1 to immunoprecipitate chromatin (Ginno et al. 2012). To determine whether a catalytically dead RNase H1 has the same localization pattern as the wild-type enzyme, we performed ChIP of the catalytically dead Rnh1-D193N in both a wild-type and *rnh201*Δ background. The enrichment of the catalytically dead Rnh1 in both backgrounds was similar to the wild-type Rnh1 (Figure 2.12 A and B), indicating that the catalytically dead enzyme does not have altered localization or level of binding to hybrids. This similarity suggests that RNase H1 binding to hybrids is independent from its enzymatic activity to degrade hybrids. These results coupled with the global localization of RNase H1 to many hybrids indicates that RNase H1 binds broadly to hybrids, but does not rapidly remove or disassociate from them.

Discussion

Why organisms from bacteria to humans harbor two distinct RNases H, both with DNA:RNA hybrid removal activity, has been an enigma. In this study, we investigated the *in vivo* roles of these two RNases H in budding yeast through the lens of R-loop induced chromosome instability. We found that cells lacking both RNases H exhibit a 15 fold greater rate of loss of heterozygosity (LOH) on the right arm of chromosome III. These events could occur through chromosome loss, chromosome rearrangement and

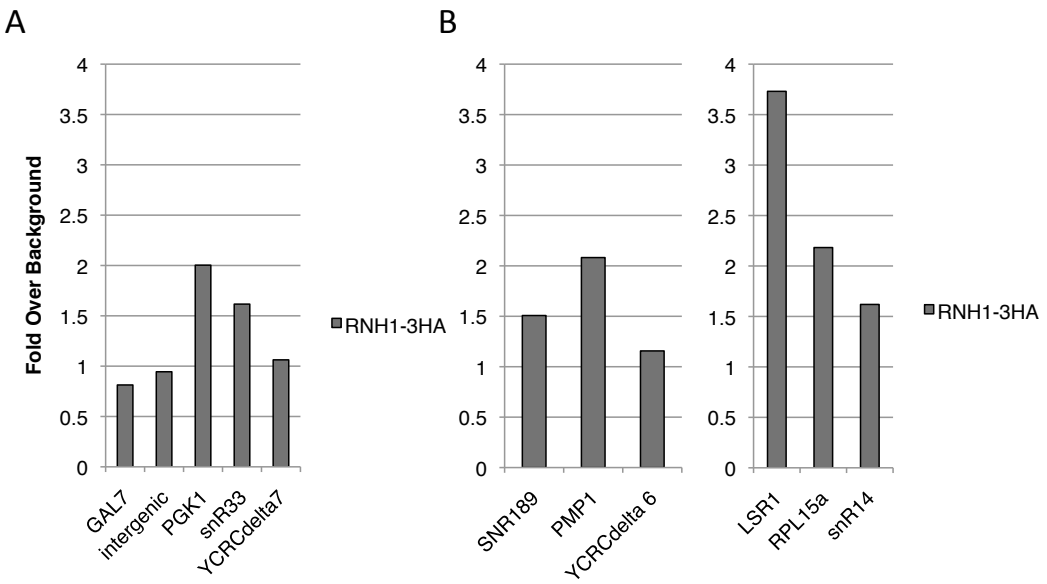


Figure 2.11
Localization of RNase H1 with HA tag
(A and B) Enrichment of Rnh1 with 3xHA at the region 4 hotspot (A) and other hybrid-prone loci (B). Fold enrichment over non-hybrid background loci is shown. Two of these background loci (GAL7 and an intergenic sequence upstream of RGS2) are shown.

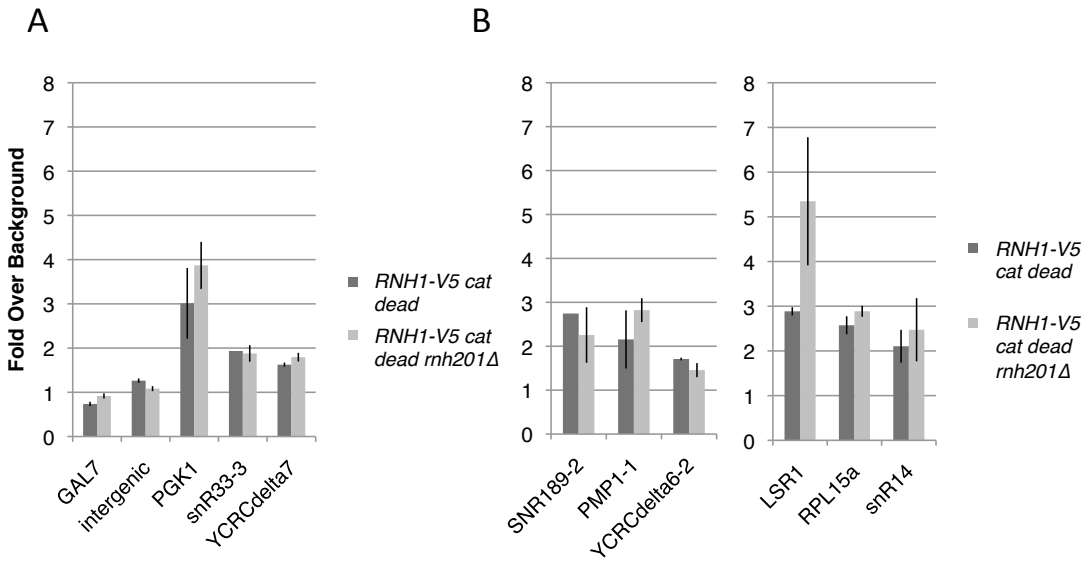


Figure 2.12
Localization of RNase H1 catalytically dead mutant
(A and B) Enrichment of Rnh1 catalytically dead mutant at the region 4 hotspot (A) and other hybrid-prone loci (B) in wild-type and *rnh201*Δ cells. Fold enrichment over non-hybrid background loci is shown. Two of these background loci (GAL7 and an intergenic sequence upstream of RGS2) are shown. Error bars represent +/- 1 standard deviation.

---

repair events that result in long tracts of LOH like mitotic recombination or break induced replication. More than 90% of these large-scale LOH events could be suppressed by restoring RNase H2, but not RNase H1 activity. Furthermore, the LOH could be suppressed by introducing a mutant of RNase H2 that retained its R-loop degradation activity but lacked its single ribonucleotide removal activity. Taken together, these results suggest that most spontaneous R-loops are inhibited from inducing large-scale LOH because they are removed by the hybrid degradation activity of RNase H2 before they can induce DNA damage.

This conclusion that RNase H2 but not H1 carries the major load of protecting cells against large scale LOH corroborates a previous study (O'Connell et al. 2015). While our work and that of O'Connell and colleagues, concluded that this protection resulted from RNase H2 removing R-loops, a different study of RNase H2's activities by Conover and colleagues (Conover et al. 2015) suggested that this protection was afforded by RNase H2's ability to remove misincorporated nucleotides. In fact, many studies have shown that misincorporated ribonucleotides can lead to DNA damage and mutation in a topoisomerase I dependent manner (Nick McElhinny et al. 2010; Clark et al. 2011; Kim et al. 2011; Williams et al. 2013). However, it has remained unclear if misincorporated ribonucleotides are a major contributor to larger scale chromosome instability. The different conclusions from the recent large-scale LOH studies (Conover et al. 2015; O'Connell et al. 2015) both relied on results from mutants of DNA polymerases that incorporated greater or fewer ribonucleotides into DNA. By using an orthogonal approach with an RNase H2 variant defective in removal of misincorporated nucleotides (*RNH201-hr*), we provide strong evidence supporting the Conover study that RNase H2 protects against large-scale chromosome instability mostly by removing R-loops. The *RNH201-hr* mutant greatly suppresses chromosome instability, however, we do note that it does not suppress all the way to wild-type levels. This indicates that removal of misincorporated ribonucleotides by RNase H2 does play a role, albeit a minor one, in preventing chromosome instability.

Further insight into the distinct functions of these two enzymes came when we mapped the position of the junctions between the regions of heterozygosity and loss of heterozygosity along a 98 kb region of chromosome III. These junctions presumably map the repair sites of lesions induced by specific R-loops in the intervals. Our map suggests that RNase H2 and H1 have distinct spatial specificity. A comparison of RNase H2 deficient and wild-type cells revealed no significant difference in the distribution of the LOH junctions. Given the 13 fold induction of overall LOH in the RNase H2 defective mutants, the LOH events induced by RNase H2 deficiency were equally distributed amongst all intervals. Therefore, these results suggest that RNase H2 acts as the major protector against LOH by degrading DNA:RNA hybrids in R-loops at many if not all sites in the genome. This spatially unconstrained function of RNase H2 in hybrid removal fits with its genome-wide activity in removing random ribonucleotide misincorporation. Consistent with this ubiquitous genomic function for RNase H2 in both single ribonucleotide and hybrid removal, it exhibited very weak binding to many sites on chromosomes as assayed by ChIP (this study). One might have imagined that this enzyme would be preoccupied with global removal of misincorporated nucleotides, thereby providing rationale for the need of a second enzyme dedicated to R-loop removal like RNase H1. However, the genome-wide load of single ribonucleotide misincorporation in

wild-type cells seems to be insufficient to generate this competition. Such a preoccupation may occur in a stress condition that increases misincorporation, thereby explaining the synergistic increase in LOH in DNA polymerase mutants lacking RNase H2.

In contrast, the mapping of the junctions of LOH events in RNase H1 deficient cells showed elevated LOH in only the fourth of the 9 contiguous intervals. This restricted spatial impact on LOH provides an explanation of why RNase H1 deficient cells did not exhibit an increase in total LOH of this chromosome III arm (the sum of all 9 intervals). Intriguingly, the ability of RNase H1 to protect primarily the fourth interval but not the other 8 intervals from hybrid induced LOH did not reflect its preferred access to the fourth interval. This conclusion was based on our RNase H1 ChIP, which showed that it localized equally well to hybrid prone regions within the fourth interval, representative hybrid prone regions outside this interval on chromosome III, and elsewhere in the genome. This equal distribution suggests a model in which RNase H1 uses its hybrid recognition activity to bind to spontaneous R-loops, but its nuclease activity is normally suppressed except at only a subset like those in the fourth interval.

Three additional observations are consistent with this hypothesis. The ChIP signal was the same at R-loops from wild-type and catalytically dead RNase H1. If RNase H1 were active on most hybrids where it was bound, one might expect it to degrade the hybrids and interact transiently, whereas the catalytically dead RNase H1 being unable to degrade the hybrids, would have a prolonged interaction and generate a higher ChIP signal (Britton et al. 2014; Gelbart et al. 2005).However, this was not the case. Second, constitutive overexpression of RNase H1 can suppress the instability of a yeast artificial chromosome in RNase H2 deficient cells (Wahba et al. 2011). Overexpression of RNase H1 may allow it to escape repression or regulation, perhaps by titrating out a repressor, and therefore to degrade spontaneous R-loops normally degraded by RNase H2. Third, biochemical characterization of RNase H activity of yeast extracts found that almost all RNase H activity was derived from RNase H2, not RNase H1, again reflecting a possible repression of H1 activity in yeast (Arudchandran et al. 2000). Why is the RNase H1 nuclease repressed at most sites of spontaneous hybrid formation? A clue may come from the fact that inactivation of RNase H1 causes a synergistic increase in hybrid induced LOH when transcription is perturbed by defects in RNA biogenesis machinery (Wahba et al. 2011). RNase H1 may be a stress-induced factor that is unleashed at sites that accumulate hybrids resulting from aberrant transcription in few loci under normal conditions but at many loci when the cells are stressed. Another possibility is that RNase H1 only resolves hybrids during distinct cell cycle stages. Previous studies have suggested that expression of RNase H2, but not RNase H1, is cell cycle regulated with two bursts of expression in S and G2 (Arudchandran et al. 2000). Intriguingly, in that same study RNase H activity appeared to cycle in *rnh201*Δ cell extracts, perhaps indicating some post-transcriptional regulation of RNase H1 activity. Given our results, additional characterization of the expression, protein levels, activity, and binding of each of the RNases H through the cell cycle would be very interesting.

Finally, our study provides important insights into the relationship of specific R-loops with R-loop induced LOH. Previous studies identified LOH events induced by RNase H deficiency but they lacked a high-resolution map of hybrids to correlate these events with specific hybrids. Our recent map of ~800 hybrid prone regions (Wahba et al.

2016) allowed us to assess various models of how R-loops lead to LOH. One possibility is that all R-loops have the potential to induce LOH because they induce damage independent of their context. Alternatively, R-loops may differ in their ability to induce damage because of unique features like their position in genes or their nucleotide content. We showed that RNase H2 deficient cells exhibit elevated LOH at multiple intervals. This broad effect suggests that, if hybrids are allowed to persist, many if not all have the potential to generate damage that leads to LOH. However, the fact that we observe a hotspot for LOH in one interval upon RNase H1 deficiency suggests that some hybrids may be more prone to lead to damage than others. Three hybrid-prone regions lie within the fourth interval, one on the *PGK1* gene and two clustered on *snR33* and *YCRCdelta7*. We showed that the elevated LOH in this interval was partially reduced by deleting individually either *PGK1* or the cluster, and eliminated completely by deleting them both. Engineered hybrids have been shown to cause chromosome instability (Wahba et al. 2013), but our results link specific natural hybrids with LOH for the first time in yeast. The fact that the elevated LOH in this interval appears to result from the sum of the events induced by each hybrid suggests that a feature in this chromosome interval makes hybrids more prone to LOH.

In summary, the experiments presented in this study provide important new examples of functional differences for the hybrid removal activities of RNase H1 and H2. Understanding the molecular basis for these differences may provide important new insights into why these two enzymes have been so highly conserved in evolution.

Materials and Methods

Yeast Strains, Media and Reagents

Full genotypes of strains used in this study are listed in Table S1. All strains are derived from the S288c or RM11, as noted in the table. Genetic markers on chromosome III and gene knock-outs were introduced by standard yeast transformation. Yeast strains were grown by in YEP or synthetic complete media supplemented with 2% glucose at 30°C according to standard yeast protocols. 5-FOA was purchased from Zymo Research.

Loss of Heterozygosity Assay

Diploid cells were dilution streaked on SC-URA plates grown at 30°C. Single colonies were resuspended in 0.25 mLs of water, diluted, and plated onto 5-FOA containing plates. 10^7 cells were plated for wild-type and *rnh1*Δ mutants and 10^6 cells were plated for *rnh201*Δ and RNase H double mutants. Plating efficiency was monitored by plating 200 cells onto YPD plates. Plates were incubated at 30°C for 2-3 days after which the number of colonies forming on each plate was counted. The number of colonies that grow on 5-FOA, normalizing for plating efficiency, is a measure of the rates of events. To determine the proportion of terminal LOH versus chromosomes loss, the colonies grown on 5-FOA plates were replica plated onto SC-LEU. The number of colonies that grew on SC-LEU divided by the number of colonies that grew on 5-FOA represented the percent terminal LOH.

Mapping terminal LOH events on Chromosome III by SNPs

Diploid cells with terminal LOH events on chromosome III were isolated on SC-LEU according to the Loss of Heterozygosity assay described above. Each colony to be mapped was derived from an independent starting colony on SC-URA. SNP containing regions (SNP locations listed in Table S2) were amplified by optimized yeast colony PCR as follows. ~5 µl of yeast colony was resuspended in 20 µl of 0.02M NaOH and boiled for 10 minutes. 1 µl of this boiled yeast colony was used as a template for PCR amplification in the following reaction conditions: 1x Q buffer (Qiagen), 1x standard Taq buffer (NEB), 0.2M dNTPs, 1uM each primer, 0.5U Taq polymerase (NEB). Touchdown PCR cycles were used: 94° for 1 min; 9 cycles of 94° for 20 sec, 62° decreasing 1° each cycle for 45 sec, 68° for 45 sec; 24 cycles of 94° for 20 sec, 52° for 45 sec, 68° for 45 sec. The PCR amplified DNA was cleaned-up using standard enzymatic clean-up conditions (1x Cutsmart buffer (NEB), 1U rSAP (NEB), 1.8U ExoI (Thermo Scientific)) and incubated at 37° for 1 hour followed by heat inactivation at 80° for 15 minutes. This reaction was diluted and the equivalent of $1/20^{th}$ of the reaction was used for standard Sanger sequencing. Heterozygosity or homozygosity of the SNP was determined by relative intensity of each base on sequencing chromatographs. For the deletion studies of hybrid-forming sequences at the hotspot (Figure 2.7), only the two SNPs flanking the hotspot were sequenced.

Pulse Field Gel Electrophoresis (PFGE) and Southern Blot Analysis

Yeast genomic DNA was prepared in 1% pulse-field grade agarose plugs (SeaPlaque 50100) and resolved as previously described (Schwartz & Cantor 1984) with a Bio-Rad CHEF-DR III system. The following parameters were used: 1.3% agarose gel in 0.5x TBE, 6 V/cm, 120° angle, 15–25 s switch times, 24 hr at 14°C. For Southern analysis, gels were transferred onto a GeneScreen Plus membrane (PerkinElmer NEF988) and probed with a 1.8 kb fragment containing LEU2 sequence.

DNA:RNA Immunoprecipitation (DRIP)

DRIP experiments were performed as described previously (Wahba et al. 2016). Briefly, genomic DNA was isolated from ~10^{10} log phase cells. 100ug of genomic DNA was digested overnight at 37°C with SpeI-HF, HindIII-HF, BsrGI, and XbaI. Digested DNA was immunoprecipitated with S9.6 antibody. The percentage hybrid signal was quantified using qPCR on DNA from immunoprecipitation and total input with the DyNAmo HS SYBR Green qPCR kit (Thermo Scientific).

Construction of tagged Rnh1 and Rnh201 strains

The *RNH201-hr* allele was amplified off the ycNPH2-FL2 plasmid harboring a C-terminally FLAG tagged *RNH201* gene with the P45D-Y219A mutations from (Chon et al. 2013) and integrated into the endogenous *RNH201* locus with a *HIS3* marker by standard PCR and recombination in yeast. Rnh1 was internally tagged with 3xV5 or 3xHA tag. The tag was placed after Proline 85. The location of the tag was chosen in a

region of low conservation within the *Saccharomyces* clade. The catalytically dead Rnh1 harbors the point mutation D193N, which was chosen as the evolutionarily conserved residue from the human RNase H1 protein of D145N (Ginno et al. 2012).

Chromatin Immunoprecipitation

Chromatin immunopreciptation was performed as described previously (Eng et al. 2015). Briefly, a 50 mL YPD culture of asynchronous cells was grown to OD 0.6. Cells were fixed for 2 hours in a final concentration of 1% formaldehyde. After mechanical cell lysis with glass beads, chromatin was sheared 24 times for 45s each (settings at duty cycle, 20%; intensity, 10; cycles/burst, 200; 30s of rest between cycles) using a Covaris S2. Immunoprecipitation of epitope-tagged proteins was isolated using 5ul of anit-V5 (R960-25; Invitrogen) or 1 ul of anti-HA (12CA5; Roche Life Sciences) monoclonal antibodies. A no-primary-antibody control and a no-tag strain was also performed to ensure specificity. Appropriate dilutions of input and immunoprecipitated DNA samples were used for quantitative PCR analysis to ensure linearity of PCR signal. A background level of no enrichment was established by averaging six sets of background primers, chosen for having very low hybrid signal in S1-DRIP-seq (Wahba et al. 2016): GAL7, RGS2 intergenic, GAL1, UCB6, CYC1, and DYN2. Fold over this average background was calculated for all loci tested. All experiments were done at least twice, and an average of the fold over background is presented. ChIP primers listed in Table S3.

References

Arudchandran, A. et al., 2000. The absence of ribonuclease H1 or H2 alters the sensitivity of Saccharomyces cerevisiae to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair. *Genes to cells : devoted to molecular & cellular mechanisms*, 5(10), pp.789–802.

Britton, S. et al., 2014. DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. *Nucleic acids research*, 42(14), pp.9047–9062.

Cerritelli, S.M. & Crouch, R.J., 2009. Ribonuclease H: the enzymes in eukaryotes. *The FEBS journal*, 276(6), pp.1494–1505.

Chan, Y.A. et al., 2014. Genome-Wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with DRIP-Chip M. Snyder, ed. *PLoS Genetics*, 10(4), p.e1004288.

Chon, H. et al., 2013. RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic acids research*, 41(5), pp.3130–3143.

Clark, A.B. et al., 2011. Mismatch repair-independent tandem repeat sequence instability resulting from ribonucleotide incorporation by DNA polymerase ε. *DNA repair*, 10(5), pp.476–482.

Conover, H.N. et al., 2015. Stimulation of Chromosomal Rearrangements by

Ribonucleotides. *Genetics*, 201(3), pp.951–961.

Eng, T., Guacci, V. & Koshland, D., 2015. Interallelic complementation provides functional evidence for cohesin-cohesin interactions on DNA. *Molecular biology of the cell*, 26(23), pp.4224–4235.

Gelbart, M.E. et al., 2005. Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. *Genes & development*, 19(8), pp.942–954.

Ginno, P.A. et al., 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular Cell*, 45(6), pp.814–825.

Huertas, P. & Aguilera, A., 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular Cell*, 12(3), pp.711–721.

Kim, N. et al., 2011. Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science*, 332(6037), pp.1561–1564.

Nick McElhinny, S.A. et al., 2010. Genome instability due to ribonucleotide incorporation into DNA. *Nature chemical biology*, 6(10), pp.774–781.

O'Connell, K., Jinks-Robertson, S. & Petes, T.D., 2015. Elevated Genome-Wide Instability in Yeast Mutants Lacking RNase H Activity. *Genetics*, 201(3), pp.963–975.

Qi, J. et al., 2009. Characterization of meiotic crossovers and gene conversion by whole-genome sequencing in Saccharomyces cerevisiae. *BMC genomics*, 10(1), p.475.

Santos-Pereira, J.M. & Aguilera, A., 2015. R loops: new modulators of genome dynamics and function. *Nature Reviews Genetics*, 16(10), pp.583–597.

Schwartz, D.C. & Cantor, C.R., 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, 37(1), pp.67–75.

Stirling, P.C. et al., 2012. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes & development*, 26(2), pp.163–175.

Wahba, L. et al., 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Molecular Cell*, 44(6), pp.978–988.

Wahba, L. et al., 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes & development*, 30(11), pp.1327–1338.

Wahba, L., Gore, S.K. & Koshland, D., 2013. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome

instability. *eLife*, 2, p.e00505.

Williams, J.S. et al., 2013. Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA. *Molecular Cell*, 49(5), pp.1010–1015.

Table 2.1 Strains Used In Chapter 2

Description	Strain	Strain Background	Genotype
LOH Assay Diploids			
WT	AZ88	S288c/RM11	*MATa/α LEU2/leu2Δ his3/HIS3 ura3/ura3-Δ0 TRP1/TRP1 lys2/LYS2 HO::KanMX BUD5/bud5Δ::URA3*
rnh1Δ/Δ rnh201Δ/Δ	AZ92	S288c/RM11	*MATa/α rnh1Δ::HYG/rnh1Δ::HYG rnh201Δ::NAT/rnh201Δ::NAT LEU2/leu2Δ his3/HIS3 ura3/ura3-Δ0 TRP1/TRP1 lys2/LYS2 HO::KanMX BUD5/bud5Δ::URA3*
RNH1+/Δ rnh201Δ/Δ	AZ174	S288c/RM11	*MATa/α RNH1:HIS3/rnh1Δ::HYG rnh201Δ::NAT/rnh201Δ::NAT LEU2/leu2Δ his3/HIS3 ura3/ura3-Δ0 TRP1/TRP1 lys2/LYS2 HO::KanMX BUD5/bud5Δ::URA3*
rnh1Δ/Δ RNH201+/Δ	AZ 203a	S288c/RM11	*MATa/α rnh1Δ::HYG/rnh1Δ::HYG RNH201:HIS3/rnh201Δ::NAT LEU2/leu2Δ his3/HIS3 ura3/ura3-Δ0 TRP1/TRP1 lys2/LYS2 HO::KanMX BUD5/bud5Δ::URA3*
rnh1Δ/Δ RNH201 hr/Δ	AZ204a	S288c/RM11	*MATa/α rnh1Δ::HYG/rnh1Δ::HYG RNH201-P45D-Y219A:HIS3/rnh201Δ::NAT LEU2/leu2Δ his3/HIS3 ura3/ura3-Δ0 TRP1/TRP1 lys2/LYS2 HO::KanMX BUD5/bud5Δ::URA3*
Hotspot Hybrid Deletion Diploids			
rnh1Δ/Δ rnh201Δ/Δ pgk1Δ	AZ207a	S288c/RM11	*MATa/α PGK1/pgk1Δ::HIS3 rnh1Δ::HYG/rnh1Δ::HYG rnh201Δ::NAT/rnh201Δ::NAT LEU2/leu2Δ his3/HIS3 ura3/ura3-Δ0 TRP1/TRP1 lys2/LYS2 HO::KanMX BUD5/bud5Δ::URA3*
rnh1Δ/Δ rnh201Δ/Δ snRNA/δΔ	AZ217a	S288c/RM11	*MATa/α SNR33-YCRCdelta7/snr33-ycrcdelta7Δ::HIS3 rnh1Δ::HYG/rnh1Δ::HYG rnh201Δ::NAT/rnh201Δ::NAT LEU2/leu2Δ his3/HIS3 ura3/ura3-Δ0 TRP1/TRP1 lys2/LYS2 HO::KanMX BUD5/bud5Δ::URA3*
rnh1Δ/Δ rnh201Δ/Δ pgk1Δ snRNA/δΔ	AZ221a	S288c/RM11	*MATa/α PGK1/pgk1Δ::TRP1 SNR33-YCRCdelta7/snr33-ycrcdelta7Δ::HIS3 rnh1Δ::HYG/rnh1Δ::HYG rnh201Δ::NAT/rnh201Δ::NAT LEU2/leu2Δ his3/HIS3 ura3/ura3-Δ0 TRP1/trp1Δ::KanMX lys2/LYS2 HO::KanMX BUD5/bud5Δ::URA3*
Haploids for DRIP			
WT	AZ61c	S288c	*MATa LEU2 bud5Δ::URA3 his3*
rnh1Δ rnh201Δ	AZ62d	S288c	*MATa rnh1Δ::HYG rnh201Δ::NAT LEU2 bud5Δ::URA3 his3 ade-*
RNase H1 ChIP			
RNH1-V5	JA151	S288c	*Mata RNH1-P85-3xV5 MIF2-3xV5:G418 his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*
RNH1-V5 rnh201Δ	JA183	S288c	*Mata RNH1-P85-3xV5 rnh201Δ::HYG his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*

RNH1-V5 Cat Dead	JA172	S288c	*Mata RNH1-P85-3xV5-D193N MIF2-3xV5:G418 his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*
RNH1-V5 Cat Dead rnh201Δ	JA184	S288c	*Mata RNH1-P85-3xV5-D193N rnh201Δ::HYG his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*
WT, No Tag	LW6836a	S288c	MATa leu2 his3 ura3 TRP1 ade-
RNH1-HA	AZ205c	RM11	MATα RNH1-F95-3HA lys2 ura3-Δ0 HO::KanMX leu2
RNase H2 ChIP			
RNH201-HA	AZ218b	S288c	*MATa RNH201-6HA:HIS3MX LEU2 bud5Δ::URA3*

Table 2.2 Locations of SNPs assayed in Chapter 2

Chromosome	Position in SGD	Allele in S288C	Allele in RM11
III	114,986	G	T
III	125,079	A	G
III	135,857	T	C
III	145,017	C	T
III	156,868	A	G
III	165,025	T	C
III	175,359	T	C
III	185,113	T	C

Table 2.3 Primers used in Chapter 2

Primer Name	Sequence
PGK1_F	TTT CGA CTT GCC ACA ACG TG
PGK1_R	ATC TTG TCA GCA ACC TTG GC
snR33_F	GCA AAT CGA TTG TCC ACA CAC
snR33_R	GCC TAG CTT TTA CAC CGG TTT G
YCRCdelta7(L)_F	CCG TTC CGC CAA ATT TTT CAT G
YCRCdelta7(L)_R	TAG CGC AAG TGG TTT AGT GG
YCRCdelta7(R)_F	AGA ACA TCC TTG AAA GGT CGA C
YCRCdelta7(R)_R	AGC TGA ATA ATG CCG TGG TG
RPL15a_F	ACC GCT GAA GAA AGA GTT GG
RPL15a_R	TGT TGA GGG TCG ACC AAG AT
GAL7_F	CCA ACC AAG AAT TTC CGA AC
GAL7_R	CGC CTC GAT TTT AAA GCA AC
PGK1_5'_F1	GAC TTC AAC TCA AGA CGC ACA G
PGK1_5'_R1	AAA GGA TTC GCG CCC AAA TC
PGK1_5'_F2	TTG CTG CTT TGC CAA CCA TC
PGK1_5'_R2	CAA GTG AGA AGC CAA GAC AAC G
PGK1_F2	TTG ATG GAA AAG GCC AAG GC
PGK1_R2	TTG GCA TCA GCA GAG AAA GC
POL4_F1	CCC AAC AAT CTT CGC TGT ACG
POL4_R1	CGA CCG AGT TGG CAA AAA TC
RGS2_intergenic_F	CGT GTC TGG CTC GGA AGT AT
RGS2_intergenic_R	CCG CAA TAA CGT ACA CAT CG
GAL1_F	GAG CTT TAC TGC CGA CGA AG
GAL1_R	CGG GAA CCA TAT GAT CCA TT
UBC6_F	GAT ACT TGG AAT CCT GGC TGG TCT GTC TC
UBC6_R	AAA GGG TCT TCT GTT TCA TCA CCT GTA TTT GC
CYC1_F	TGA ATT CAA GGC CGG TTC TG
CYC1_R	TTA TGT GGG CCA CCC TTT TC
DYN2_F	ACA TTG CTG GGA CGG TAA AG
DYN2_R	AAT GGC CCT TTT CGT GTG TC
SNR189_F	CGT AAG TAC TCC AAA GCA GTC TC
SNR189_R	ACG GGC CTG ACA TCT CTA TTC
PMP1_F	AAA GGG TAT CGC ACA CAC AC
PMP1_R	CGG AGC GAG CCA TTT TAT TTC C
YCRCdelta6_F	GTG AGG AAT TAT CGG GCA TCT TG
YCRCdelta6_R	GCC ATT TCA TGA GGA CGG AAT AC
LSR1_F	TTT TGG TTT GCA AGG AAA GG
LSR1_R	TGT AGA CCA ACC CCA CCC TA
snR14_F	CCT TAT GCA CGG GAA ATA CG
snR14_R	ATT CAA AAG CGA ACA CCG AAT

Chapter 3: Chromosomal context effects of native hybrid-forming sequences

Abstract

S1-DRIP-seq (S1 nuclease DNA:RNA immunoprecipitation followed by deep sequencing) was used to create a high-resolution map of DNA:RNA hybrids in the yeast genome. This study found that hybrids form at the same loci in wild-type and RNase H deficient cells, and corroborated previous observations in the field that hybrids form at higher levels in RNase H mutants. This study also determined that high transcription is both strongly correlated with and sufficient for hybrid formation. In the experiments presented in this chapter, we investigate if the increased R-loop levels observed in the RNase H mutants are due to changes in transcript levels. We found that the transcript levels at hybrid-forming loci were not increased in RNase H deficient cells. We also investigated whether genes that are highly transcribed and form hybrids at their endogenous locus can form hybrids at an ectopic locus. We found that *HSP150* and *CIS3* did form hybrids when inserted onto a yeast artificial chromosome (YAC). The presence of these hybrid-forming sequences did not cause instability of the YAC, even when overexpressed. These experiments show that hybrid-prone sequences can form hybrids outside their native chromosomal context.

Background

R-loops are a potent source of genome instability, and modulators of some cellular activities. However, when we embarked on this study, there were no genome-wide maps of DNA:RNA hybrids in yeast. As yeast had proven to be a powerful organism in which to study R-loops and their consequences, we sought to identify where R-loops form *in vivo.*

A low resolution map of R-loops in human cells (Ginno et al. 2012) provided a foundation on which we were able to build a technique with good signal preservation and high resolution mapping. Ginno et al. used restriction enzymes to gently fragment genomic DNA before immunoprecipitation with the S9.6 antibody. The use of restriction enzymes was a major reason why their study produced a low-resolution map. In order to gain resolution, we used sonication, or physical acoustic shearing, to randomly shear our DNA. Sonication is a technique common in chromatin immunoprecipitation (ChIP) procedures. However, we found that sonication lead to the loss of ~80% of R-loops (Figure 1B in (Wahba et al. 2016)). Accordingly we sought to prevent the loss of hybrids during shearing. Use of crosslinking agents or digestion by micrococcal nuclease or Fragmentase (New England Biolabs), did not help retain R-loops during fragmentation. However, pre-treating the genomic DNA with S1 nuclease prior to sonication resulted in retention of the majority of DNA:RNA hybrids. S1 nuclease is an endonuclease which degrades single stranded DNA, and presumably it degrades the single strand of DNA excluded in an R-loop. This could prevent that DNA strand from reanneling to its complementary strand and displacing the RNA molecule during sonication.

We termed this improved strategy S1-DRIP-seq and used it to map hybrids in wild-type and RNase H deficient (*rnh1*Δ*rnh201*Δ*)* cells. We identified 781 peaks in *rnh1*Δ*rnh201*Δ, all of which were found in wild-type cells as well. Our map confirmed previously identified hybrid-forming loci: ribosomal DNA, Ty elements, telomeres, and tRNAs. We were also able to identify additional classes of loci prone to hybrid formation: snRNAs, snoRNAs and ncRNAs. The high-resolution nature of our map also allowed us to find sequences prone to hybrid formation. We found that hybrid regions had high AT skew. We also found that long poly A tracts were extremely prone to hybrid formation.

A major factor that influences hybrid formation at open reading frames (ORFs) is transcription level. We found that of the hybrids associated with an ORF, 42% were in the top 10% highest transcribed ORFs. Of the top 5% highest transcribed ORFs, 46% were hybrid-prone. If the stringency of peak-calling was relaxed, then 82% of these highest transcribed ORFs were associated with a hybrid. Beyond these correlative studies, we found that if we took genes that were normally lowly expressed and non-hybrid-forming and induced high levels of transcription, hybrids formed. Taken together this shows that transcription level has a strong influence on hybrid formation.

After identifying some of the features which impact hybrid formation, we wanted to determine if there are additional factors that govern hybrid formation that could not be discovered with our mapping technique. One candidate that we explore here is the chromosomal context in which hybrids form. There have been some reports in the field that suggest that hybrids play a role in global chromosome architecture. In budding yeast, R-loop accumulating mutants have elevated H3S10 phosporylation, a mark associated with chromatin condensation (Castellano-Pozo et al. 2013). In fission yeast, mutants with elevated R-loops were able to rescue cells with a Condensin deficiency (Legros et al. 2014).

Given a potential link between R-loops and chromosome architecture, we wanted to determine if chromosomal context influenced R-loop formation. Additionally, since high transcription was such a potent inducer of hybrid formation, we wanted to see if high transcription of a hybrid-prone sequence would form hybrids in an exogenous context.

Results

Increased hybrid levels in RNase H mutants are not due to increased transcription

In the S1-DRIP-seq study, we found that hybrids formed at the same locations in wild-type and *rnh1*Δ*rnh201*Δ cells, but at higher levels. Additionally, we found that high transcript levels were highly correlated with hybrid-formation. We wanted to determine if the increased hybrid formation in the *rnh1*Δ*rnh201*Δ cells could be due to increased transcription of hybrid-prone loci in this mutant.

We extracted RNA from exponentially growing wild-type and *rnh1*Δ*rnh201*Δ cells. We then quantified the transcript levels of 5 loci. For our analysis of transcription levels in the S1-DRIP-seq study, we binned all genes into 20 equal categories based on

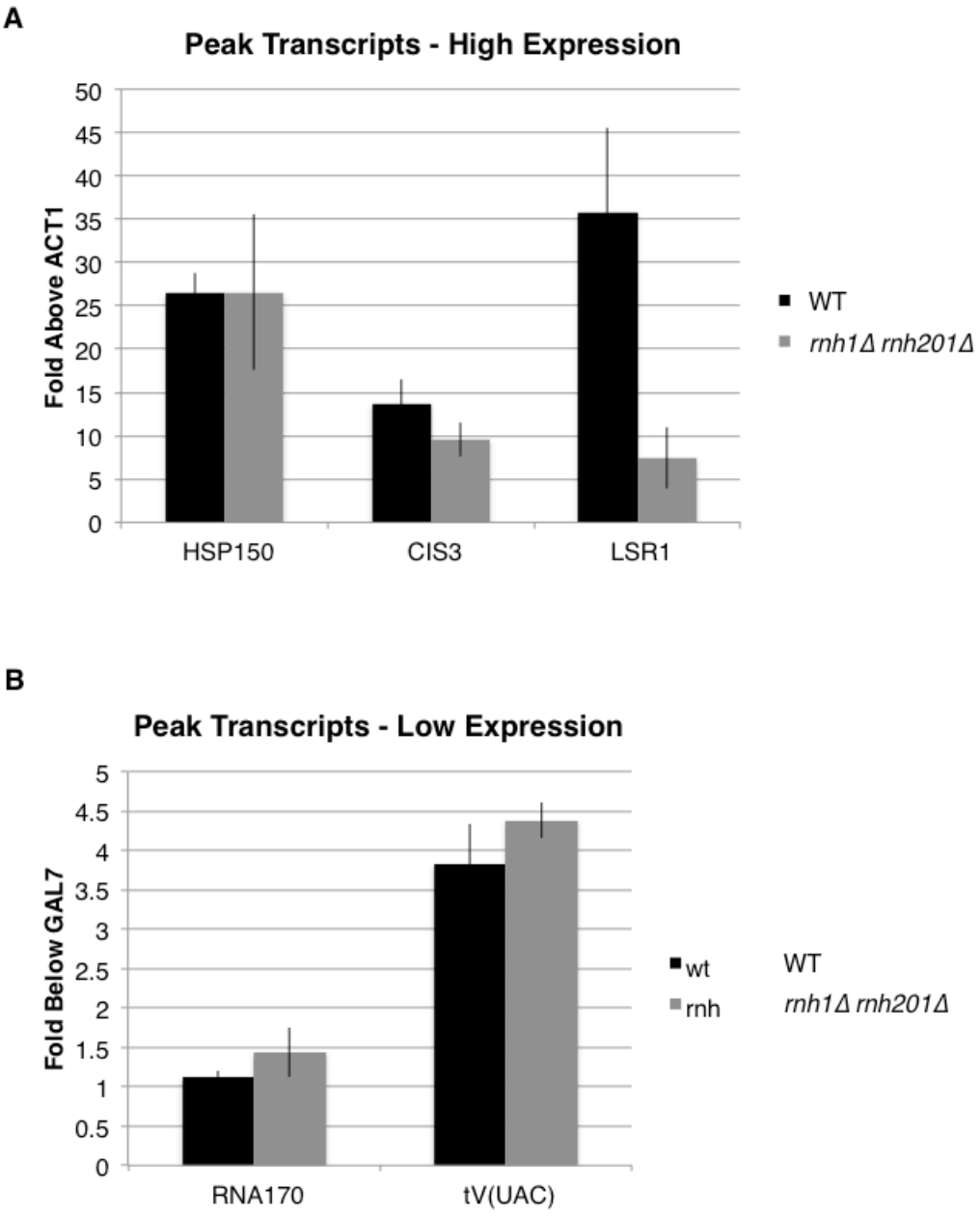


Figure 3.1
Transcript levels of peak-forming loci in wild-type and RNase H deficient cells

RT-PCR quantification of transcript abundance. Transcript levels for high expression genes were normalized to *ACT1* in (A) and lower expression genes were normalized to *GAL7* in (B). Error bars represent one standard deviation of 2 biological replicates.

their transcript abundance measured in an RNA-seq study (Van Dijk et al. 2011). Here, we tested two loci that were classified as highly expressed, in the second highest bin: *HSP150* and *CIS3*. We also tested three non-coding transcripts whose transcript levels were not present in the RNA-seq study: *LSR1*, *RNA170*, and *tV(UAC)*. *HSP150*, *CIS3*, *LSR1* and *tV(UAC)* were all found to be hybrid-prone by S1-DRIP-seq. *RNA170* was a control, non-hybrid-forming locus.

Using RT-PCR we measured the relative abundance of these five transcripts and found none of the five increased in transcript levels from wild-type to *rnh1*Δ*rnh201*Δ (Figure 3.1 A and B). Four of the five tested loci had the same levels of transcripts. *LSR1* had lower levels of transcript in the RNase H deficient cells (Figure 3.1A), even though it has moderately increased levels of hybrids in that mutant (Figure S1.6 in (Wahba et al. 2016)). From these experiments we can conclude that the increased levels of hybrid formation observed in the *rnh1*Δ*rnh201*Δ cannot be attributed to an increase in transcript levels at these loci.

Sequences that form hybrids in their native chromosomal context can form hybrids at an ectopic locus

Our S1-DRIP-seq experiments identified a hybrid-prone region on chromosome X that contained two neighboring, highly expressed hybrid-forming ORFs: *HSP150* and *CIS3* (Figure 3.2A). We wanted to determine which was more important for hybrid formation: sequence or chromosomal context. To explore this we moved parts or the entire *HSP150-CIS3* locus onto a yeast artificial chromosome (YAC) (Figure 3.2B).

We first determined if the *HSP150-CIS3* locus could form hybrids in the chromosomal context of the YAC. After inserting the genes under their native promoters onto the YAC, we performed restriction enzyme DRIP (RE-DRIP). Using restriction enzymes as the method of fragmentation allowed us to determine if the hybrid signal we detect at the *HSP150* or *CIS3* sequence derives from the endogenous chromosomal locus or from the YAC locus. We chose restriction enzymes that maintained an intact piece of DNA with both the hybrid-prone sequence and unique YAC sequence. Using this strategy with two distinct restriction enzyme cocktails, RE-DRIP followed by qPCR showed that the hybrid-prone sequences *HSP150* and *CIS3* could form hybrids at an ectopic locus on the YAC (Figure 3.3 A and B). We measured hybrid signal at the YAC specific primer set number 4 (primer set positions diagrammed in Figure 3.2B). Importantly, strains with a YAC without the *HSP150* and *CIS3* insertion did not have a detectible hybrid signal at primer set 4 or another non-hybrid forming locus on the YAC (Figure 3.4).

Our S1-DRIP-seq analysis revealed that all hybrid-prone regions found in *rnh1*Δ*rnh201*Δ cells were also found in wild-type cells. However, as mentioned above, the hybrids formed at these loci at increased levels in *rnh1*Δ*rnh201*Δ compared to wild-type (Wahba et al. 2016). The RE-DRIP qPCR experiments performed here confirmed this observation. At the *HSP150-CIS3* loci, the level of hybrid formation was greater in RNase H deficient cells (Figure 3.3 A and B).

Similar observations of context-independent hybrid formation were observed for a Ty element expressed on the YAC (Wahba et al. 2016). In these experiments, Lorenzo Costantino also showed that Ty elements form hybrids at their genomic locus by

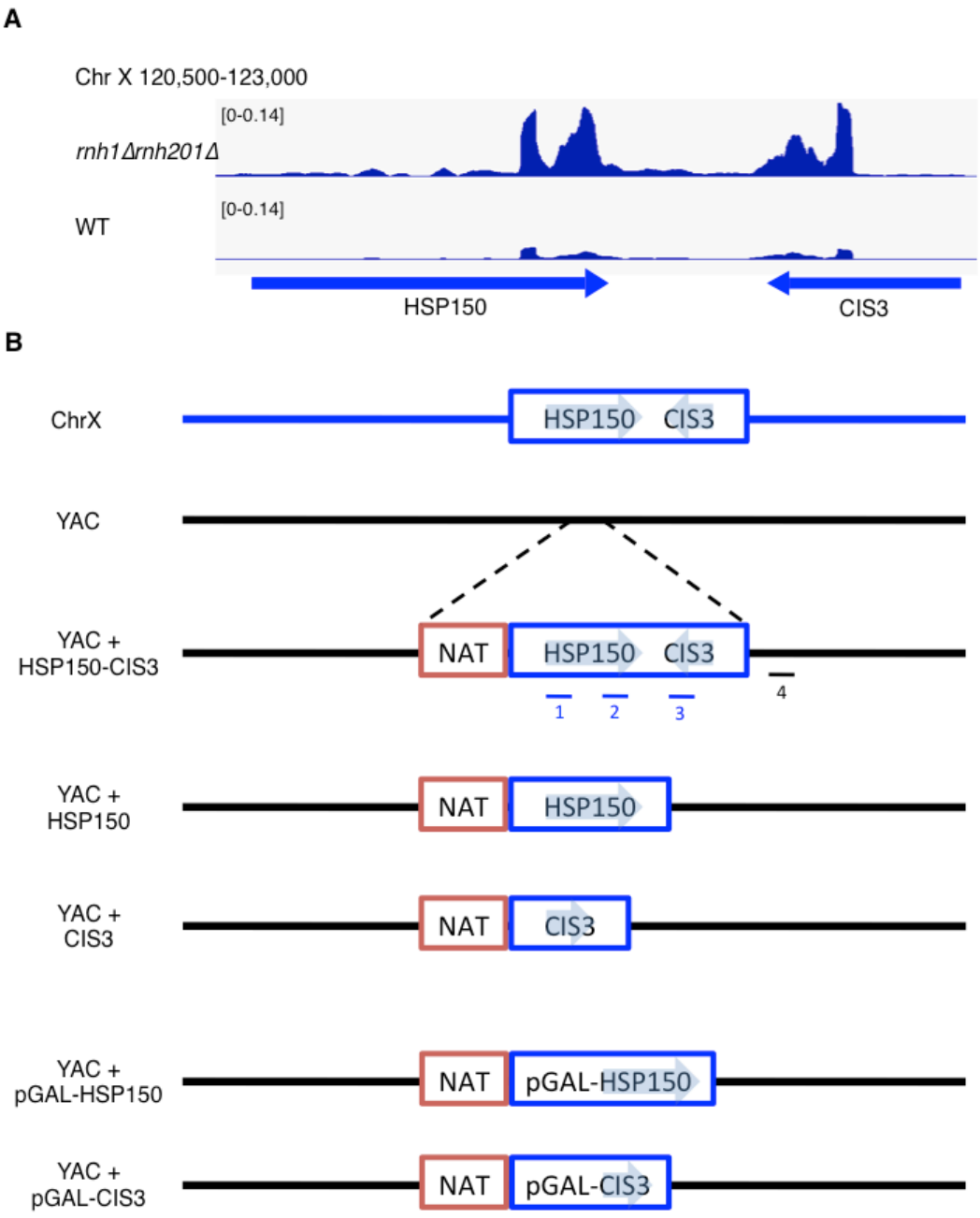


Figure 3.2
Diagram of the hybrid forming loci *HSP150* and *CIS3*
(A) Snapshot of S1-DRIP-seq reads mapping to the *HSP150* and *CIS3* loci in *rnh1*Δ*rnh201*Δ and wild-type.
(B) Schematic representation of integration of the loci onto the YAC. All cells contained an intact copy *HSP150-CIS3* at the endogenous locus (first line, diagrammed in blue). Various sequences were introduced into the YAC, diagrammed in black. The locations of the qPCR primers used in Figure 3.3 are shown.

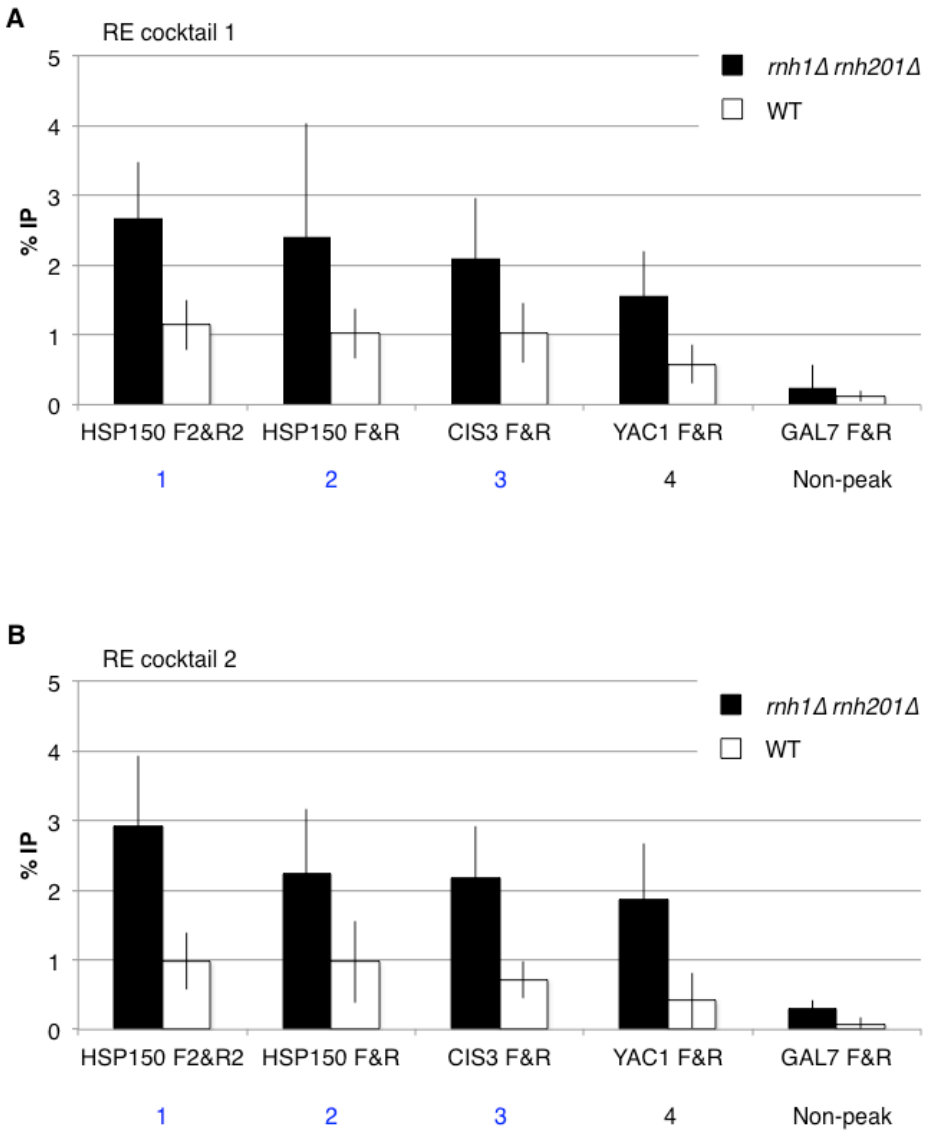


Figure 3.3
Detection of hybrids at *HSP150-CIS3* locus on the YAC

RE-DRIP-qPCR was used to detect hybrids in RNase H deficient and wild-type cells. Locations of primer sets are indicated in Figure 3.2B. Primers 1-3 are common to both the endogenous and YAC loci while primer set 4 is unique to the YAC. Restriction enzymes used for fragmentation in (A) are HindIII, BsrGI, XbaI and SpeI and in (B) are EcoRV, NdeI, SacI and SnaB1. Error bars represent +/- 1 standard deviation.

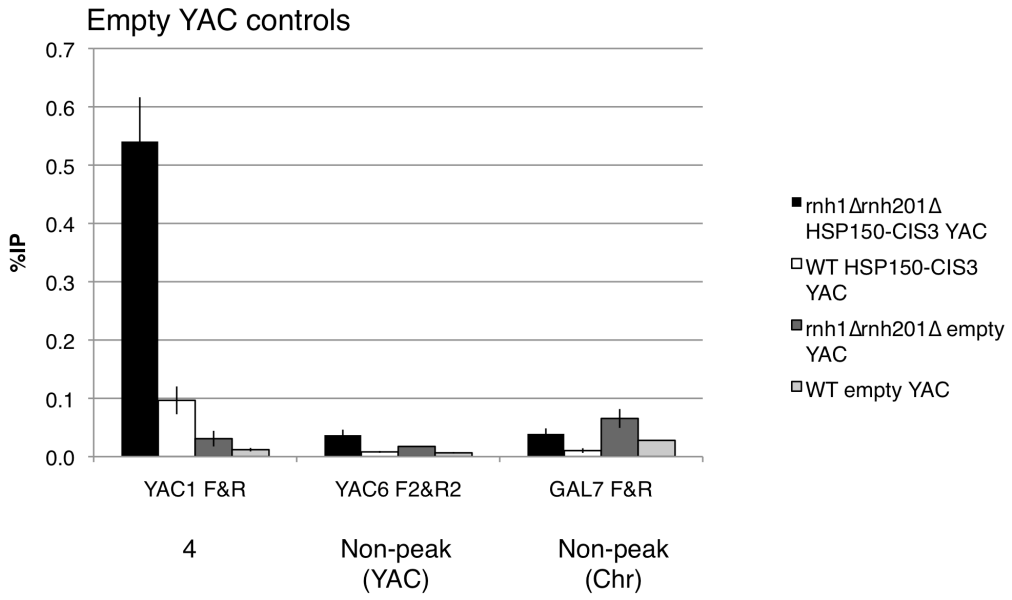


Figure 3.4
YAC DRIP controls

Control loci for DRIP of HSP150-CIS3 insertions on the YAC and YACs without insertion are shown. YAC6 F2&R2 primer set is located on the YAC on a different restriction fragment separate from the hybrid-forming locus. GAL7 F&R is a non-hybrid forming locus on chromosome II.

---

performing DRIP using restriction enzymes to fragment the DNA followed qPCR using primers that amplify unique sequences adjacent to the Ty locus that are on the same restriction fragment as the Ty. This was done at both endogenous Ty loci and for a Ty element placed on the YAC. In both of these experiments hybrids were detected at the genomic Ty locus.

Hybrid-forming sequences do not increase YAC instability

Previous work in the Koshland Lab found that inducing hybrid formation on the YAC could induce its instability (Wahba et al. 2013). The YAC is comprised of non-essential sequence mainly derived from human chromosome VII, as well as yeast derived selectable genetic markers that can be used to monitor its presence in cells. The right arm of the chromosome carries the *URA3* marker, and cells that have lost this marker can be selected for by plating on the drug 5-FOA. Given that the *HSP150* and *CIS3* sequences form hybrids when introduced to the YAC, we wanted to determine their affect on YAC stability.

We measured rates of 5-FOA resistance for YACs bearing *HSP150* and *CIS3* expressed under their endogenous promoters alone and together. We found no difference

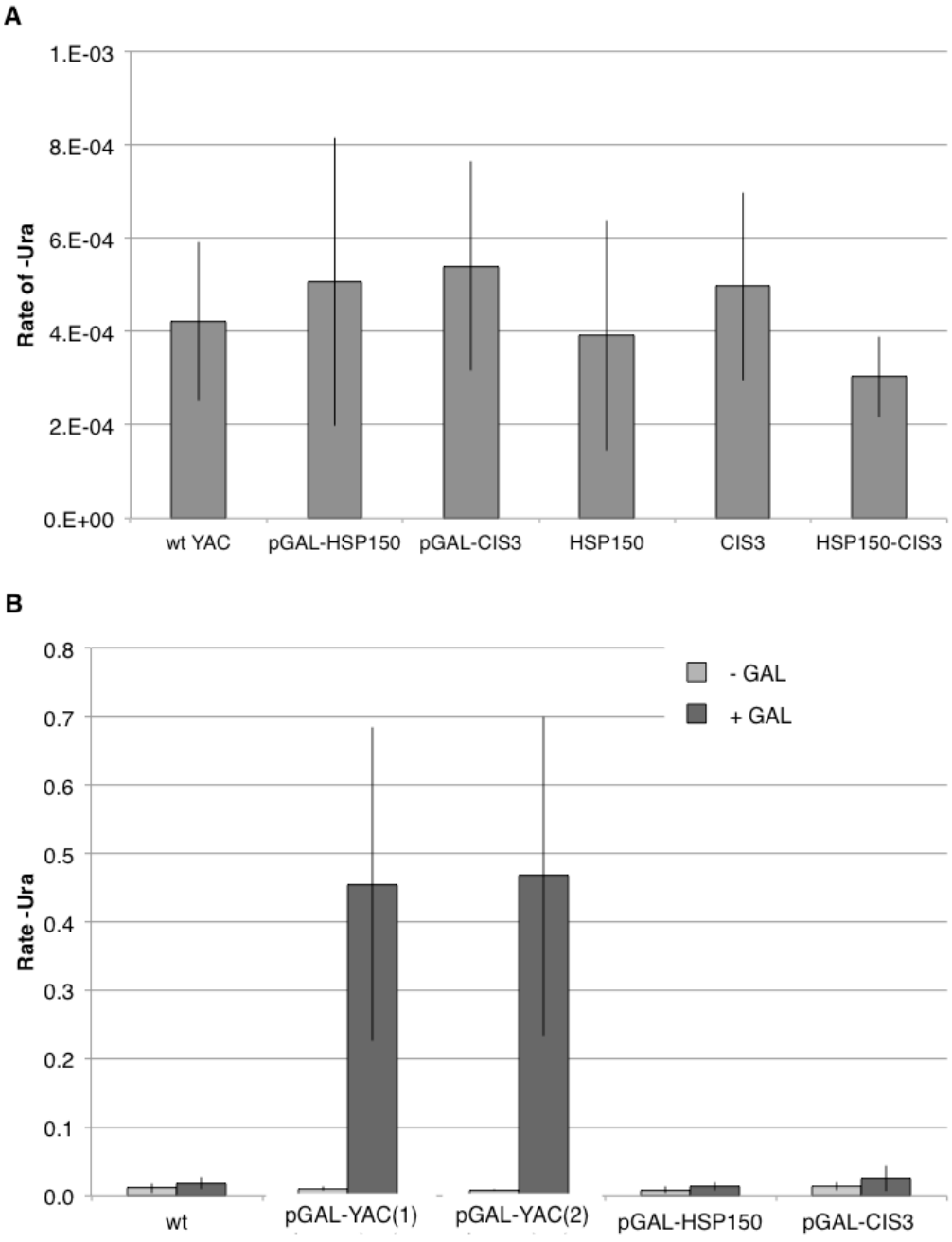


Figure 3.5
Instability of YACs containing *HSP150* and *CIS3*

Rates of YAC instability of YACs containing *HSP150* or *CIS3* under their endogenous promoters (A) or pGAL promoter (B). Error bars represent +/- 1 standard deviation.

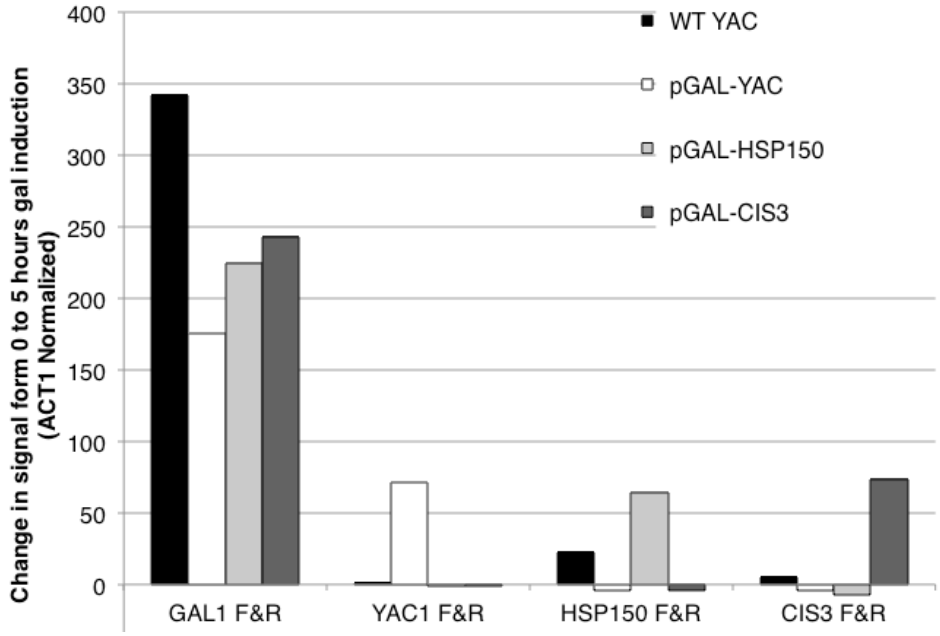


Figure 3.6
RNA abundance of pGAL overexpressed transcripts

Fold induction of pGAL overexpressed transcripts determined by RT-PCR and expressed as a fold change between basal levels at timepoint 0 and 5 hours of galactose induction.

---

in YAC stability when these hybrid-forming sequences were present on the YAC as compared to a wild-type YAC with no sequences inserted (Figure 3.4A). A similar phenomenon was observed when a hybrid-forming Ty element was inserted on the YAC (Costantino and Koshland, unpublished). Together, these experiments show that hybrid-forming sequences do not necessarily increase overall YAC instability.

Overexpression of hybrid-forming sequences does not induce YAC instability

Previously, hybrid-associated YAC instability was observed upon overexpression of human derived YAC sequence (Wahba et al. 2013)(Figure 3.5B). This overexpression was accomplished using a galactose-inducible promoter system (pGAL). For this study, we placed the hybrid-prone *HSP150* and *CIS3* genes under control of the pGAL (Figure 3.2B). We induced transcription of the pGAL driven genes by transferring cells to galactose containing media. After 5 hours of pGAL driven overexpression, no increase in instability was observed for the pGAL-HSP150 or pGAL-CIS3 containing YACs (Figure 3.5B).

To ensure that the genes were indeed overexpressed in galactose, we measured RNA levels in these strains. After 5 hours in galactose, all the strains induced the endogenous GAL1 transcript ~200 fold (Figure 3.6). This strongly induced locus was

used as a positive control locus. Overexpression of all of the pGAL driven transcripts was also observed. In the corresponding strains, the YAC sequence, *HSP150* and *CIS3* were all induced ~60 fold (Figure 3.6).

Discussion

The mapping of hybrids using S1-DRIP-seq gave us a high-resolution map of hybrid formation genome-wide. This view into the *in vivo* formation of hybrids provided insight into factors that influence hybrid formation: high levels of transcription, high AT skew, and long polyA tracts. Armed with this information, we wanted to further explore how other features influenced hybrid formation. The experiments presented in this chapter explored the chromosomal context in which endogenous sequences form hybrids.

First we verified that the increase in hybrid formation we observed in RNase H mutants was not due to an increase in transcription at these loci. Previously it was reported that *rnh201*Δ mutants had altered expression levels of 349 genes, 123 of which were upregulated and 226 were downregulated (Arana et al. 2012). However, the effects observed in this study were modest, mostly 1.5-2 fold changes. Nonetheless, given that we showed that greatly increasing the expression of a gene is sufficient to induce hybrid formation, we wanted to be sure transcript levels did not influence the increased hybrid formation in *rnh1*Δ*rnh201*Δ cells. This was unlikely as we mapped hybrids to 477 ORFs, which is many more ORFs than those whose expression was altered by loss of RNase H2. Indeed, we did not detect any transcripts with increased abundance in the RNase H deficient cells.

Since the increased hybrids in RNase H mutants were not due to increased transcript levels, we propose that mapping hybrids in the RNase H mutants is simply a sensitized way to detect hybrids that form endogenously in wild-type cells. All of the 781 hybrid peaks detected in the *rnh1*Δ*rnh201*Δ were also detected in wild-type. This shows that hybrids do form transiently in at least 8% of the genome but are usually rapidly degraded by the RNases H. Perhaps it is not surprising that deletion of the RNases H does not alter the locations of hybrid formation; the RNases H can be viewed as a surveillance system that globally removes hybrids that are liable to form at specific loci. This is consistent with the RNases H localizing to all hybrid-forming loci by ChIP presented in chapter 2 of this dissertation. Only when these hybrids are allowed to persist in the absence of the RNases H do hybrids induce elevated levels of genomic instability.

We next sought to determine if chromosomal context affects hybrid formation. For this we selected the highly expressed genes *HSP150* and *CIS3*. We placed these genes on the YAC and found that they did form hybrids at this ectopic location. From this we can conclude that simple expression of this DNA sequence is sufficient to form a hybrid, independent of the chromosomal context. The context-independence of hybrids is consistent with the recently reported evolutionary conservation of hybrid-forming sequences between human and mouse (Sanz et al. 2016). Interestingly, in that study they observed that some orthologous genes that did not have conserved R-loops had differences in levels of RNA expression. This provides further evidence that expression level is a key determinant in R-loop formation.

As reviewed extensively in chapter 1, hybrid formation is linked to genomic instability. However, our S1-DRIP-seq study, as well as many other genome-wide

hybrid-mapping studies, has found that R-loops are a prevalent genomic feature in wild-type cells. In our study we found 8-12% of the genome is hybrid prone, depending on the stringency of the parameters used. Additionally, recent work in the field has uncovered biological functions that utilize R-loops (reviewed in chapter 1). The pervasive presence of R-loops might require the cell to prevent hybrid-associated instability at loci that naturally form R-loops. In agreement with that hypothesis, we observed that YACs expressing the hybrid-forming *HSP150* and *CIS3* sequences did not have increased chromosomal instability.

In contrast, a non-endogenous hybrid can be induced by pGAL driven transcription of the human-derived YAC sequence (Wahba et al. 2013). Driving the production of this "unscheduled" R-loop induces the instability of the YAC. However, driving this higher level of transcription of *HSP150* or *CIS3* does not induce instability of the YAC. This further supports the idea that endogenous, scheduled R-loops are somehow prevented from causing instability. The hypothesis is additionally supported by the loss of heterozygosity data presented in chapter 2. The monitored segment of chromosome III contained many loci that form hybrids in wild-type cells. However, in wild-type cells no clustering of instability at hybrid loci was observed, suggesting that they are resolved without deleterious consequences.

Materials and Methods

Yeast stains, media and reagents

Full genotypes of the strains used in this study are listed in Table 3.1. All strains were built with standard yeast methods. All stains containing a YAC had their chromosome sizes confirmed by pulse field gel electrophoresis. 5-Fluoroorotic (5-FOA) was purchased from Zymo Research. The primers used in this study are listed in Table 3.2.

Measuring RNA abundance using RT-PCR

RNA was extracted from ~10^8 log phase cells using the RNeasy Kit (Qiagen) with mechanical lysis. The resultant RNA sample was digested with 1 ul DNase I (New England Biolabs) in 1x DNase I buffer (NEB) for 1 hour at 37°C. RT-PCR was performed One-Step RT-PCR kit (Qiagen) with the addition of SYBR green.

RE-DRIP

RE-DRIP was performed as previously described (Wahba et al. 2016) with restriction enzyme cocktails indicated in the figure legends.

YAC instability assay

For the YAC instability assay without Galactose induction, single colonies were grown on media lacking uracil at 30°C, resuspended in water and 10^6 cells were plated on 5-FOA and 200 cells were plated on YPD to determine colony forming units (CFUs) and

grown for 2-3 days at 30°C. Rate of –Ura presented is the colonies able to grow on 5-FOA normalized to CFUs on YPD.

For YAC instability assays with Galactose induction, single colonies grown on media lacking uracil were used to inoculate 5 mls of SC-URA liquid media and grown overnight (~20 hours) at 30°. These cells were washed once and used to inoculate 25 mLs of YEP media with 2% lactic acid and 3% glycerol to a density of $2.8x10^6$ cells/mL. Cells were grown for ~17 hours at 30°C to complete 2.5-3 doublings. Galactose was then added to 2% final. Samples were taken at 0 and 5 hours to determine YAC instability (as described above) and RNA transcript abundance (as described above).

Table 3.1

Yeast strains used in Chapter 3

Description	Strain	Genotype
Strains from S1-DRIP-seq for RNA abundance of hybrid-prone transcripts		
WT	LW6836a	*MATa leu2 his3 ura3 TRP1 ade-*
rnh1Δrnh201Δ	LW6838a	*MATa leu2 his3 ura3 TRP1 ade- rnh1Δ::HYG-B rnh201Δ::CLONAT*
Insertion of hybrid-prone sequences on the YAC		
WT YAC HSP150	AZ82a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 NAT:HSP150)*
WT YAC CIS3	AZ84a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 NAT:CIS3)*
WT YAC HSP150-CIS3	AZ86a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 NAT:HSP150-CIS3)*
WT YAC pGAL-HSP150	AZ80a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 NAT:pGAL-HSP150)*
WT YAC pGAL-CIS3	AZ81a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 NAT:pGAL-CIS3)*
WT YAC pGAL-YAC	AZ36a&b	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 NAT:pGAL)*
YAC strains for DRIP		
WT YAC HSP150-CIS3	AZ 156b	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 NAT:HSP150-CIS3)*
rnh1Δrnh201Δ YAC HSP150-CIS3	AZ153b	*MATa ura3Δ0 his3Δ met15Δ0 leu2Δ0 rnh1Δ::HYG rnh201Δ::KanMX YAC (URA3 MFA1pr-HIS3 TRP1 MET15 NAT:HSP150-CIS3)*
WT empty YAC	LW6732c	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1)*
rnh1Δrnh201Δ empty YAC	LW5031	*MATa rnh1Δ:G418 rnh201Δ:NAT ura3Δ0 his3Δ met15Δ0 leu2Δ0 YAC (URA3 MFA1pr-HIS3 TRP1 MET15)*

Table 3.2

Primers used in Chapter 3

Primer Name	Sequence
LSR1_F	TTT TGG TTT GCA AGG AAA GG
LSR1_R	TGT AGA CCA ACC CCA CCC TA
RNA170_F	GCG CTG CAG ATC TAT CCA A
RNA170_R	ATG CAC ATT CCT GCC CTT AC
tV(UAC)_F	TAG TAT CGA GTT CCG GGT CC
tV(UAC)_R	CTA CTC TTT TCG AAC GCA GAA
HSP150_F	CGG TAA CTT GGC TAT TGG TGA
HSP150_R	CGA TAG CTT CCA AGT GGA CTG
HSP150_F2	GGG CGG TAT CTT AAC TGA CG
HSP150_R2	GAC CAA CCA GCA GCG TAG AT
CIS3_F	CGA CCA AAG AAA CAG CTT CC
CIS3_R	GGT TAC TTG GCT TTG GGT GA
YAC1_F	GAG GAA ATG AGC TGC ATT TTC
YAC1_R	GAG GCA TTA AAC ACA TGG TAG
GAL7_F	CCA ACC AAG AAT TTC CGA AC
GAL7_R	CGC CTC GAT TTT AAA GCA AC
YAC6_F2	TGG ATG TCT GGA AAA CAG CA
YAC6_R2	TAG GTC AAT GCA GCA TCA GC

Table 3.3

Plasmids generated for Chapter 3

Plasmid Name	Description
pAZ2	Integrating plasmid to integrate NAT:pGAL fragment into YAC
pAZ3	Integrating plasmid to integrate NAT fragment into YAC
pAZ8	Integrating plasmid to integrate NAT:pGAL-HSP150 fragment into YAC
pAZ10	Integrating plasmid to integrate NAT:pGAL-CIS3 fragment into YAC
pAZ9A	Integrating plasmid to integrate NAT:HSP150 fragment into YAC
pAZ11A	Integrating plasmid to integrate NAT:CIS3 fragment into YAC
pAZ12A	Integrating plasmid to integrate NAT:CIS3-HSP150 fragment into YAC

References

Arana, M.E. et al., 2012. Transcriptional responses to loss of RNase H2 in Saccharomyces cerevisiae. *DNA repair*, 11(12), pp.933–941.

Castellano-Pozo, M. et al., 2013. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Molecular Cell*, 52(4), pp.583–590.

Ginno, P.A. et al., 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular Cell*, 45(6), pp.814–825.

Legros, P. et al., 2014. RNA processing factors Swd2.2 and Sen1 antagonize RNA Pol III-dependent transcription and the localization of condensin at Pol III genes. G. P. Copenhaver, ed. *PLoS Genetics*, 10(11), p.e1004794.

Sanz, L.A. et al., 2016. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Molecular Cell*, 63(1), pp.167–178.

Van Dijk, E.L., Chen, C.L. & d'Aubenton-Carafa, Y., 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature*.

Wahba, L. et al., 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes & development*, 30(11), pp.1327–1338.

Wahba, L., Gore, S.K. & Koshland, D., 2013. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *eLife*, 2, p.e00505.

Chapter 4: Concluding Remarks

Over the past 15 years, our knowledge of genomic R-loops has grown from functional intermediates in bacterial DNA replication to a pervasive genomic feature associated with both the deleterious outcome of genome instability and positive functions in transcription and chromatin regulation. The work presented here furthers our understanding of this darker side of R-loops, by exploring how their formation affects chromosome instability and how the cellular RNases H act differentially to prevent this. We found that not all R-loops induce chromosome instability in wild-type cells, however the increased genome instability observed in RNase H2 deficient cells derived from many genomic hybrids. Additionally RNase H1 seems to play a more specialized role resolving some but not all hybrids.

We now have a high-resolution map of hybrid formation in the yeast genome. The studies of chromosome instability presented here suggest that a select few hybrids are more mutagenic than others. It will therefore be interesting in the future to investigate if there are specific hybrids throughout the genome that are more prone to DNA damage. This might be accomplished by ChIP-seq of the DNA damage mark H2A-Ser129 phosporylation, which is analogous to γ-H2AX in human cells, as was previously performed by microarray analysis in the R-loop forming mutants in mRNA cleavage and polyadenylation (Stirling et al. 2012). These studies found that more highly transcribed genes and genes near replication origins had a slight increase H2A-Ser129 phosphorylation. A similar mapping study of damage in RNase H mutants could help segregate R-loops that may cause damage from those that avoid damage, and perhaps serve functional roles. Additionally, using proteins that bind directly to damaged DNA, such as the virally derived double strand end binding protein Gam (Shee et al. 2013) could provide more precision.

In a similar vein it will be interesting to have a genome-wide map of RNase H localization. In some ways it was very surprising that we were able to determine RNase H localization with the chromatin immunoprecipitation technique, as one may have expected an enzyme like RNase H1 to interact only transiently with the chromatin. Our studies provide a proof of principle that RNase H1 can be detected. In fact, we found RNase H1 localized to all hybrid-forming regions that we tested. It will be nice to know if this is true genome-wide or if there are certain loci or classes of hybrids that recruit or avoid RNase H1. This could shed light on how cells maintain useful hybrids and eliminate dangerous ones. Additionally, profiling RNase H1 localization throughout the cell cycle could provide insight into its regulation.

The study of RNase H mutants has been very useful in both understanding how the R-loop removing enzymes work and where spontaneous R-loops form in cells. However, future study of how other R-loop forming mutants affect the hybrid landscape will be of much interest. The low-resolution DRIP-chip study, whose limitations have been discussed in previous chapters, mapped R-loops in *hpr1*Δ and *sen1-1* mutants and found some differences in R-loop formation patterns (Chan et al. 2014). However, given the limited differences they observed between wild-type and *rnh1*Δ*rnh201*Δ profiles, it would be worth reinvestigating with the improved S1-DRIP-seq methodology, as well as expanding the analysis to additional R-loop forming mutants. Conversely, it could be interesting to profile R-loops in cells overexpressing either of the RNases H in order to

identify potential functional hybrids that escape RNase H degradation, although these experiments may be technically difficult.

As discussed in chapter 1, there is still a large unknown in the field: the mechanism by which genomic R-loops are converted into DNA damage, double-strand breaks and genome instability. There have been several studies providing evidence that R-loops can lead to replication stress and S-phase damage. However, hybrid-dependent damage has been observed outside of S-phase. Whether this represents two distinct damage mechanisms is unknown. A better understanding of if and how the RNases H and other R-loop removal factors act throughout the cell cycle could further our understanding of the cell cycle timing and resolution of damage. Additional genetic dissection through genetic screens and synthetic genetic interactions could also help illuminate this apparent black box. However, preliminary overexpression screens in the Koshland Lab have not elucidated additional key players.

The study of R-loops in budding yeast has provided a strong foundation for the field. It is clear that R-loops can induce genome instability, and going forward it will be interesting to see if and how R-loops play into the more complex physiology of human cells. While it has been often suggested that R-loops could be a potent instigator of oncogenesis, there has been little in the way of concrete evidence. Additionally R-loops are emerging as intermediate structures in transcriptional regulation. How R-loops function in the more complex transcription landscape of mammalian genomes is open for exploration.

References

Chan, Y.A. et al., 2014. Genome-Wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with DRIP-Chip M. Snyder, ed. *PLoS Genetics*, 10(4), p.e1004288.

Shee, C. et al., 2013. Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. *eLife*, 2, p.e01222.

Stirling, P.C. et al., 2012. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes & development*, 26(2), pp.163–175.

Appendix: Sequence contribution to hybrid formation and chromosome instability

Background

Transcription can lead to DNA:RNA hybrid formation and genome instability. However, not all transcription leads to hybrid formation. As discussed in chapter 1, the first indication of sequence-dependence came from studies of hybrid formation using a plasmid-based transcription associated recombination (TAR) assay in yeast. In the transcription elongation mutant *hpr1*Δ, high levels of TAR were observed when transcribing through the bacterially derived lacZ sequence, but not the yeast derived *PHO5* sequence (Chávez & Aguilera 1997). The TAR was later determined to be due to the formation of a hybrid (Huertas & Aguilera 2003), however the sequence dependence of hybrid formation remained unclear.

Subsequent studies have found that deregulation of transcription can lead to hybrid-mediated genome instability. Yeast mutants that increase overall transcription or allow cryptic transcription, such as those lacking the histone deacetylase Sin3, have increased hybrid formation and chromosome instability (Wahba et al. 2011). Therefore, previous work in the Koshland lab explored the consequences of inducing high levels of transcription of an exogenous, human derived sequence in yeast (Wahba et al. 2013). They utilized a yeast artificial chromosome (YAC) in order to monitor the instability of a non-essential chromosome. They found that inducing high levels of transcription of a human-derived YAC sequence, using a galactose-inducible promoter, resulted in the formation of a hybrid at the locus and increased YAC instability (Wahba et al. 2013). In subsequent unpublished work, Lamia Wahba found that driving overexpression of the yeast-derived *LEU2* transcript from the same locus on the YAC did not induce chromosome instability.

In this work, I sought to determine the underlying factors governing why overexpression of one transcript, but not another, drove chromosome instability. Are there specific sequences in the human-derived transcript that promote hybrid formation? Alternatively are there sequences within the *LEU2* transcript that prevent hybrid formation or the associated chromosome instability? Was this protection against hybrid formation unique to *LEU2* or did it extend to other yeast genes?

I initially reconfirmed the previously reported instability induced by overexpression of the human-derived YAC sequence and lack of increased instability upon overexpression of *LEU2*. However, through the course of the experiments described here, I ran into some difficulties that ultimately resulted in my inability to replicate experiments. I was therefore not confident enough in the results to publish, but I will record the results here in this appendix. The data presented here was reproducible initially. I will discuss the inconsistencies I observed and my attempts to troubleshoot at the end of the results section.

Results

Transcription of multiple human-derived sequences can induce YAC instability

I first sought to determine if the human-derived YAC sequence randomly chosen for overexpression was unique in its ability to induce chromosome instability. For these studies I utilized the same YAC instability assay previously published (Wahba et al. 2011). In this assay, wild-type haploid cells harbor a YAC comprised mostly of sequence from human chromosome VII (Figure 6.1A) (Huang & Koshland 2003). The YAC is similar in size to endogenous yeast chromosomes, but is dispensable for viability of yeast. The chromosome is very telocentric with a short left arm and long right arm, and bears the genetic markers *HIS3* and *URA3*, which can be used to select for chromosome instability events. Cells that have lost the entire YAC will be phenotypically His- and Ura-. Cells that have a terminal deletion on the right arm will be His+ Ura- (Figure 6.1A).

To investigate if additional human derived YAC sequences could induce instability, I inserted a galactose inducible promoter (pGAL) into two additional sites on the YAC. The original site of pGAL insertion, termed pGAL #1, was 10 kb from the right telomere. I inserted pGAL #2 telomeric to pGAL #1, 9 kb from the telomere, and pGAL #3 centromeric to pGAL #1, 158.5 kb from the telomere (Figure 6.2 B).

To induce transcription, cells were grown in bulk and transiently supplied with galactose (Figure 6.2). Initially, cells were grown in media lacking uracil to maintain the YAC. This initial culture was expanded into rich media with lactic acid and glycerol as the carbon source. Unlike glucose, lactic acid glycerol media is non repressive and allows for rapid induction of the GAL promoter. After 3 generations of growth in this non-repressive condition, galactose was added to induce transcription for 5 hours. At the 0 and 5 hour time point cells were plated on media containing 5-FOA to select for cells that had lost the *URA3* marker. Cells were also plated on rich media (YPD) as a control for colony forming units.

Galactose-induced transcription lead to YAC instability in all three pGAL constructs (Figure 6.3 A). The YACs bearing pGAL #1 were the most unstable, about 7 fold over a wild-type YAC with no promoter. The other two constructs also induced instability: pGAL #2 was 5 fold and pGAL #3 was 3.8 fold over wild-type. From this we conclude that additional human-derived YAC sequences can induce chromosomal instability.

I next investigated the nature of the transcript induced at the pGAL #1 insertion locus. I prepared RNA from cells collected at the 0 and 5 hour time points and determined the relative abundance of transcripts at the locus, normalized to ACT1. Upon galactose induction a transcript of around 700 bp was induced to the right of the promoter (Figure 6.3 A). I next performed DRIP to detect hybrid formation at the locus. Although the transcript was only induced to the right of the promoter, hybrid formation could be detected across the locus after 2 hours of galactose induction (Figure 6.3 B). This implies that the transcript from the CloNAT drug marker placed to the left of the promoter may also induce hybrid formation after galactose induction.

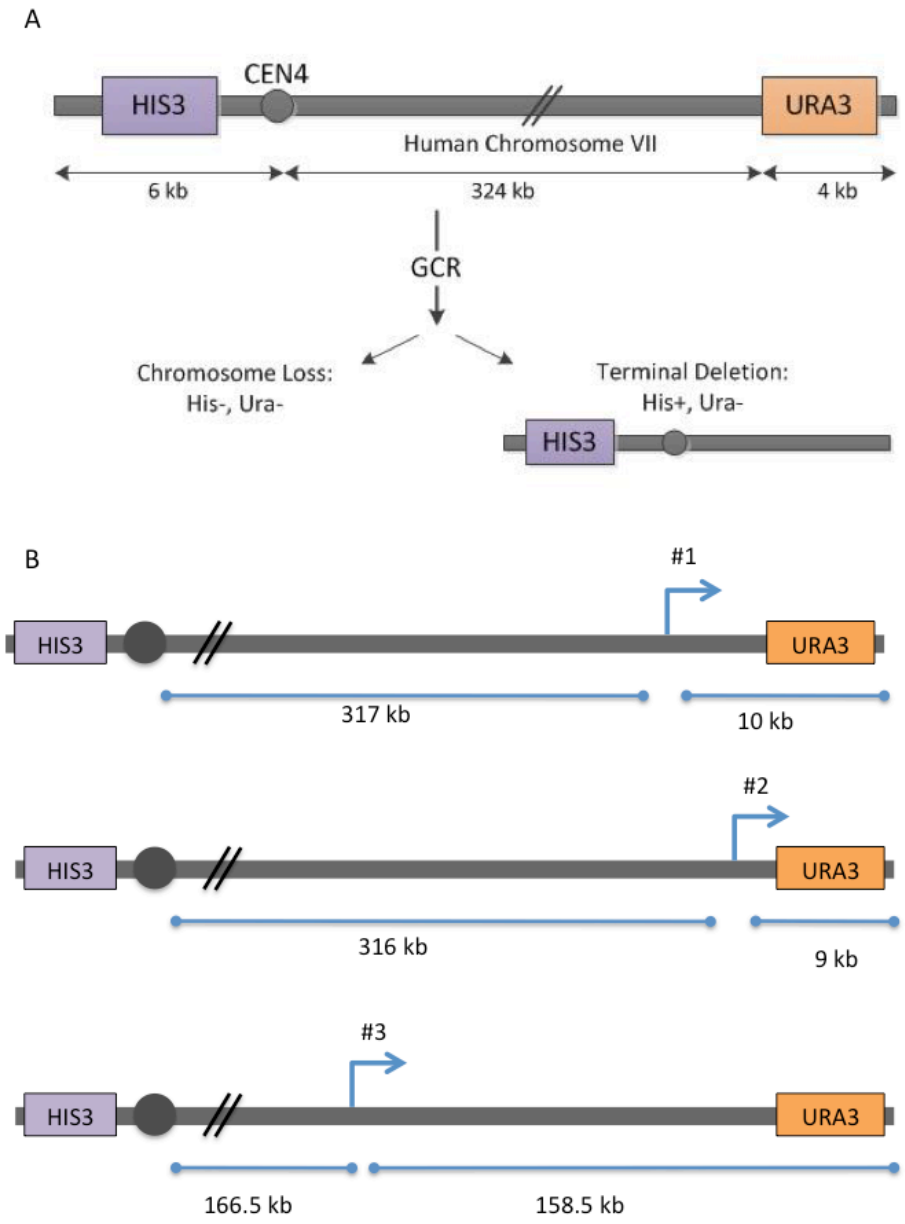


Figure 6.1
YACs used in instability assay

(A) Diagram of YAC used in these studies indicating genetic markers. The nature of a gross chromosomal rearrangement (GCR) can be determined phenotypically as indicated. (B) Diagram indicating the three locations of insertion of pGAL (indicated by blue arrow).

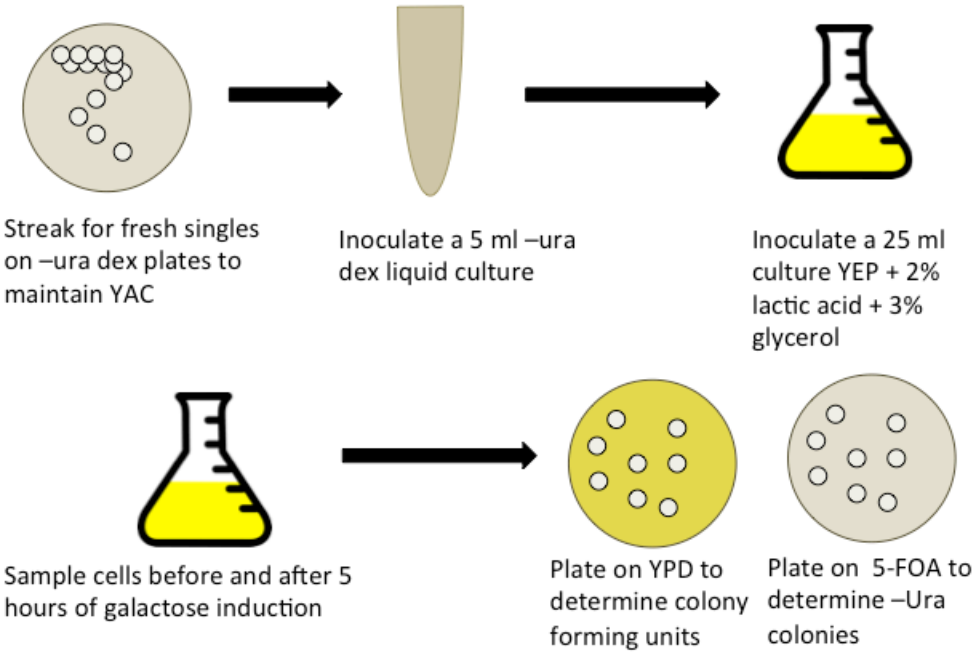


Figure 6.2
Transcription induced YAC instability assay

Schematic overview of assay conditions used to drive galactose induced transcription on the YAC. Cells were subsequently assayed to select for YAC instability.

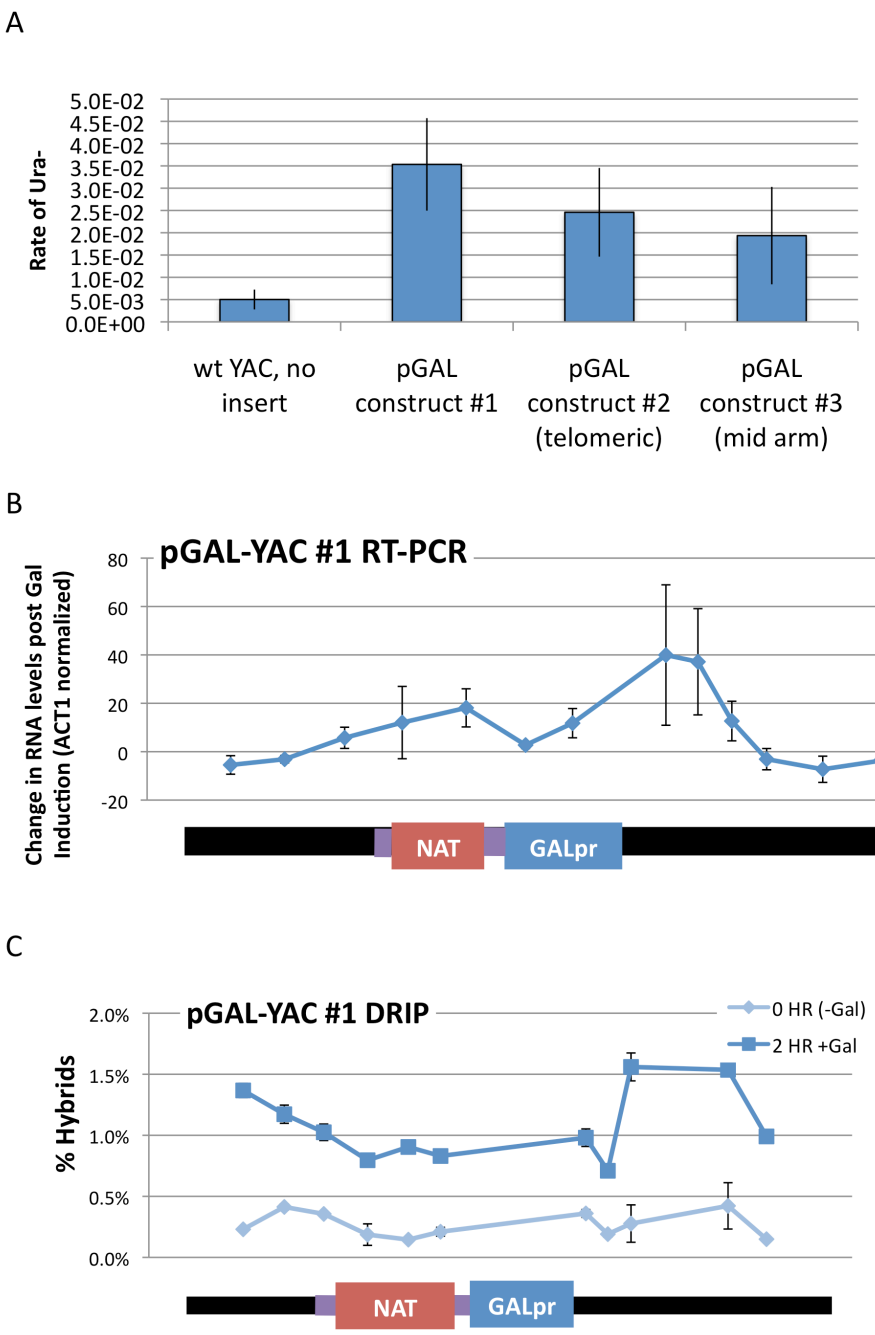


Figure 6.3
Overexpression of human-derived sequences drives YAC instability

(A) YAC instability plotted as rate of –Ura colonies.
(B) RNA abundance of pGAL-YAC#1 as determined by RT-PCR. Plotted as a change in RNA levels from uninduced to Gal induced, normalized to *ACT1* transcript levels.
(C) DNA:RNA hybrid levels at the pGAL-YAC #1 locus as determined by DRIP in uninduced and Gal induced conditions.

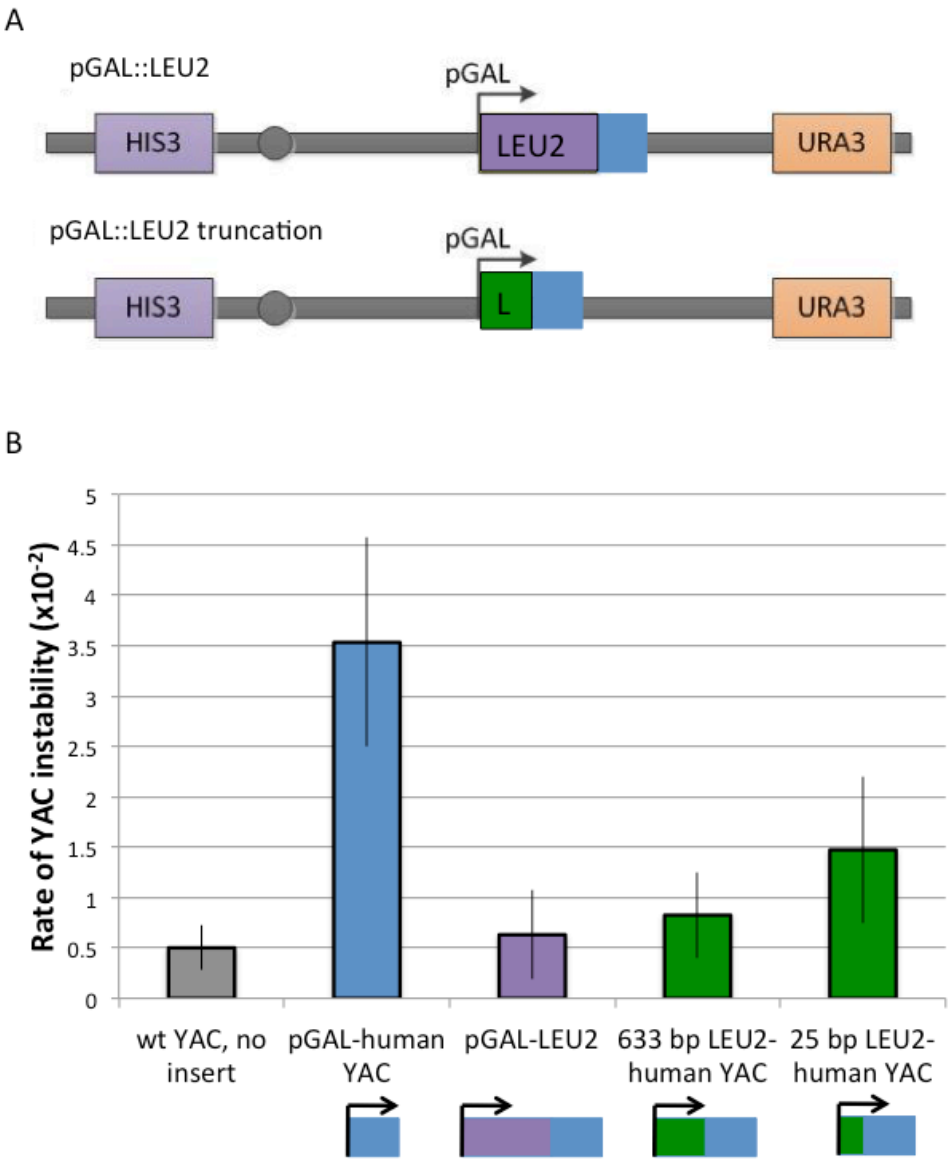
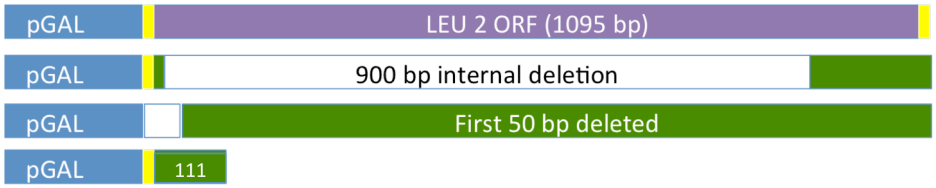


Figure 6.4
Transcription induced YAC instability is suppressed by *LEU2* sequence fusions

(A) Diagram of *LEU2* fusion sequences inserted on the YAC.
(B) YAC instability plotted as rate of –Ura colonies

A



B

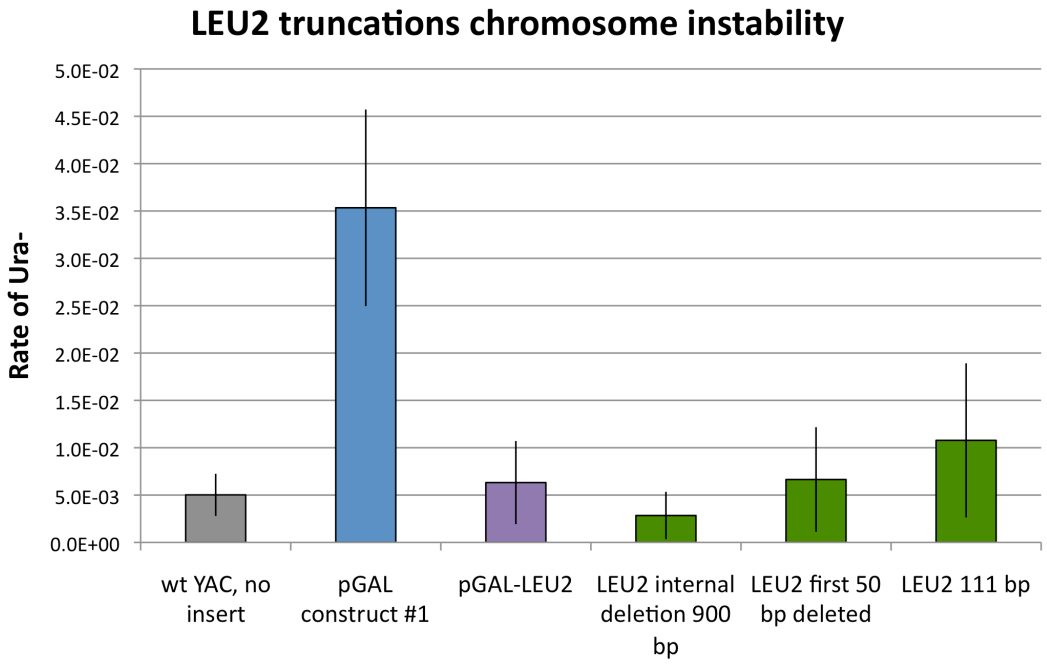


Figure 6.5

Fusion of many *LEU2* derived sequences suppresses YAC instability

(A) Diagram of *LEU2* fusion sequences inserted on the YAC.
(B) YAC instability plotted as rate of –Ura colonies.

Fusion of LEU2 transcript to YAC sequence suppresses chromosome instability

To determine if the pGAL #1 transcript was sufficient to induce instability, I next investigated the effects of fusing *LEU2* to the transcript. I fused the full-length *LEU2* open reading frame (without a transcription terminator) to the YAC sequence (Figure 6.4 A). pGAL driven overexpression of this fusion transcript did not induce YAC instability (Figure 6.4 B). However, when the transcript was detected by RT-PCR, it was found to include both the *LEU2* sequence and the YAC derived sequence (Figure 6.6A) induced upon galactose addition. Therefore, the *LEU2* sequence at the 5' end of the message conferred a stabilizing property that the pure YAC-derived transcript lacked.

I next investigated if part or all of the 1095 bp of the *LEU2* sequence needed to be fused the YAC sequence in order to prevent chromosome instability. Fusing the first 633 bp of the transcript still prevented chromosome instability (Figure 6.4 B). Remarkable, fusing just the first 25 bps also prevented chromosome instability. I verified that the YAC sequence was still being transcribed in this 25 bp fusion (Figure 6.6 B). I additionally tested other deletions of *LEU2:* and internal deletion of 900 bp, deletion of the first 50 bp, and a truncation leaving 111 bp (Figure 6.5 A). Overexpression of all of these fusion transcripts did not result in chromosome instability (Figure 6.5 B).

The lack of chromosome instability in these *LEU2* fusion constructs could either be suppressing hybrid formation or preventing hybrids from being converted into chromosome instability. To distinguish these possibilities, I performed DRIP to detect hybrids in the 25 bp fusion construct. No hybrid signal was detected at the locus before or after galactose induction (Figure 6.6 C). This is consistent with the chromosome stability phenotype of this construct, indicating that this small portion of the *LEU2* transcript somehow prevented hybrid formation and chromosome instability.

Transcription of additional yeast genes does not induce chromosome instability

I next tested whether the transcription of other yeast genes on the YAC could induce chromosome instability. These studies were carried out before any genome-wide maps of hybrid forming sequences in yeast were available. I chose two candidate genes to interrogate: *DYN2* and *CYC1*. *DYN2* was initially chosen because it is one of the few intron-containing genes in yeast. Intron containing genes were hypothesized to be hybrid-prone because hybrid formation increased in cells lacking splicing factors (Li & Manley 2005). However, the subsequent hybrid mapping studies have revealed that *DYN2* is not a hybrid-prone gene. *CYC1* was chosen because its transcription and termination have been well characterized. It also was not found to be hybrid-prone in subsequent hybrid mapping studies. Experiments inserting the hybrid-prone genes *HSP150* and *CIS3* on the YAC are described in chapter 3.

DYN2 and *CYC1* were inserted on the YAC under their native promoters and pGAL. These sequences contained transcription terminators, and the transcripts were not fusions to the human-derived YAC sequences. Under condition of galactose induction, none of these transcripts lead to chromosome instability (Figure 6.7). These experiments, in combination with the study of *HSP150* and *CIS3*, show that transcription of many yeast genes on the YAC do not induce chromosome instability.

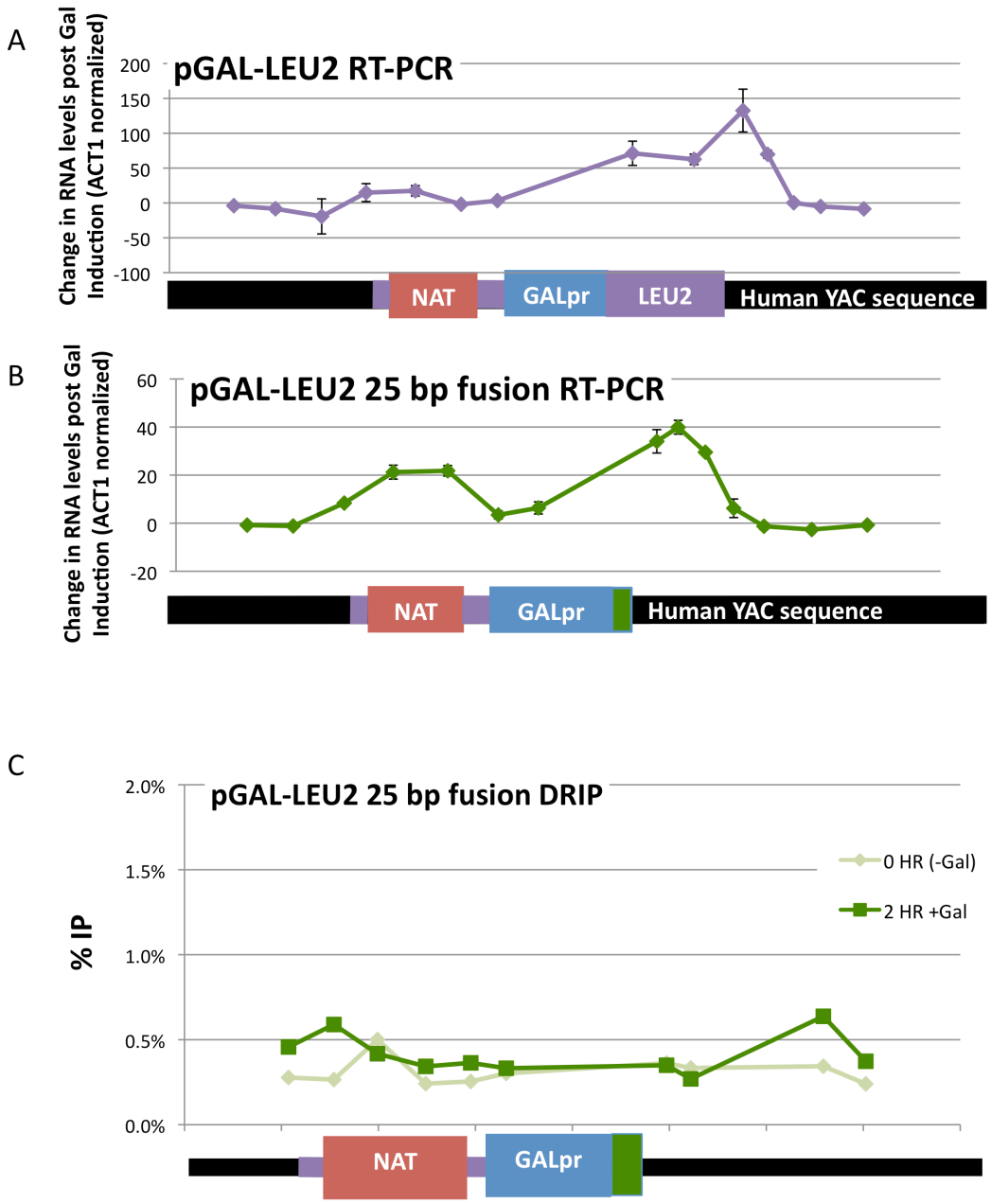


Figure 6.6
Expression and hybrid formation of *LEU2* fusion constructs

RNA abundance of full-length *LEU2* (A) and 25 bp (B) fusion constructs as determined by RT-PCR. Plotted as a change in RNA levels from uninduced to Gal induced, normalized to *ACT1* transcript levels.
(C) DNA:RNA hybrid levels at the 25 bp *LEU2* fusion construct as determined by DRIP in uninduced and Gal induced conditions.

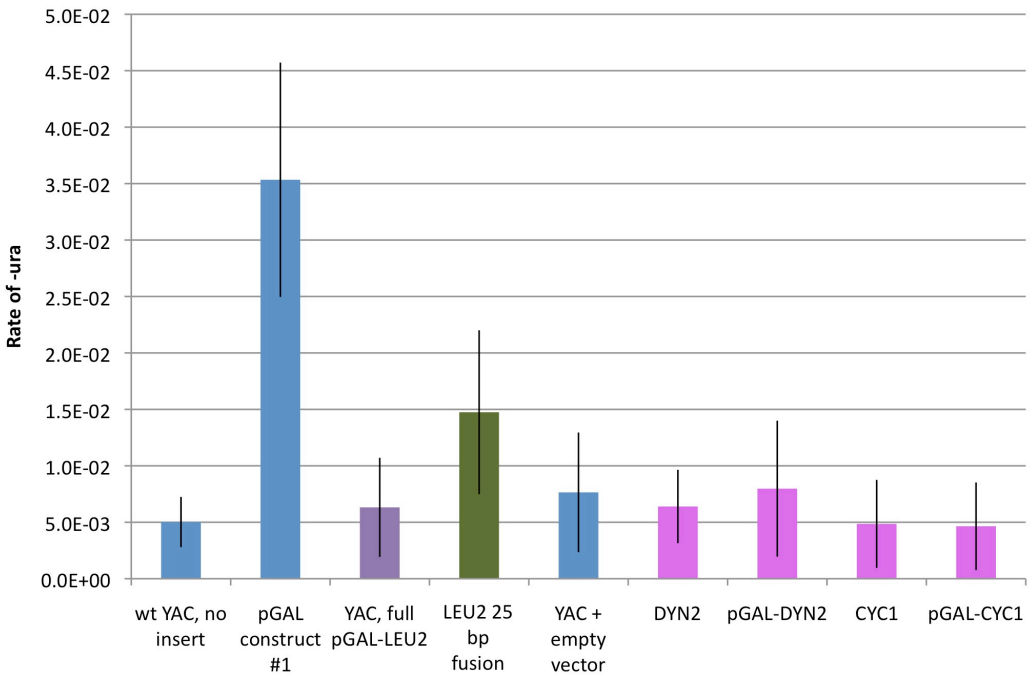


Figure 6.7
Expression of *DYN2* or *CYC1* do not induce YAC instability

YAC instability plotted as rate of –Ura colonies.

Technical difficulties

The results presented here are representative, repeated experiments. However, as I was performing the experiments I observed increasingly sporadic phenotypes in the pGAL YAC #1 construct, which was used as a positive control for galactose inducible instability in all experiments. Sometimes I observed hyper-instability phonotypes where nearly 100% of cells lost the *URA3* marker. More and more often I observed no galactose-inducible chromosome instability. In all cases, the level of induction of RNA was similar, around 70-100 fold induction.

I pursued many avenues to troubleshoot these inconsistencies. I rebuilt all of the strains fresh and assayed them without storing at -80C. I modulated the time of pre-growth in the non-repressive lactic acid glycerol from just one cell division to 5-6 cell divisions. I pre-grew the cells in an alternate non-repressive media using raffinose as a carbon source. I performed the entire assay in defined minimal media supplemented with either all amino acids or lacking uracil. I extended the assay to up to 24 hours of galactose induction and reduced it to as few as two hours. As part of the strain building process, all strains that carry a YAC have their chromosome sizes checked by pulse field gel electrophoresis. In addition to confirming chromosome sizes at the time the strain was

built, I additionally check the chromosome sizes of strains after storage at -80C before use in subsequent assays and found no changes in chromosome sizes. Unfortunately, none of these attempts at troubleshooting clarified the sporadic transcription-inducible chromosome instability phenotype.

Discussion

Here I investigated the role that DNA and/or RNA sequence plays in hybrid formation and subsequent chromosome instability. To do so I drove transcription of different sequences on the YAC, allowing me to monitor the instability of a nonessential chromosome in yeast. Subsequently I performed DRIP to detect DNA:RNA hybrid formation at the transcribed locus. I found that multiple human-derived YAC sequences were able to drive transcription-associated YAC instability. However, transcription of a fusion of part or all of the *LEU2* transcript to a human sequence did not induce chromosome instability. Additionally, this fusion transcript did not drive hybrid formation. Lastly, the overexpression of two additional yeast genes, *DYN2* and *CYC1*, did not drive chromosome instability.

It is intriguing that transcription of three different human derived YAC sequences drove chromosome instability. Transcription occurs pervasively throughout the yeast genome (Nagalakshmi et al. 2008)(reviewed in (Jensen et al. 2013)), however, our S1-DRIP-seq study found that only about 8-12% of the yeast genome forms hybrids. Furthermore, not all hybrid-forming loci induce chromosome instability, as I have shown in chapter 2 and 3 of this dissertation. This implies that cells have evolved ways to deal with many transcripts and many hybrids in a way that does not result in chromosome instability. However, when an exogenous sequence is transcribed, the cell is unable to properly process it, to deleterious consequences. Importantly, all these studies were carried out in otherwise wild-type cells, so how these YAC transcripts have the somewhat unique ability to evade the multiple RNA degradation and hybrid surveillance mechanisms is mysterious. In the screen of non-essential yeast genes important for YAC stability (Wahba et al. 2011), mutants in numerous RNA biogenesis pathways were found to increase hybrid mediated chromosome instability. These results taken with the overexpression of YAC transcripts presented here show that small changes in RNA physiology in the cell can have large repercussions on genome integrity.

The series of *LEU2* fusion constructs that were able to suppress the transcription associated chromosome instability confound this view of a precarious transcriptome susceptible to minor changes. A fusion of just the first 25 bp of *LEU2* to the human YAC sequence was sufficient to prevent hybrid formation and instability; perhaps this small amount of yeast-derived sequence was able to engage RNA processing factors that the human-derived sequence could otherwise evade. We know that perturbing RNA polymerase elongation leads to hybrid formation even as it deceases the overall abundance of transcripts (Huertas & Aguilera 2003). This suggests that small changes in RNA is created can have large effects. Perhaps the human sequence is a difficult template for yeast RNA polymerases, and the presence of yeast derived sequences eases transcription at the locus. Unfortunately the technical difficulties encountered here precluded further study of questions like this.

Materials and Methods

Galactose induced YAC instability assay

Performed as described in chapter 3.

RNA transcript abundance by RT-PCR

Performed as described in chapter 3.

DRIP

Performed as described in (Wahba et al. 2013).

Table 6.1

Strains used in Appendix

Description	Strain	Genotype
YAC strains		
WT empty YAC	LW6732c	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1)*
pGAL-YAC #1	LW6811c	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL1-10)*
pGAL-YAC #2	AZC25a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL1-10)*
pGAL-YAC #3	AZC31a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL1-10)*
pGAL-LEU2	LW6809	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL-LEU2)*
pGAL-633 bp LEU2	AZ22a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL-633 bp LEU2)*
pGAL-25 bp LEU2	AZ34a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL-25 bp LEU2)*
pGAL-LEU2 internal deletion 900 bp	AZ23a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL-LEU2 900bp internal deletion)*
pGAL-LEU2 first 50 bp deleted	AZ26a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL-LEU2 first 50 bp deleted)*
pGAL-111 bp LEU2	AZ30a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL-111 bp LEU2)*
YAC + empty vector	AZ37a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT)*
YAC pGAL #1 remake with NotI	AZ36a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL)*

cloning site		
YAC DYN2	AZ38a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-DYN2)*
YAC pGAL-DYN2	AZ43a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL-DYN2)*
YAC CYC1	AZ41a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-CYC1)*
YAC pGAL-CYC1	AZ40a	*MATa his3Δ ura3Δ0 met15Δ0 leu2Δ0 /YAC (MFA1pr-HIS3 URA3 MET15 TRP1 CLONAT-pGAL-CYC1)*

Table 6.1

Primers used in Appendix

Primer Name	Reverse Primer sequence	midpoint of primer (bp)	Forward Primer sequence	midpoint of primer (bp)	product size (bp)	midpoint of primer set in bp
YAC -5 F-R	GGG AAC TAA ATG TGT AGG TGG T	321544	GAG AGG ATG CTG CAA AGA GC	321759	215	**321651.5**
YAC -4 F2-R2	AGA GTT CCA GGG CTG TCA AA	321958	TCA GGA AGG ATG AAG ACC AGA	322187	229	**322072.5**
YAC -3a F-R	TAT GGA ATT CAA CTT ACC TTC	322432	AGG GAA TGG AGA CAT AAA CC	322649	217	**322540.5**
YAC -3 F2-R2	TGG ATG CAG TAG TGG GGA GT	322877	TGC AGG AAA CCT GGA AAC AT	323102	225	**322989.5**
YAC -2 F-R	GTT AGG ATT TGC CAC TGA GG	323363	CCA CCG GCA CCT CCC GCA GG	323613	250	**323488**
YAC -1a F-R	ACC TCT GGC TGG AGG TCA C	323878	CAG GGC ATG CTC ATG TAG AG	324025	147	**323951.5**
YAC -1 F2-R2	TGG TCG CTA TAC TGC TGT CG	324222	GCC TGA TGC GGT ATT TTC TC	324414	192	**324318**
YAC 0 F-R	TGC AAG GCG ATT AAG TTG GG	324506	CCC GCT CGG CGG CTT CTA ATC	324771	265	**324638.5**
LEU2 F-R	ATG TCT GCC CCT AAG AAG ATC	325352	TCC TGA AAT TGC TGA TCC TT	325452	100	**325402.5**
YAC 1 F-R	GAG GCA TTA AAC ACA TGG TAG	325373	GAG GAA ATG AGC TGC ATT TTC	325619	246	**325496**
YAC 2 F-R	CTA AAT CTT CAT TGC TCC AC	325620	GAA AAT GTG CTA GGC ACC GTA C	325871	251	**325745.5**
YAC 3 F2-R2	AGG CAA GTA AGC CTT TTC CA	325916	CAA GCA TGC CAT AAA TGT	326104	188	**326010**

			TCA			
YAC 4 F-R	CAT AAT GTC CCT AAT CCT ACC	326139	CCC ATA TTT CCC CAA ATA AAG	326424	285	**326281.5**
YAC 5 F-R	TAG CCC TTT TCA GAC TCT GC	326595	AGC CAA TCT ACA GAC TGG CC	326843	248	**326719**
YAC 6 F2-R2	TAG GTC AAT GCA GCA TCA GC	327060	TGG ATG TCT GGA AAA CAG CA	327283	223	**327171.5**

References

Chávez, S. & Aguilera, A., 1997. The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes & development*, 11(24), pp.3459–3470.

Huang, D. & Koshland, D., 2003. Chromosome integrity in Saccharomyces cerevisiae: the interplay of DNA replication initiation factors, elongation factors, and origins. *Genes & development*, 17(14), pp.1741–1754.

Huertas, P. & Aguilera, A., 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular Cell*, 12(3), pp.711–721.

Jensen, T.H., Jacquier, A. & Libri, D., 2013. Dealing with pervasive transcription. *Molecular Cell*, 52(4), pp.473–484.

Li, X. & Manley, J.L., 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell*, 122(3), pp.365–378.

Nagalakshmi, U. et al., 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881), pp.1344–1349.

Wahba, L. et al., 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Molecular Cell*, 44(6), pp.978–988.

Wahba, L., Gore, S.K. & Koshland, D., 2013. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *eLife*, 2, p.e00505.