UNIVERSITY OF CALIFORNIA, SAN DIEGO

Identification and Characterization of the Cdc73 Transcription Factor as a Suppressor of Genome Instability

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Rahul V. Nene

Committee in charge:

Professor Richard D. Kolodner, Chair
Professor Arshad Desai
Professor Steven F. Dowdy
Professor Jan Karlseder
Professor Albert R. La Spada

2015

The Dissertation of Rahul V. Nene is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

DEDICATION

I dedicate this dissertation to all of my science mentors, teachers, collaborators, and peers, who have fostered my interest in science and encouraged me to pursue this challenging and rewarding field. I also dedicate this to my parents and the rest of my family in Kenya for their incredible support, and to the Fern Glen community, my family in San Diego, for helping to keep me sane during these many years of training.

EPIGRAPH

The most exciting phrase to hear in science, the one that heralds new discoveries, is not “Eureka!” but “That’s funny…”

Isaac Asimov

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CHEF	Clamped Homogeneous Electric Field
DSB	Double Strand Break
GCR	Gross Chromosomal Rearrangement
HR	Homologous Recombination
NHEJ	Non-homologous end joining
PFGE	Pulse field gel electrophoresis
TCGA	The Cancer Genome Atlas

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would first like to thank my advisor, Richard Kolodner, for his amazing guidance and support. I cannot believe how much I learned from him during these past four years and I cannot imagine getting a better foundation for a career in the life sciences. I would also like to thank Christopher Putnam for being a superb mentor and an incredible and patient teacher; he has really helped me to think more critically about anything and everything. I would also like to thank Anjana Srivatsan, Hans Hombauer and Catherine Smith, for initiating me into the world of yeast genetics, and the other members of the Kolodner laboratory, past and present, including Katie Pallis, Sara Bell, Sandra Martinez, Sarah Clotfelter, Nikki Bowen, Binzhong Li, Bill Graham, Matt Duprie, Elaine Guo, and Betsy Van Ness for their incredible support over the years. I would also like to thank Christopher Campbell in the Desai laboratory for his collaboration, which helped to expand the scope of my project. I thank my thesis committee members, Arshad Desai, Steve Dowdy, Jan Karlseder, and Al La Spada, for their support, advice, feedback, and insightful questions during my thesis committee meetings. I would like to recognize Paul Insel, Mary Alice Kiisel, and the rest of the faculty and support staff within the Medical Scientist Training Program for their invaluable support and advice during my training at UCSD. I would also like to thank my parents for their support and encouragement to pursue this field. I would also like to thank my previous research mentors, particularly Vishva Dixit and Herve Tettelin, for fostering my interest in the biomedical sciences.

Chapter 1 is an adaptation of a manuscript that has been submitted for publication. The authorship and title are as follows: Putnam CD*, Srivatsan A*, Nene RV*, Martinez SL*, Clotfelter SP, Bell SN, Somach S, de Souza JES, Fonseca AF, de Souza SJ, Kolodner RD. An extended genetic network suppresses genome rearrangements in *S. cerevisiae*. (*Indicates first co-authorship). The work presented in this chapter was done during the first half of my graduate training. Christopher D. Putnam and Richard D. Kolodner conceived the overall experimental design and the systematic mating strategy. Anjana Srivatsan, Rahul V. Nene, Sandra L. Martinez, Sarah P. Clotfelter, and Sara N. Bell did strain construction, patching, and quantitative rate measurements. Christopher D. Putnam analyzed the resulting patch scores and rates. Sandro J. de Souza, Jorge E.S. de Souza, André F. Fonseca, Steven Somach, and Richard D. Kolodner analyzed the TCGA data. Christopher D. Putnam and Richard D. Kolodner wrote the manuscript and all other authors revised and modified the manuscript. The authors of the manuscript also thank Vincent Pennaneach and Jorrit Ensernik for assistance in construction of early versions of the query strains and Renan Valieris for assistance in the construction of the MySQL database containing TCGA data. This work was supported by NIH grant R01-GM26017 to Richard D. Kolodner, NIH grant F30-CA177240 to Rahul V. Nene, CAPES (Brazil) grant (23038.004629/2014-19) to Sandro J. de Souza and support from the Ludwig Institute for Cancer Research to Richard D. Kolodner and Christopher D. Putnam.

Chapter 2 is an original document that is being prepared for publication. As of May 19th 2015, the authorship and tentative title of this manuscript are: Nene RV, Putnam CD, Campbell CS, Desai A, Kolodner RD. Cdc73 Suppresses Genome Instability by Mediating Telomere Homeostasis and RNA:DNA Hybrid Formation. The dissertation

author was the primary author of this material and contributed to the conception and design of experiments, execution of experiments, data analysis, and manuscript writing. The dissertation author would like to thank Christopher S. Putnam and Richard D. Kolodner for assisting with the conception and design of the research, data analysis, manuscript writing and mentorship.

Chapter 3 is an original document that is being prepared for publication. As of May 19th 2015, the authorship and tentative title of this manuscript are: Nene RV, Putnam CD, Campbell CS, Desai A, Kolodner RD. Identification of a Domain of Cdc73 That is Necessary and Sufficient for the Suppression of Genome Instability. The dissertation author was the primary author of this material and contributed to the conception and design of experiments, execution of experiments, data analysis, and manuscript writing. The dissertation author would like to thank Christopher S. Campbell for conducting the confocal microscopy and image analysis. The dissertation author would also like to thank Christopher S. Putnam and Richard D. Kolodner for assisting with the conception and design of the research, data analysis, manuscript writing and mentorship.

VITA

2005 – 2009	Brown University Bachelor of Science, Biology *Magna Cum Laude* *Phi Beta Kappa Honor Society* *Sigma Xi Honor Society*
2009 – 2011	University of California, San Diego, School of Medicine Medical Scientist Training Program, Year 1-2 Medical Student
2011 – 2015	University of California, San Diego, School of Medicine Medical Scientist Training Program, Years 3-6 Biomedical Sciences Graduate Student
2011 – 2015	Ludwig Institute for Cancer Research, San Diego Cancer Genetics Laboratory Graduate Research
2013 – 2016	National Research Service Award Training Fellowship
2015	University of California, San Diego Doctor of Philosophy, Biomedical Sciences
2015 – 2017	University of California, San Diego, School of Medicine Medical Scientist Training Program, Years 7-8 Medical Student

PUBLICATIONS

Peer-Reviewed Original Research:

Putnam CD*, Srivatsan A*, **Nene RV***, Martinez SL*, Clotfelter SP, Bell SN, Somach S, de Souza JES, Fonseca AF, de Souza SJ, Kolodner RD. An extended genetic network suppresses genome rearrangements in *S. cerevisiae*. Manuscript submitted. 2015.

* Indicates first co-authorship.

Nene RV, Putnam CD, Kolodner RD. Working title: Cdc73 Suppresses Genome Instability by Mediating Telomere Homeostasis and RNA:DNA Hybrid Formation. Manuscript in preparation.

Nene RV, Putnam CD, Campbell CS, Arshad Desai, Kolodner RD. Working title: Identification of a Domain of Cdc73 That is Necessary and Sufficient for the Suppression of Genome Instability. Manuscript in preparation.

W. Ittiprasert, **R. Nene**, A. Miller, N. Raghavan, F. Lewis, J. Hodgson, and M. Knight, "Schistosoma mansoni infection of juvenile Biomphalaria glabrata induces a differential stress response between resistant and susceptible snails," *Exp. Parasitol.*, vol. 123, pp. 203–211, 2009.

J. C. Dunning Hotopp, M. E. Clark, D. C. S. G. Oliveira, J. M. Foster, P. Fischer, M. C. Muñoz Torres, J. D. Giebel, N. Kumar, N. Ishmael, S. Wang, J. Ingram, **R. V. Nene**, J. Shepard, J. Tomkins, S. Richards, D. J. Spiro, E. Ghedin, B. E. Slatko, H. Tettelin, and J. H. Werren, "Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes.," *Science*, vol. 317, pp. 1753–1756, 2007.

ABSTRACT OF THE DISSERTATION

Identification and Characterization of the Cdc73 Transcription Factor as a Suppressor of Genome Instability

by

Rahul V. Nene
Doctor of Philosophy In Biomedical Sciences
University of California, San Diego, 2015

Professor Richard D. Kolodner, Chair

Genome instability is an increasingly well-appreciated aspect of tumorigenesis, and genome rearrangements, such as translocations, copy number changes, and aneuploidy, are seen in many cancers. The Kolodner Laboratory has developed a variety of assays to study the formation of Gross Chromosomal Rearrangements (GCRs) in the model organism *Saccharomyces cerevisiae*, and our goal based on this work is to leverage the power of yeast genetics to gain insights in to the mechanisms by which increased genome instability contributes to cancer in humans. In Chapter 1, we describe

collaborative efforts that used a large-scale screen to identify all the genes and pathways that interact to suppress genome instability in budding yeast. This led to the identification of 183 genes that directly suppress the accumulation of GCRs, 65 of which have not been previously identified. Bioinformatic analysis of cancer genome databases found that the human homologs of these genes are mutated in 90% and 70% of ovarian and colorectal cancers, respectively. Among the 65 novel suppressors was *CDC73*, a member of the Paf1 Complex that functions in transcription elongation and whose human homolog is a tumor suppressor. In Chapter 2, we investigate the mechanisms by which this gene suppresses genome instability. We demonstrate loss of *CDC73* synergizes with mutations in telomere maintenance genes, and we show that increased GCRs rates are due to defects in telomerase and the accumulation of recombinogenic RNA:DNA hybrids. In Chapter 3, we expand this analysis to the rest of the Paf1 Complex members and demonstrate there is not a direct correlation between loss of complex function and an increase in GCR rate. We also defined an approximately 100 residue region of Cdc73 that is necessary and sufficient for its function and determined this region is necessary for nuclear localization and binding to Paf1. These findings on the structure and function of Cdc73 provide insights into how the human homolog functions as a tumor suppressor.

Introduction

Background and Significance

Cancer is the second leading cause of mortality in the United States (Heron, 2013). Although recent advances in immunotherapy and targeted therapies are beginning to impact cancer treatment, the three long-standing major forms of treatment are surgery, radiation therapy, and chemotherapy. Radiation therapy and many types of chemotherapy share a common mechanism of introducing DNA damage into the cancer cells, resulting in increased cell killing or decreased cell growth relative to normal cells (Madhusudan and Middleton, 2005). The rationale for this approach is that neoplastic cells typically have a rapid turnover rate, and are therefore more likely to be susceptible to DNA damage. However, numerous normal cells in the human body also share a rapid cell cycle and many normal tissues exhibit some level of cell division, which leads to the common side effects of these therapies, such as immune suppression, nausea and vomiting, and alopecia (Lemieux et al., 2011). The promise for the future of cancer therapy is the development of treatments that are more specific for targeting only cancer cells.

A number of common and rare cancer susceptibility syndromes, such as Ataxia Telangiectasia, Lynch Syndrome, and familial breast and ovarian cancers, are caused by

inherited mutations in genes in homologous recombination (HR), DNA damage response (DDR), and other DNA repair pathways (Canman and Lim, 1998; Fishel et al., 1993; Yoshida and Miki, 2004). Numerous studies have also shown that sporadic tumors develop defects affecting many of these same genes (Kolodner et al., 2011; Lengauer et al., 1998). Large scale projects such as The Cancer Genome Atlas (TCGA) hold promise to discover many more mutations that link genome instability to cancer (Chin and Gray, 2008). This has led to the hypothesis that it may be possible to exploit the inherent genome instability of cancers by introducing enough stress in these cancer cells to specifically kill them without harming normal healthy cells (Bunting et al., 2010).

The overall implication from large scale cancer genome analysis efforts is that defects in DNA repair are common. Thus, there is the potential to expose these cancers to increased sensitivity by stressing or inhibiting compensatory pathways (The Cancer Genome Atlas Research Network, 2013). However, analysis of the data generated by TCGA and other cancer genome projects is limited by a lack of rapid functional assays. The work presented in this dissertation is significant because it demonstrates that the use of the budding yeast *Saccharomyces cerevisiae* as a model organism can facilitate the study of the genes, pathways, and mechanisms that suppress genome instability in higher eukaryotes, particularly in the setting of cancer. We hope this understanding will eventually lead to the discovery of therapeutic targets that can ultimately be exploited in the development of new cancer treatments.

Studying Genome Instability in Yeast

The utility of *S. cerevisiae* as a model to study genome rearrangements was recognized over 30 years ago, with studies on engineered homology-directed genome rearrangements (Jinks-Robertson and Petes, 1986; Mikus and Petes, 1982; Sugawara and Szostak, 1983). More recently, a variety of assays have been developed in the Kolodner laboratory that allow for detection of a broad range of gross chromosomal rearrangements (GCRs) and the genes that suppress them. Chen and Kolodner described the first genetic assay for measuring the rate of gross chromosomal rearrangement (GCR) formation (Chen and Kolodner, 1999). The assay involved placing a *URA3* gene adjacent to *CAN1* on the non-essential terminal region of the left arm of Chromosome V. *CAN1* encodes a plasma membrane arginine permease, and mutation of this gene confers resistance to the drug canavanine (Broach et al., 1979), while *URA3* encodes orotidine-5'-phosphate (OMP) decarboxylase, which converts 5-fluoroorotic acid (5-FOA) into a toxic compound, so loss of *URA3* confers resistance to 5-FOA (Boeke et al., 1984). Spontaneously occurring GCRs that result in the loss of these two genes allow the yeast to grow on media containing both canavanine and 5-FOA. Differences in the number of colonies that grow on nonselective yeast extract-peptone-dextrose (YPD) agar versus media containing both canavanine and 5-FOA can be used to calculate the GCR rate (Lea and Coulson, 1949). In the original assay, breakpoints leading to the formation of GCRs occurred in the region between the *CAN1* gene and the *PCM1* gene, which is the first centromeric essential gene (Putnam et al., 2004a, 2005). This so-called breakpoint region contained single-copy DNA sequences; i.e. it shared little homology with any other

region in the genome. The features of the assay selected for GCRs mediated by single copy sequences, and were capable of detecting GCRs such as broken chromosomes healed by *de novo* telomere addition and various translocations mediated by both homologous recombination (HR) and non-homologous end joining (NHEJ). Alternative versions of the GCR assay have been developed that vary based on the sequence located within the breakpoint region. A modified GCR assay was developed to query both single-copy GCRs and GCRs mediated by non-allelic HR between large (>1 kb) imperfect homologies that resemble the segmental duplications found in highly repetitive mammalian genomes (Putnam et al., 2009). In summary, this modified assay gives a higher baseline GCR rate, identifies genes that suppress other types of GCRs besides those mediated by single copy sequences, and interrogates DNA structures that closely mimic those seen in mammalian genomes. Another GCR assay was recently developed in which a Ty element, along with the *CAN1* and *URA3* genes, were placed on the telomeric end of chromosome V (Chan and Kolodner, 2011). Ty elements are retrotransposons analogous to human Alu and LINE elements and are found at a high rate throughout the yeast genome. In contrast to the low-copy repeat sequence mediated GCRs in the segmental duplication GCR assay, the Ty assay allows the study of GCRs mediated by high copy number repeats. The Kolodner laboratory has developed a variety of methods for determining the structure of the GCRs that are formed and has discovered that mutations in certain genes can have a strong effect on both the rate and the type of GCRs formed (Chan and Kolodner, 2012; Chen and Kolodner, 1999; Motegi and Myung, 2007; Pennaneach and Kolodner, 2009a; Schmidt et al., 2006). The first part of this dissertation

describes the systematic analysis that was conducted to identify all the genes and pathways that suppress genome instability in *S. cerevisiae*. This screen, to which this author contributed, led to the identification of a novel suppressor of genome instability, the transcription factor *CDC73*. The rest of the dissertation is dedicated to the analysis of the mechanism and structure by which Cdc73 suppresses genome instability, as primarily conducted by this author.

Cdc73 and the Paf1 Complex

CDC73 is a member of the Paf1 complex (Paf1C), which binds to and modifies the activity of RNA polymerase II during transcription (Jaehning, 2010). In the budding yeast *Saccharomyces cerevisiae*, the proteins Paf1 and Cdc73 were identified as forming a complex with RNA polymerase II independently of Srbps (Shi et al., 1997; Wade et al., 1996). Further analysis led to the identification of Rtf1, Ctr9 and Leo1 as additional components of this complex (Koch et al., 1999; Krogan et al., 2002; Mueller and Jaehning, 2002; Squazzo et al., 2002). Paf1C has been implicated in a variety of cellular processes, including transcription elongation, 3'-end mRNA maturation, and histone modification (Jaehning, 2010; Krogan et al., 2002, 2003; Nordick et al., 2008; Tomson and Arndt, 2013). It is well conserved among eukaryotes, from budding yeast and *Drosophila* to zebrafish and humans (Newey et al., 2009).

The precise structure and function of the complex is poorly understood. No enzymatic activity has been detected in any of the subunits, consistent with the hypothesis that the complex functions as a scaffold and aids in the recruitment of other

proteins (Jaehning, 2010). The best defined domain of any subunit is the Plus3 domain of Rtf1, a 90 amino acid region that is essential for promoting histone modification (Piro et al., 2012). The only structural data for the complex is of the well-conserved C-terminus of Cdc73 (Amrich et al., 2012; Chen et al., 2012). This region was shown to play a modest role in transcription elongation and histone modification, and was independently shown to be important for the association of Paf1C with chromatin (Qiu et al., 2012).

Although the complex contains 5 subunits, only for the human homolog of *CDC73* is there strong evidence for a role in preventing tumorigenesis. The function of *CDC73* as a tumor suppressor has been linked to breast, renal, and gastric cancers, is often mutated in parathyroid cancer, and germline mutations of *CDC73* cause the cancer susceptibility syndrome hyperparathyroidism-jaw tumor syndrome (HPT-JT) (Newey et al., 2010). The human homolog has also been implicated in Wnt/β-catenin signaling and cell cycle regulation, but little else with regard to its function has been demonstrated (Mosimann et al., 2006; Zhang et al., 2006). Thus, we sought here to leverage the power of yeast genetics to analyze how the structure and function of Cdc73 contributes to the suppression of genome instability, with the hope that this may provide insights in to how the human homolog may function as a tumor suppressor. As explained in Chapter 2, we determined that loss of *CDC73* causes a large synergistic increase in GCR rate with mutations in genes that function in telomere homeostasis, in particular, *TEL1* and *YKU8* We demonstrate this interaction is due to defects in telomerase and the accumulation of RNA:DNA hybrids at subtelomeric sites of transcription. Provided here is a brief introduction to telomere biology.

Telomeres Protect the Ends of Chromosomes and Promote Genome Stability

Telomeres are nucleoprotein structures that perform an important capping function at the ends of linear chromosomes (Wellinger and Zakian, 2012). They prevent the chromosome ends from being degraded by nucleases and from being identified as the site of a double strand break by the DNA damage response (Garvik et al., 1995; Sandell and Zakian, 1993). Telomere biology is very well conserved throughout eukaryotes, and many of the pathways were first determined in *S. cerevisiae*. Numerous proteins contribute to telomere structure and function, many of which are essential, highlighting the importance of telomeres for maintaining viability. A few of the genes discussed in this dissertation are described here. *YKU70* and *YKU80* encode subunits of the Ku complex, which plays a role in telomere capping, telomere length maintenance, and the nonhomologous end joining (NHEJ) pathway of double strand break repair (Bertuch and Lundblad, 2003; Boulton and Jackson, 1996). *TEL1* encodes a protein kinase involved in DNA damage checkpoint and also plays a role in telomere length regulation (Greenwell et al., 1995; Morrow et al., 1995). *EXO1* encodes an exonuclease that acts in multiple aspects of DNA metabolism, including resection at deprotected telomeres (Fiorentini et al., 1997; Tran et al., 2004). And, *SIR2*, *SIR3*, and *SIR4*, encode members of a complex that is required for mating, but also play a role in telomere maintenance and regulating transcription at telomeric sites (Moretti and Shore, 2001; Moretti et al., 1994; Rine and Herskowitz, 1987).

In yeast, telomeres are maintained as ~300 bp of simple repeats, typically abbreviated $C_{1\text{-}3}A/TG_{1\text{-}3}$. The telomeres also contain X and/or Y' elements, repetitive

sequences that are found just centromeric to the $C_{1-3}A/TG_{1-3}$ repeats. Virtually all telomeres contain X elements, though only half the telomeres contain Y' elements (Horowitz et al., 1984; Zakian et al., 1986). Telomeres are subject to the end-replication problem as defined by James Watson, whereby the necessity of a 5' RNA primer to initiate replication necessarily results in the shortening of telomeres during every round of replication (Watson, 1972). This contributes to the concept of the Hayflick limit, which describes how many times a human cell can divide before it undergoes senescence (Hayflick and Moorhead, 1961). However, cells can also activate telomerase, a complex composed of RNA and proteins that can extend the 3' end of the telomere (Greider and Blackburn, 1989). In yeast, telomerase is composed of *EST1*, *EST2*, and *EST3*, which encode the protein subunits, and *TLC1*, which encodes the RNA template. Telomerase is constitutively active in yeast, so even though mutations affecting different telomere maintenance functions may cause shorter telomeres, the cells are still able to maintain viable chromosomes. But, if telomerase is fully inactivated, telomeres undergo shortening, ultimately triggering growth arrest and the cells eventually die (Enomoto et al., 2002; Khadaroo et al., 2009; Lundblad and Blackburn, 1993). However, rare survivors may emerge from the arrested cultures, and these cells rely on recombination to maintain viable telomeres. These cells fit in one of two classes: type I survivors rely on amplification of Y' elements (Lundblad and Blackburn, 1993), while type II survivors show only minor amplifications of subtelomeric repeats, but instead show large, variable amplification of $C_{1-3}A/TG_{1-3}$ telomeric repeats (Teng and Zakian, 1999) .

Contrary to the traditional view of telomeres as being sites that are transcriptionally repressed, there is growing evidence that non-coding telomeric repeat-containing RNA (TERRA) is transcribed by RNA polymerase II (Azzalin et al., 2007; Luke et al., 2008). TERRA levels are tightly regulated and are thought to play a role in the nucleoprotein structure of telomeres (Azzalin and Lingner, 2015). TERRA transcription is regulated by many of the same factors that control telomeric silencing, such as the *SIR2/3/4* complex (Iglesias et al., 2011), and TERRA transcription is stimulated at shortened telomeres to aid in the recruitment of telomerase in a telomere specific fashion (Cusanelli et al., 2013). However, artificial overexpression of TERRA promotes Exo1 dependent telomere resection, demonstrating how fine the balance is between TERRA levels and telomere function (Pfeiffer and Lingner, 2012). Defects in TERRA transcription have also been implicated in telomere dysfunction, and this is most well described for mutants of the THO complex, which is composed of Hpr1, Tho2, Thp2, and Mft1, and plays a role in pre-mRNA processing and mRNA export (Chávez et al., 2000; Jimeno et al., 2002; Rondón et al., 2010). Deletion of *HPR1* results in the accumulation of RNA:DNA hybrids, or r-loops, a compromised structure whereby one strand of DNA is paired with RNA, while the other DNA strand is left exposed. Throughout the genome, these r-loops can promote genome instability, as they are recombinogenic and can block incoming DNA replication forks (Bermejo et al., 2012; Chávez and Aguilera, 1997; Gómez-González et al., 2011) This genome instability can be suppressed by the overexpression of RNase H1, an enzyme that specifically resolves RNA:DNA hybrids (Santos-Pereira et al., 2013; Wahba et al., 2011). R-loops can also

form at sites of TERRA transcription, as has been shown through analyses of mutations affecting another THO complex component, *THP2* (Pfeiffer et al., 2013). These telomeric r-loops can affect the regulation of telomerase, expose telomeres to increased nucleolytic degradation, and form structures that promote homologous recombination (Aguilera and García-Muse, 2013; Arora et al., 2014; Balk et al., 2013; Maicher et al., 2014; Pfeiffer and Lingner, 2012).

The Paf1 complex is known to play a role in transcription elongation and 3' end mRNA maturation (Nordick et al., 2008; Sheldon et al., 2005), and recent evidence has also implicated the PAF1 Complex in transcription-induced genome instability; the Koshland laboratory showed that *cdc73Δ* and *leo1Δ* cause a ~8 fold increase in yeast artificial chromosome (YAC) instability that could be reversed by RNase H1 overexpression (Wahba et al., 2011). We demonstrate in Chapter 2 that deletion of *CDC73* results in telomere defects that synergize with deletion of *TEL1* and *YKU80*, and we find that overexpression of RNase H1 or *TLC1* substantially decrease the GCR rates in these cells, suggesting *CDC73* suppresses genome instability by playing a role in both telomere maintenance and r-loop formation. In Chapter 3, we expand our analysis to test what role the other members of the Paf1 Complex play in the suppression of genome instability, and identify a minimal region of Cdc73 that is necessary for its function, providing clues about how the complex as a whole is assembled.

Dissertation Organization

The first chapter is an adaptation of a manuscript that has been submitted for publication. The authorship and title are as follows:

> Putnam CD, Srivatsan A, Nene RV, Martinez SL, Clotfelter SP, Bell SN, Somach S, de Souza JES, Fonseca AF, de Souza SJ, Kolodner RD. An extended genetic network suppresses genome rearrangements in *S. cerevisiae*.

The manuscript describes a large scale screen to identify all the genes in yeast that contribute to the suppression of genome instability. The work presented in this chapter was done during the first half of my graduate training, in collaboration with Christopher S. Putnam and Anjana Srivatsan, who designed and conducted many of the experiments in this chapter. I acknowledge their essential and leading contribution to this work and thank them for their guidance and direction. The project was overseen by Richard D. Kolodner, who provided invaluable support during all stages of the project. We also received tremendous help and assistance from Sandra L. Martinez, Sarah P. Clotfelter, and Sara N. Bell. We also acknowledge the contribution of Sandro J. de Souza, Jorge E.S. de Souza, André F. Fonseca, and Steven Somach, who collaborated on the analysis of mutations of the human homologs in cancers.

Among the 65 novel suppressors of genome instability identified in Chapter 1 was *CDC73*, a transcription factor whose human homolog is a tumor suppressor commonly mutated in head and neck cancers. The study of this gene is the primary focus for the rest of the dissertation. Chapter 2 is an original document that is being prepared for publication and deals specifically with the characterization of *CDC73* as a suppressor of

genome instability. As of May 19th 2015, the authorship and tentative title of this manuscript are:

> Nene RV, Putnam CD, Campbell CS, Desai A, Kolodner RD. Cdc73 Suppresses Genome Instability by Mediating Telomere Homeostasis and RNA:DNA Hybrid Formation.

The dissertation author was the primary author of this material and contributed to the conception and design of experiments, execution of experiments, data analysis, and manuscript writing. The dissertation author would like to thank Christopher S. Putnam and Richard D. Kolodner for assisting with the conception and design of the research, data analysis, manuscript writing and mentorship.

As mentioned above, Cdc73 is a subunit of the five protein Paf1 complex. Chapter 3 is an original document that is being prepared for publication and investigates the role that the Paf1 complex as a whole plays in suppressing genome instability and demonstrates how a structural analysis of Cdc73 can contribute to the understanding of how complex assembly affects this function. As of May 19th 2015, the authorship and tentative title of this manuscript are:

> Nene RV, Putnam CD, Campbell CS, Desai A, Kolodner RD. Identification of a Domain of Cdc73 That is Necessary and Sufficient for the Suppression of Genome Instability.

The dissertation author was the primary author of this material and contributed to the conception and design of experiments, execution of experiments, data analysis, and manuscript writing. The dissertation author would like to thank Christopher S. Campbell for conducting the confocal microscopy and image analysis. The dissertation author

would also like to thank Christopher S. Putnam and Richard D. Kolodner for assisting with the conception and design of the research, data analysis, manuscript writing and mentorship.

Chapter 1:

An extended genetic network suppresses genome rearrangements in *S. cerevisiae*

By

Christopher D. Putnam[1,2*], Anjana Srivatsan[1*], Rahul V. Nene[1*], Sandra L. Martinez[1*], Sarah P. Clotfelter[1], Sara N. Bell[1], Steven Somach[1#], Jorge E.S. de Souza[6-7], André F. Fonseca[8], Sandro J. de Souza[8,] and Richard D. Kolodner[1,3,4,5]

From

Ludwig Institute for Cancer Research[1], Departments of Medicine[2] and Cellular and Molecular Medicine[3], Moores-UCSD Cancer Center[4] and Institute of Genomic Medicine[5], University of California School of Medicine, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0669
and
Instituto de Bioinformática e Biotecnologia[6] , Instituto Metrópole Digital – UFRN[7], and Instituto do Cérebro – UFRN[8], Natal, Brazil.

*, Co-first authors

ABSTRACT

Gross Chromosomal Rearrangements (GCRs) play an important role in human diseases, including cancer. The identity of all Genome Instability Suppressing (GIS) genes is not currently known. Here multiple *Saccharomyces cerevisiae* GCR assays and query mutations were crossed into arrays of mutants to identify progeny with increased GCR rates. This resulted in the identification of 183 GIS genes in which mutations cause increased GCR rates. In addition, 438 cooperatively acting GIS genes were identified in which mutations did not cause increased GCR rates but which enhanced the genome instability caused by individual query mutations. Analysis of the TCGA data using the human genes predicted to act in the pathways implicated by the 183 *S. cerevisiae* GIS genes for loss-of-function mutations, copy number changes with reduced expression and silencing demonstrated that a minimum of 93% of ovarian and 66% of colorectal cancer cases had defects affecting one or more predicted GIS genes.

INTRODUCTION

Genetic instability is seen in most cancers and is thought to play a critical role in the development and progression of tumors (Loeb, 2001). There are two general types of genetic instability seen in cancer (Vogelstein et al., 2013). One type of instability is the accumulation of large numbers of mutations due to environmental mutagens, defects in DNA mismatch repair (MMR) genes, defects in the fidelity of DNA polymerases, and possibly defects in the regulation of cytosine deaminases (de la Chapelle, 2004; Lawrence et al., 2013; Loeb and Harris, 2008; Nik-Zainal et al., 2012; Palles et al., 2013). The other type of instability is the accumulation of genome rearrangements such as translocations, copy number changes and aneuploidy (Inaki and Liu, 2012; Vogelstein et al., 2013). Our understanding of the genes that suppress genome rearrangements in cancer comes from the study of inherited defects causing cancer susceptibility syndromes including Fanconi Anemia and the *BRCA1*- and *BRCA2*-defective breast and ovarian cancer syndromes (D'Andrea, 2010; Kobayashi et al., 2013). However, our understanding of the causes of genome rearrangements is incomplete in part because most relevant studies have focused on a limited number of candidate genes and because it has not been possible to perform unbiased genetic screens to identify Genome Instability Suppressing (GIS) genes in mammalian cells.

Genetic studies in *Saccharomyces cerevisiae* have provided considerable information about the spontaneous formation of genome rearrangements. Such studies have been made possible by the development of quantitative genetic assays that can detect gross chromosomal rearrangements (GCRs) (Chan and Kolodner, 2011; Chen and

Kolodner, 1999; Hackett et al., 2001; Kanellis et al., 2007; Myung et al., 2001a; Putnam et al., 2009). The types of GCRs that have been observed depend in part on the features of the specific GCR assay, but include 1) terminal deletions healed by *de novo* telomere addition, 2) monocentric translocations that can be mediated by different types of re-joining of broken chromosomes or non-allelic recombination between repeated sequences, 3) interstitial deletions, 4) complex translocations initially mediated by inverted repeat formation at the site of a broken chromosome end or during the chromosome breakage process, 5) chromosome fusions and other types of dicentric translocation chromosomes involving joining of degraded telomeres, re-joining of broken chromosomes or non-allelic homologous recombination (HR) between repeated sequences, and 6) complex GCRs resulting from multiple cycles of rearrangement usually as a result of the formation of dicentric translocations (Chan and Kolodner, 2011, 2012; Chen and Kolodner, 1999; Pennaneach and Kolodner, 2009b; Putnam et al., 2004b, 2005, 2014; Schmidt and Kolodner, 2006). Overall, the types of genome rearrangements selected for in GCR assays parallel those being identified by whole genome analysis in human diseases including cancer. In addition, GCR assays have been used to identify genes in *S. cerevisiae* that prevent GCRs from occurring as well as alter the types of GCRs formed (Banerjee et al., 2008; Chan and Kolodner, 2011; Chen and Kolodner, 1999; Huang and Kolodner, 2005; Huang et al., 2003; Kanellis et al., 2007; Motegi et al., 2006; Myung et al., 2001a, 2001b; De Piccoli et al., 2006; Putnam et al., 2009, 2010, 2014; Schmidt and Kolodner, 2006; Smith et al., 2004; Stirling et al., 2011). These studies have shown that a combination of oxidative defense pathways, the replication

machinery, DNA repair and HR pathways including the Post Replication Repair (PRR) pathways, cell cycle checkpoint pathways, telomere maintenance pathways, RNA processing pathways and chromatin modification and assembly pathways function in concert to prevent GCRs.

Our knowledge of the genes and pathways that prevent and form GCRs in model organisms like *S. cerevisiae* is also incomplete. This is in part because most of the genes that suppress GCRs have been identified through limited candidate-gene approaches (Albuquerque et al., 2013; Chan and Kolodner, 2011; Chen and Kolodner, 1999; Myung et al., 2001a, 2001b; Putnam et al., 2009, 2010, 2014). Some studies have screened collections of arrayed *S. cerevisiae* mutants for mutations that cause increased GCR rates and have identified a small number of additional genes of interest (Huang et al., 2003; Kanellis et al., 2007; Smith et al., 2004; Stirling et al., 2011). Reasons for the limited success of these latter screens include: 1) that some of the screens did not use assays that are specific for GCRs, 2) that GCR assays are tedious and not well-suited for large scale screens resulting in the identification of only a few mutations that cause large increases in GCR rates, and 3) that not enough different GCR assays and interacting mutations were used in the studies reported. Here, we used a 2-stage screen design in which a bioinformatics approach was used to develop a highly enriched candidate gene list sorted into candidate pathways (Putnam et al., 2012) followed by an extensive genetic interaction screen utilizing three different GCR assays and 43 genetically diverse query mutations to screen for genes and interacting pairs of genes that act to suppress GCRs. Our results have provided a much more detailed picture of the genetic network that acts

to prevent GCRs than previously available, and analysis of The Cancer Genome Atlas (TCGA) data for ovarian and colorectal cancers (Ciriello et al., 2013; The Cancer Genome Atlas Research Network, 2012, 2013) has suggested that the genes in this network are potentially altered in a large proportion of ovarian and colorectal cancers.

RESULTS

Design of the systematic mating screen for mutations and genetic interactions causing increased genome instability. Our strategy for identifying mutations that cause increased genome instability involved using an adaptation of the Synthetic Genetic Array (SGA) method (Tong and Boone, 2006) to cross one of three GCR assays (Figure 1.1a) as well as 43 query mutations into a collection of candidate mutants followed by testing the mutants for increased accumulation of GCRs. The GCR assays were all based on the observation that when the *CAN1* and *URA3* genes are inserted on the nonessential left arm of chromosome V between the telomere and the first centromeric essential gene (*PCM1*), haploid cells containing GCRs that result in the loss of both *CAN1* and *URA*3 can be identified due to their resistance to both canavanine (Can) and 5-fluoroorotate (5FOA). The first GCR assay contained single copy sequences between the *CAN1 URA3* cassette and *PCM1*, as well as a short repeated sequence (sGCR assay) due to ~100 bp of *YCLWdelta5* in the *can1::*P_{LEU2}*-NAT* locus that has homology to the long-terminal repeats from Ty1 and Ty2 retrotransposons present in the genome. The second GCR assay contained the ~4 kb *DSF1-HXT13* segmental duplication in the interval between the *CAN1 URA3* cassette and *PCM1* (dGCR assay) in addition to the *YCLWdelta5*

fragment. The third GCR assay contained a full-length copy of the ~5 kb Ty912 sequence (tyGCR assay) between a *CAN1 URA3* containing region and *PCM1*. Using this strategy, we screened 1,056 mutant strains selected based on the 1,041 *S. cerevisiae* genes identified by an *in silico* screen for genes likely to be involved in suppressing genome stability (Figure 1.1b) (Putnam et al., 2012) and 16 additional genes that act in the pathways identified by the *in silico* screen.

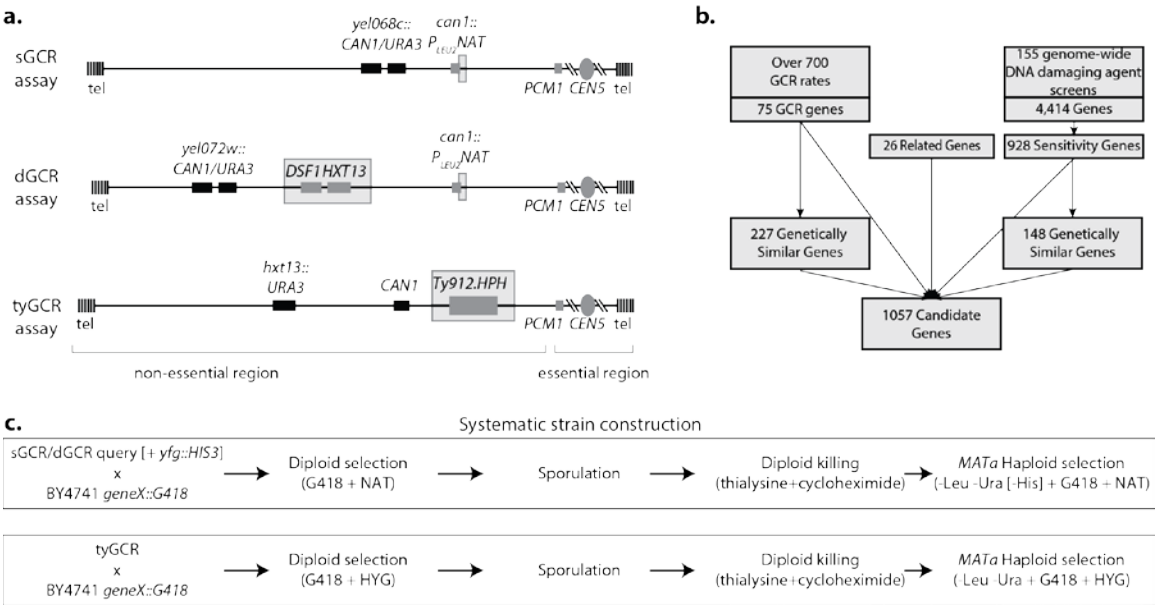


Figure 1.1 Details of the systematic screen.
a. The sGCR, dGCR, and tyGCR assays involve selection against the *CAN1* and *URA3* genes inserted into the terminal non-essential region of the left arm of chromosome V. For the sGCR assay, the only sequence with homology to other regions of the genome in the breakpoint region (between *CAN1/URA3* and *PCM1*) is a ~100 bp fragment of *YCLWdelta5* introduced by *can1::*P_{LEU2}*-NAT* (grey outline). The dGCR assay can form rearrangements mediated by non-allelic HR between the *DSF1/HXT13* segmental duplication (grey outline) and regions on chromosomes IV, X, and XIV and in addition contains the *YCLWdelta5* fragment in *can1::*P_{LEU2}*-NAT*. The tyGCR assay can form rearrangements mediated by non-allelic HR between the inserted Ty912 element (grey outline) and other Ty-related elements throughout the genome. **b.** Summary of how the 1,057 mutations in the first generation bait strain collection were identified. **c.** Summary of the procedure for crossing the query strains containing the GCR assays and query mutations with bait strains to obtain haploid strains containing GCR assays and deletions of interest for further analysis.

Generation and analysis of single mutant strains. Wild-type query strains containing each GCR assay were crossed to the first-generation set of verified mutant strains and haploid *MATa* progeny containing a GCR assay and the relevant mutation of interest were selected (Figure 1.1c). Crossing the dGCR, sGCR, and tyGCR query strains to the deletion mutants generated 1,002, 995, and 1,009 single mutant strains,

respectively. GCR strain scores for all strains were generated based on the number of Can^R $5FOA^R$ papilla observed after replica plating independent patches onto GCR selection media (Figure 1.2a). The GCR strain scores for the *leu2Δ* control strains were 1, 94, and 2.67 for sGCR, dGCR, and tyGCR assays, respectively. These scores were consistent with the results of GCR rate measurements and with the distribution of GCR strain scores for all of the single mutant strains generated, which peaked around the GCR strain scores for the *leu2Δ* control strains, suggesting that most mutations did not strongly affect the GCR rate as single mutations (Figure 1.2b-d).

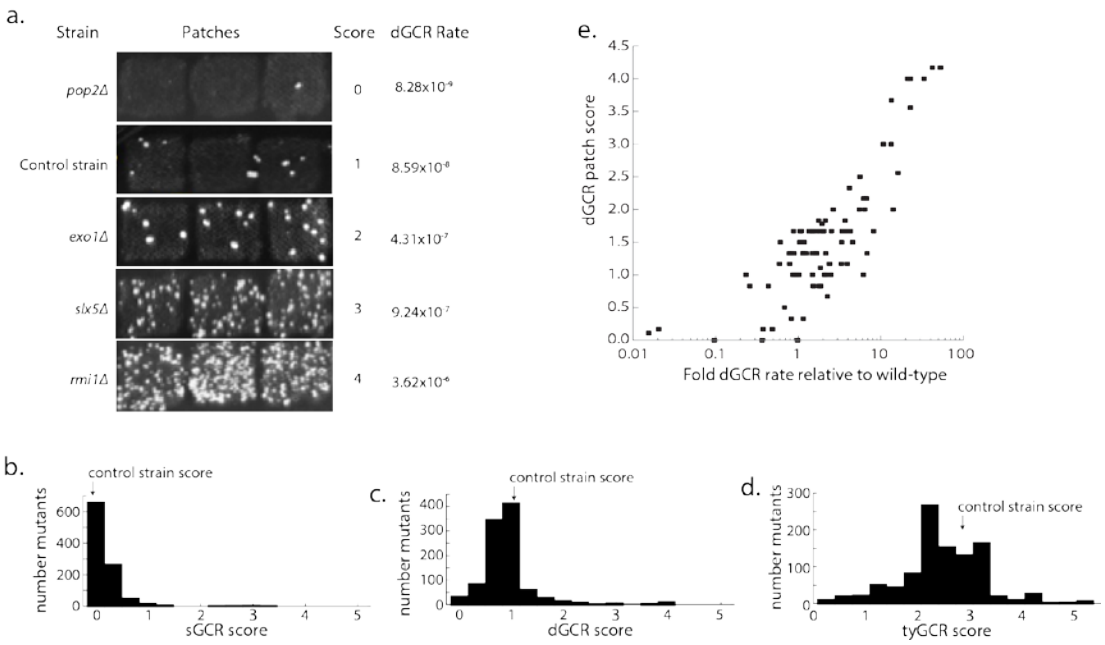


Figure 1.2. Assaying single mutant strains using a semi-quantitative GCR strain score. **a.** The semi-quantitative scoring strategy assigns a number to each patch of between 0 and 5 depending on the number of papilla. A score of 1 was assigned to the number of papilla in the *leu2Δ* control strain in the dGCR assay. For each strain, individual GCR patch scores were averaged to calculate the GCR strain score. Increases in the GCR strain score were paralleled by increases in GCR rates measured by the fluctuation method. **b.-d.** Histograms of GCR strain scores for single mutant strains in the sGCR (panel b), dGCR (panel c), and tyGCR (panel d) assays reveal that the average GCR strain score increases with the GCR rate for each GCR assay and that the score of the *leu2Δ* control strain generally lies at the peak of each histogram, suggesting that many of the mutations do not substantially affect the GCR rate as single mutations. **e.** The fold increase in the quantitative GCR rate is correlated with the GCR strain score for systematically generated strains containing the dGCR assay.

To determine a cutoff score to identify mutations causing increased GCR rates, we determined GCR rates for 101 strains resulting from the cross with the wild-type dGCR assay query strain and 43 strains resulting from crosses of mutant dGCR query strains with the *leu2Δ* control strain (see below). Strains chosen for rate measurements either had high GCR strain scores and/or contained single mutations that had not been

previously studied. For mutations that had been previously studied in strains constructed by transformation-mediated gene deletion and traditional crosses, we found a small but consistent increase in dGCR rates in strains derived from the systematic crosses: the wild-type *leu2Δ* control strain had a 4.4-fold increase in GCR rate, and 48 systematically generated single mutant strains had an average of a 3.0-fold increase. This small increase in GCR rate could have been due to the *YCLWdelta5* fragment (Figure 1.1); however, the 21-fold increase of the GCR rate of the *leu2Δ* dGCR strain relative to that of the *leu2Δ* sGCR strain indicated that most GCRs detected by the dGCR assay were mediated by the *HXT13-DSF1* region and not the *YCLWdelta5* fragment. Despite this small increase in the dGCR rate, we found a robust correlation between the GCR strain scores from patches and the GCR rates determined by fluctuation analyses (Figure 1.2e). Using the dGCR rates of the 144 systematically generated strains, we determined that a cutoff score of 1.4 (4 above the wild-type score) balanced the false-positive and false-negative errors in when GCR strain scores to identify mutations that caused an accumulation of GCRs.

We identified the *S. cerevisiae* GIS genes by combining all previously known and newly identified genes in which mutations caused increased accumulation of GCRs. Initially, we selected all single mutations that caused GCR strain scores that were 4 or more above the wild-type score in any GCR assay. We then eliminated false-positive and false-negative mutations from this list identified by GCR rate measurements. Finally, we included previously identified mutations that increased the GCR rate by 3-fold or more, including essential genes not studied here and genes identified in studies using GCR assays lacking any duplication in the GCR breakpoint region (single-copy or unique

sequence GCR assays; uGCR assays). Together, these data implicate 74 genes as suppressing GCRs in the dGCR assay, 72 genes in the tyGCR assay, 81 genes in the sGCR assay, and 105 genes in the uGCR assays. The higher number of suppressing genes implicated in the uGCR assay primarily reflects the identification of genes in candidate gene studies in which mutations cause small but significant increases in quantitative GCR assays that were too small to reliably detect by the semi-quantitative scoring method used here. Together, these data implicate 183 *S. cerevisiae* genes as GIS genes; 51 of which suppress genome instability in at least 3 of the 4 GCR assays (Figure 1.3).

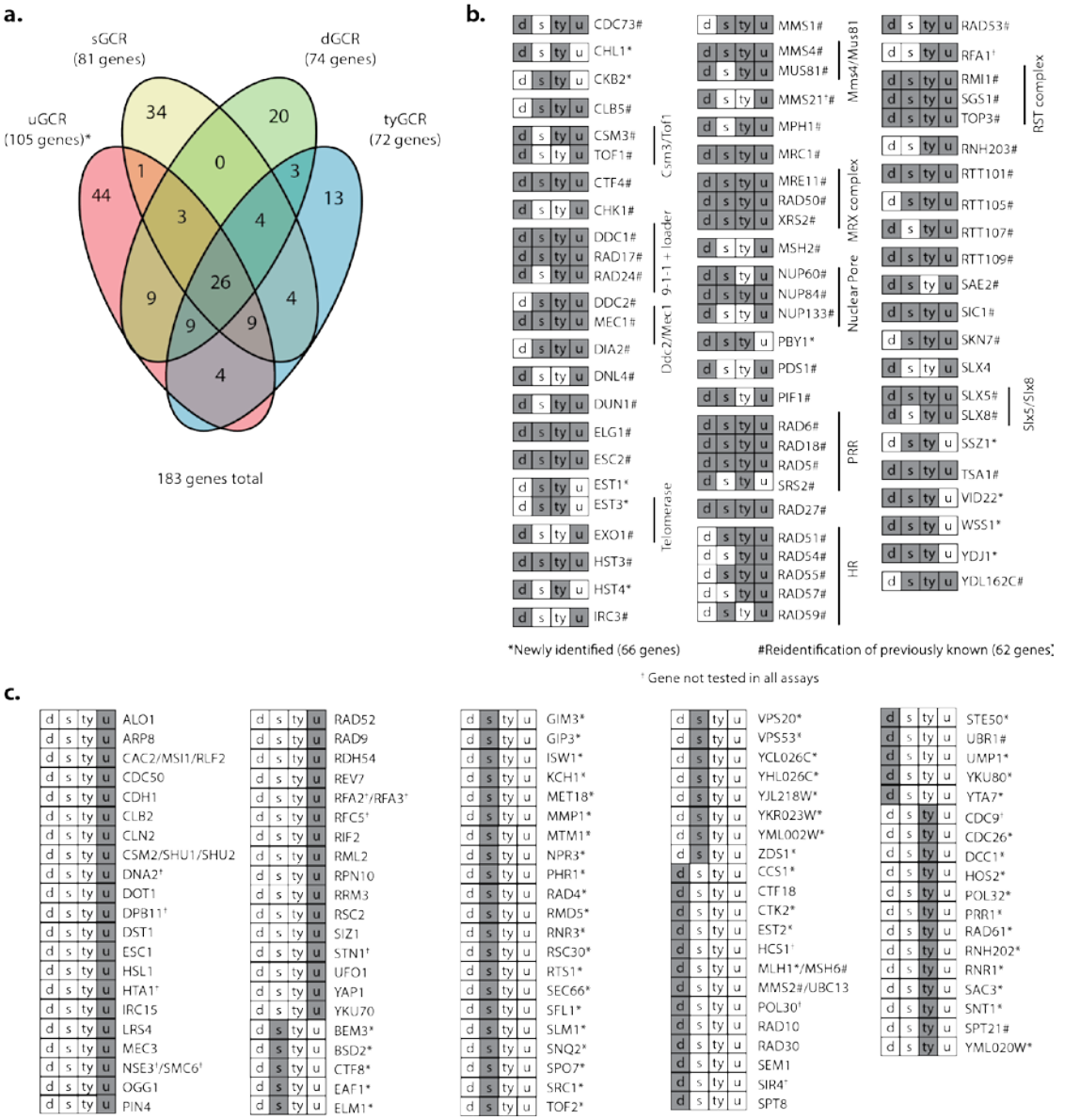


Figure 1.3. Summary of the increased GCR rates of single mutant strains identified using patch tests.

a. Venn diagram indicating the number of genes that suppress GCRs in each of the GCR assays used. **b.** Genes implicated in suppressing GCRs in more than one GCR assay. The boxes indicate which assays (d=dGCR, s=sGCR, ty=tyGCR, u=uGCR) in which the listed gene suppresses (grey) or does not suppress (white) GCRs. Marked genes were newly identified (*) or re-identified (#). Many genes unique to the uGCR assay are primarily genes in which mutations cause small but significant increases in GCR rates that were identified using fluctuation assays but are difficult to identify by the semi-quantitative patch score method used here. **c.** Genes implicated in suppressing GCRs in only one GCR assay, annotated as in panel b.

This analysis re-identified 62 previously known genes and 65 genes not previously identified as suppressing increased GCR rates. Of the 56 genes that were not re-identified in this screen, 13 were not discoverable, as these genes were either essential for viability or mating. Mutations in most of the remaining 43 genes caused only a small increase in GCR rates that could not be easily identified by patch scores; however, 42 of these 43 genes were subsequently re-identified in our GCR enhancer screen (see below). Examples of newly identified GCR suppressing genes included *PBY1*, *VID22*, *WSS1*, and *YDJ1*. *PBY1* encodes a tubulin-tyrosine ligase homolog that lacks this activity *in vitro* (Badin-Larcon et al., 2004) but localizes to RNA Processing Bodies (P-bodies), which form in a Mec1- or Tel1-dependent fashion under conditions of DNA damage (Sweet et al., 2007; Tkach et al., 2012). *WSS1* encodes a SUMO-dependent isopeptidase (Mullen et al., 2010), consistent with the observation that dysregulation of SUMOylation causes increased GCR rates in the dGCR assay (Albuquerque et al., 2013). *VID22* encodes a partner of the essential protein Tbf1 that insulates genes from telomeric silencing (Fourel et al., 1999), drives transcription of small nucleolar RNAs (Preti et al., 2010) and promotes the resection of DSBs and their repair by Non-homologous End Joining (Bonetti et al., 2013). And *YDJ1* encodes the major cytosolic Hsp40/DnaJ co-chaperone that plays important roles in protein maturation and stabilization (Sahi and Craig, 2007). The imperfect overlap of mutations causing increased GCR strain scores in the dGCR, tyGCR, and sGCR assays suggests the existence of mutations that can have different effects on genome stability in different GCR assays, which we verified by determining GCR rates.

To determine the efficiency of our bioinformatic strategy to pre-select candidate GIS genes, we crossed the dGCR assay to five randomly selected 96 well plates of mutant strains from the *S. cerevisiae* deletion collection and determined GCR strain scores for the progeny. Of the 463 single mutants scored, we found only one unexpected deletion, *ydl118wΔ*, that caused an increased GCR strain score (1.83); this deletion was present in our bioinformatically generated list of strains, but scores from the initial cross did not reveal an increased GCR strain score, most likely because this deletion causes only a small, borderline increase in GCR strain score. Extrapolating to the entire deletion collection, we estimate that the method potentially missed identifying ~8 GIS genes and that we identified 96% of the GIS genes. However, it should be noted that *ydl118wΔ* was identified in the enhancer mutation screen described below, which would be expected to identify weak but significant alleles as interacting mutations.

Identification of enhancer mutations using the dGCR assay. Some mutations can be observed to cause increased accumulation of GCRs only when they are combined with other mutations (Myung et al., 2001a). Thus, we screened for enhancer mutations by crossing our set of mutant strains against strains containing the dGCR assay and a query mutation. Using our previously generated hierarchical clustering of candidate GIS genes (Putnam et al., 2012), 43 query mutations were selected (Figure 1.4a). These query mutations broadly surveyed the genetic diversity within each of the gene clusters that had strong similarities to known GIS genes by sampling one or a small number of genes within each subcluster (Figure 1.4b). These query mutations were crossed to a second-generation set of mutant strains in which the number of bait mutations was reduced from

1,057 to 638 by eliminating bait mutations that did not affect GIS genes and did not synergistically interact with one of four query mutations, *dia2Δ*, *exo1Δ*, *rrm3Δ*, and *rtt107Δ*, that together interacted with most mutations from the 1,057 first-generation mutation collection.

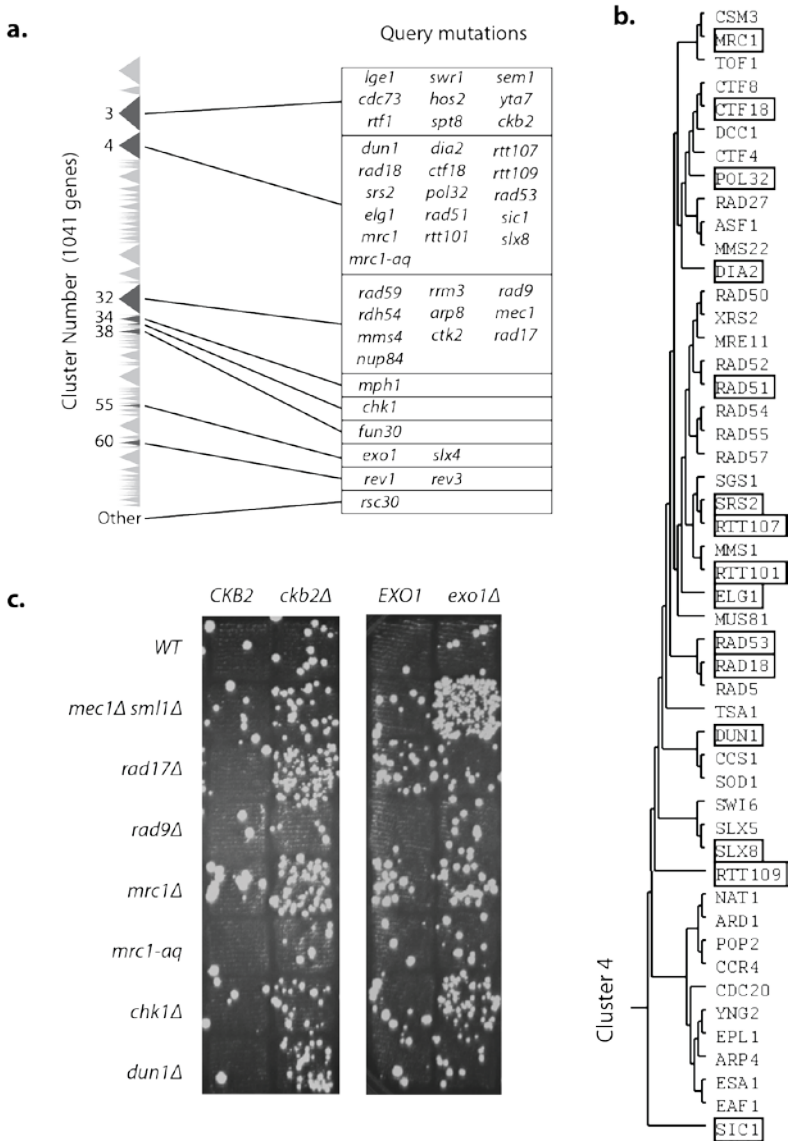
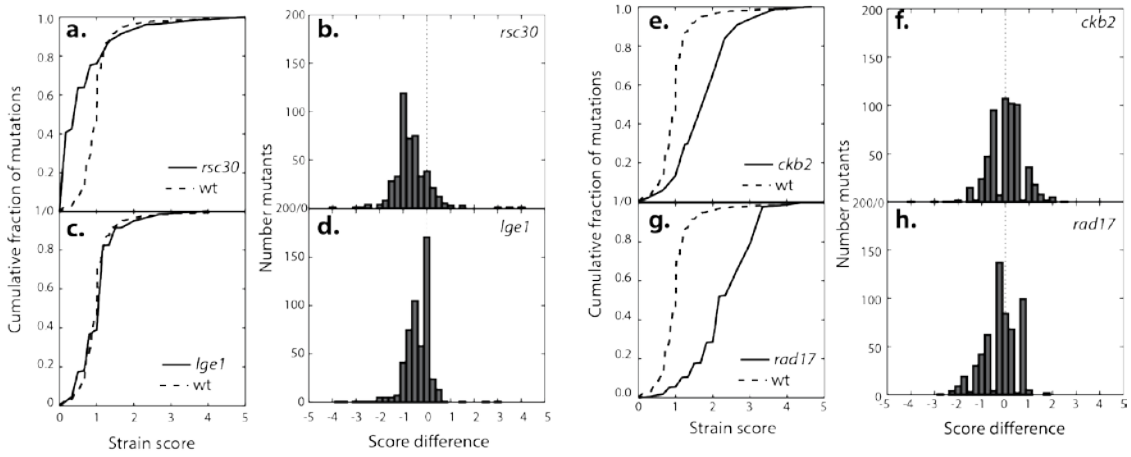
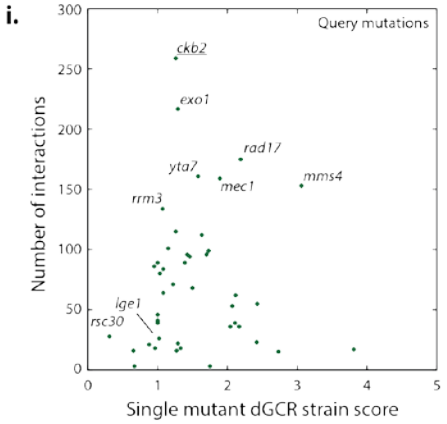


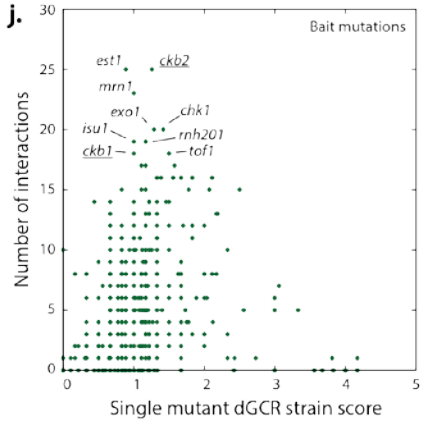
Figure 1.4. Selection of query mutations for use in analyzing genetic interactions.
a. The query mutations selected were primarily selected from the previously described gene clusters 3, 4, 32, 55, and 60 generated by clustering the candidate GCR suppression genes by genetic interactions (Putnam et al., 2012). Clusters 3, 4, and 32 had the greatest number of GCR suppressing genes. Triangles indicate the relative size of the cluster in terms of the number of genes and the darker triangles are the clusters from which query mutations were selected. **b.** Query mutations in non-essential genes selected from cluster 4 (indicated by the boxes) were selected to provide the greatest genetic diversity by picking 1 or 2 mutations from most sub-clusters in cluster 4. Mutations were similarly selected from clusters 3 and 32. **c.** Patch tests documenting genetic interactions involving mutations in either *CKB2* or *EXO1*. The status of *CKB2* or *EXO1* is indicated across the top of each set of patches and the bait mutations tested are indicated along the left side of each set of patches.

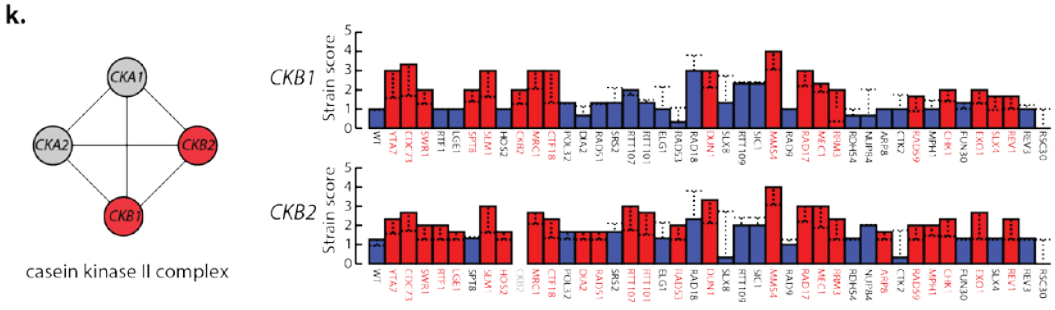
The 25,974 double mutants resulting from the crosses were then tested for increased GCR strain scores relative to those of the respective single mutants (Figure 1.4c). Because the 43 query mutations were also present in the set of bait mutations, there were 903 possible pairs of double mutant strains generated as query × bait or bait × query combinations. We obtained 801 of the 903 pairs, and the GCR strain scores measured for these pairs were highly consistent. The scores of the double mutant strains were distributed about the score of the query mutants very much like that seen in the analysis of the single mutant strains, including mutations causing reduced scores (e.g. *rsc30Δ*), scores essentially identical to wild-type (e.g. *lge1Δ*), or increased scores (e.g. *ckb2Δ* and *rad17Δ*) (Figure 1.5a-h). 3,149 (~13%) double mutant strains had a GCR strain score that was 4 or higher than either of the two single mutant scores. Determination of the GCR rates of selected double mutants revealed that increased double mutant GCR strain scores were a good indicator for synergistic interactions in the dGCR assay. Double mutant GCR strain scores identified many genetic interactions that had been previously identified (Myung et al., 2001a; Putnam et al., 2010), such as the redundancy between the *REV1-REV3-REV7*- and *MMS2-UBC13*-dependent branches of post-replication repair (PRR), the dependence of *rad18Δ* and *rad5Δ* increases in GCR rates on *SRS2*, and the redundancy of *MEC1*- and *TEL1*-mediated suppression of GCR formation as well as many new interacting mutations.

Figure 1.5 (See Next Page). Identification of genetic interactions involved in suppressing genome instability.
a, c, e, g. Plots of the cumulative fraction of mutations below specific strain scores for crosses with strains containing *rsc30Δ*, *lge1Δ*, *ckb2Δ*, and *rad17Δ* query mutations (solid line) compared with the distribution from the crosses to the wild-type strain (dashed line). **b, d, f, h.** Histograms of the number of mutations in combination with *rsc30Δ*, *lge1Δ*, *ckb2Δ*, or *rad17Δ* as a function of the score difference, which is the score of the double mutant strain (*aΔ bΔ*) minus the score of the higher of the two single mutant strains (*aΔ* or *bΔ*). **i.** Plot of the number of GCR-based interactions as a function of the single mutant GCR strain score for the 44 query strains including the wild-type strain and 43 mutant query strains. Query mutations with large numbers of interactions or those displayed in panels a-h are labeled. **j.** Plot of the number of GCR-based interactions as a function of the single mutant GCR strain score for bait mutations. Bait mutations with large numbers of interactions are labeled. **k.** Analysis of physical interaction data for the casein kinase II complex is shown (left) with reported physical interactions in BioGrid (lines) between complex components (circles). Components with known GCR interations are red; untested components (*CKA2*) or those tested with only 4 query mutations (*CKA1*) are in grey. Display of the genetic interactions between the *ckb1Δ*, and *ckb2Δ* bait mutations and the 43 query mutations (right). Bar heights indicate the strain score for the double mutant, and bar colors correspond to the presence (red) or absence (blue) of an increased level of genome instability in the double mutant as indicated in patch tests; the horizontal dashed line corresponds to the GCR strain score of the higher of the two single mutations. Missing bars and query names in grey correspond to double mutant strains that were not generated.


a.
b.
rsc30
c.
d.
lge1
wt
Cumulative fraction of mutations
Number mutants
Strain score
Score difference
e.
f.
ckb2
g.
h.
rad17


i.
Query mutations
ckb2
exo1
rad17
yta7
mec1
mms4
rrm3
lge1
rsc30
Number of interactions
Single mutant dGCR strain score


j.
Bait mutations
est1
ckb2
mrn1
exo1
chk1
isu1
rnh201
ckb1
tof1
Number of interactions
Single mutant dGCR strain score


k.
CKA1
CKA2
CKB2
CKB1
casein kinase II complex
CKB1
CKB2
Strain score

The query mutations with the largest number of interacting mutations in the interaction dataset were *ckb2Δ*, *exo1Δ*, *rad17Δ*, *yta7Δ*, *mec1Δ*, *mms4Δ*, and *rrm3Δ* (Figure 1.5i), and the bait mutations that interacted with the largest number of query mutations were *est1Δ*, *ckb2Δ*, *mrn1Δ*, *exo1Δ*, *chk1Δ*, *isu1Δ*, *rnh201Δ*, *ckb1Δ*, and *tof1Δ* (Figure 1.5j). Two mutations illustrating the complexity of these interactions were *ckb2Δ* and *exo1Δ,* which both interacted with checkpoint defects and also interacted with each other, indicating that casein kinase II and Exo1 function in different GCR suppressing pathways. Mutations causing very high (>3) GCR strain scores as single mutations tended to have fewer interactions than those causing lower GCR strain scores; one possible explanation for this is that these single mutations come close to saturating the assay making interactions that increase genomic instability difficult to detect by patch scores. In total, we identified 595 mutations as interacting with at least one of the query mutations; 438 of these mutations were distinct from mutations in the 183 GIS genes and hence were termed Cooperating Genome Instability Suppressing (CGIS) genes. In total, mutations in 621 genes (183 GIS genes and 438 CGIS genes; 13% of the 4,848 non-essential *S. cerevisiae* ORFs) were identified as causing or enhancing genome instability.

To focus on the most robust interactions, we identified shared interactions between a query mutation and mutations in genes that encode components of an annotated complex or pathway (Figure 1.5k). We identified 77 complexes and pathways where mutations in multiple genes encoding these complexes or pathways shared interactions with query mutations (Table 1.1) and an additional 91 complexes and pathways where a mutation in only a single gene was involved in interactions with query

mutations or where mutations in multiple genes were involved in interactions with query mutations but not with the same query mutations. The latter category includes complexes such as the Elg1-Rfc2-5 complex where the *elg1Δ* mutation is an enhancer mutation whereas the other genes in the complex were essential and could not be analyzed. We observed only 2 complexes, Mre11-Rad50-Xrs2 and Sgs1-Top3-Rmi1, in which mutations caused significant increases in GCR rates but lacked any interactions; however, single mutations affecting these complexes caused high GCR strain scores (~4.0) making it difficult to identify interactions with these mutations.

Table 1.1. Complexes with shared interactions in the dGCR enhancer screen.

Process	**Complexes and pathways***
DNA repair	core mitotic homologous recombination, Slx1-Slx4 complex, Mms4-Mus81 complex, base excision repair, DNA ligase IV complex, Ku complex, ribonuclease H2 complex, Shu complex, Msh2-Msh6 complex, Mlh1-Pms1 complex, Mlh1-Mlh3 complex, Mlh1-Mlh2 complex, Cul8-RING ubiquitin ligase complex, nucleotide-excision repair factor 1 (NEF1) complex, Rad1-Rad10-Saw1 complex, nucleotide-excision repair factor 4 complex, DNA polymerase zeta – Rev1 complex, Mms2-Ubc13 complex, Shu complex
DNA replication	DNA polymerase epsilon complex, telomerase, ribonucleoside-diphosphate reductase complex
Chromosome cohesion and segregation	Ctf18 RFC-like complex, Ctf19 complex (includes COMA complex), monopolin, prefoldin complex, dynactin complex, Msh4-Msh5 complex
Cell cycle checkpoints	Mec1-Ddc2 complex, Tof1-Csm3 complex, Rad17-Ddc1-Mec3 complex + Rad24-Rfc2-5 clamploader, protein phosphatase (PP4) complex, spindle checkpoint, Cdc28 cyclin-dependent kinase complexes, anaphase-promoting complex (APC/C)
Chromatin/Transcription/ mRNA processing	Swr1 complex, SAGA complex, Rpd3L complex, SLIK (SAGA-like) complex, cytoplasmic mRNA processing body, NuA4 histone acetyltransferase complex, mediator complex, HIR complex, Rpd3S complex, Set3C complex, COMPASS complex, carboxy-terminal domain protein kinase complex, cytoplasmic Sm-like complex, Elogin-Cullin-Socs (ECS) ligase complex, NuA3 histone acetyltransferase complex, chromatin assembly complex, Chz1-Htz1-Htb1 complex, Cdc73/Paf1 complex, RSC complex, Ino80 complex, RNA polymerase I complex, Spt3-Spt8 SAGA subunit of SAGA complex, ISW1a chromatin remodeling complex, CCR4-NOT core complex, U6 snRNP
Nuclear pore	nuclear pore outer ring, nuclear pore nuclear basket
Proteasome/Protein degradation	proteasome 19/22S regulator, 20 proteasome + Ump1 chaperone, Doa10 ubiquitin ligase complex, Hrd1p ubiquitin ligase ERAD-L complex, Ula-Uba3 complex, Rad6-Ubr1 complex
Other	Kel1-Kel2 complex, Sod1-Ccs1 complex, Golgi transport complex, HMC complex, NatA complex, Chs5p/Arf-1 binding proteins (ChAPs), Ssk1-Ssk2 complex, AP-3 adaptor complex, ESCRT III complex, casein kinase II complex

Inactivation of GIS genes in ovarian cancer, colorectal cancer, and acute myeloid leukemia. To determine if defects in GIS genes might occur in cancer, the TCGA data for ovarian cancer, colorectal cancer and acute myeloid leukemia (AML) were analyzed (Ciriello et al., 2013; The Cancer Genome Atlas Research Network, 2012, 2013). The genes of interest were the human homologs of the 183 *S. cerevisiae* GIS genes plus 13 additional genes that act in pathways and protein complexes defined by the GIS genes (hGIS1, 219 genes) and an expanded list of genes (hGIS2, 284 genes) that also included human DNA repair genes that function in pathways identified in *S. cerevisiae* but do not have an *S. cerevisiae* homolog (e.g., *BRCA1* and *BRCA2*) or have an *S. cerevisiae* homolog that was not identified in the initial screen because of a borderline score (e.g., *NHEJ1* and *H2AFX*).

To identify potential cancer genes, we used a scoring system (S-score) recently developed by us (de Souza et al., 2014) that integrates in an unbiased way genome-wide data (copy number variation, expression, methylation and mutations) generated from a set of tumor samples. In the first analysis, human GIS genes were analyzed for signatures consistent with tumor suppressors (S-scores $\leq$ -2) or for signatures consistent with proto-oncogenes (S-scores $\geq$ 2). Genes from hGIS1 and hGIS2 with S-scores $\leq$ -2 were significantly enriched in ovarian cancer cases (hGIS1, 27 genes, p=0004; hGIS2, 42 genes, p<0001). In contrast, there was no enrichment in hGIS1 or hGIS2 genes with S-scores $\geq$ 2 in the ovarian cancer cases (hGIS1, 43 genes, p=37; hGIS2, 54 genes, p=47). The 42 genes from hGIS2 with S-scores $\leq$ -2 and 4 additional genes with S-scores between -2 and -1.95 in ovarian cancer cases were analyzed for reduced copy number

(GISTIC scores of -1 or -2) associated with reduced expression relative to controls (Z scores < -2). Reduced copy number associated with reduced expression of 1 to 19 of these 46 genes was observed in 97% of 527 ovarian cancer cases. A box plot of the data for one such gene, *RAD17*, and the frequency of occurrence for the top 20 such genes in ovarian cancer are shown in Figures 1.6a & 1.6b. There were also 3 genes that appeared to be silenced in 12% of 537 ovarian cancer cases. For the colorectal cancer cases, genes with S-scores ≤ -2 were significantly enriched (hGIS1, 18 genes, p=0001; hGIS2, 18 genes. P=0014). In contrast, hGIS1 genes with S-scores ≥ 2 were not significantly enriched in both hGIS1 (12 genes, p=12), and hGIS2 (16 genes, p=071) and were not studied further. The 18 genes with S-scores ≤-2 and 2 genes with S-scores between -2 and -1.95 in colorectal cancer cases were further analyzed. Reduced copy number associated with reduced expression of from 1 to 8 of these 20 genes was observed in 54% of 456 colorectal cancer cases, and 4 genes had apparent silencing in 10% of 463 colorectal cancer cases. In the case of AML, there was no enrichment of genes with S-scores ≤ -2 in the hGIS1 (p=071) and a marginal significance in hGIS2 (p=047). There was no enrichment for genes with S-scores ≥ 2 in both hGIS1 (p=084) and hGIS2 (p=205) nor were any genes with apparent silencing identified.

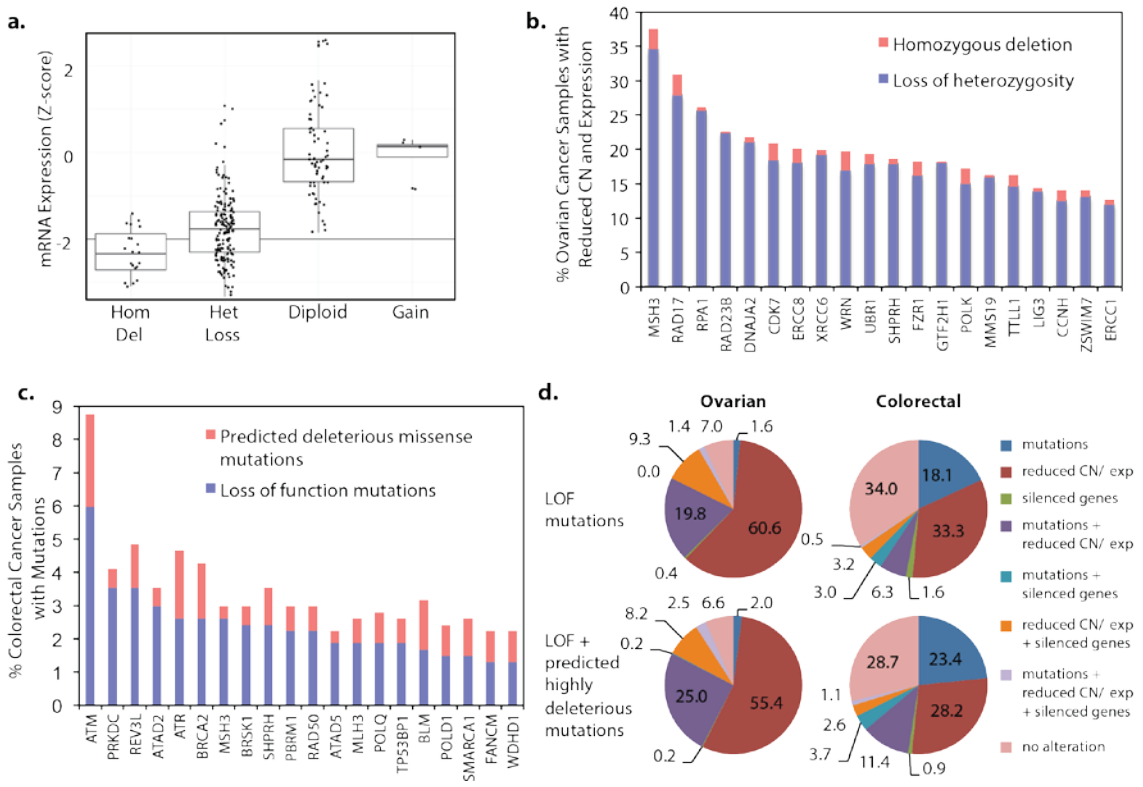


Figure 1.6. Analysis of the ovarian and colorectal cancer TCGA data for alterations in GIS genes.

a. Box plot of the RNA Seq data for the copy number (GISTIC -2, Homozygous Deletion; GISTIC -1, Heterozygous Loss; GISTIC 0, Diploid; GISTIC 1, Gain) vs. the Z-score for mRNA expression of *RAD17* in ovarian cancer. **b.** Histogram of the frequency of reduced copy number with reduced mRNA expression for the top 20 most altered GIS genes in ovarian cancer. **c.** Histogram of the frequency of mutations in the top 20 most altered GIS genes in colorectal cancer. Data for *MSH2*, *MSH6* and *MLH1* were excluded as defects in these genes predominantly cause rates of accumulation of point mutations. Predicted deleterious missense mutations are those that scored as deleterious in 5 or 6 out of 6 functional prediction tests. **d.** Pie charts showing the % of ovarian (left) and colorectal (right) cancer samples with different combinations of mutations, reduced copy number with reduced expression and silencing among all samples for which any type of genomics data were available. Analysis of LOF mutations alone (Top) and LOF + predicted deleterious missense mutations (Bottom) are presented separately. Note that 19% of the ovarian and 25% of the colorectal cancer cases were not analyzed for all types of potential alterations, and consequently the values presented are an underestimate.

In the second analysis, the number of potential loss-of-function (LOF) mutations (nonsense mutations, frameshift insertion/deletions, in-frame insertion/deletions, and splice site mutations) in the hGIS1 and hGIS2 genes was tabulated for 476 ovarian cancer cases and 537 colorectal cancer cases. For ovarian cancer, LOF mutations were not significantly enriched for the hGIS1 genes (p=78) but were enriched for hGIS2 genes (p<0001); this increase in significance was due to the presence *BRCA1* and *BRCA2* in the hGIS2 gene list, which accounted for 70% of the LOF mutations in hGIS2 genes.. Analysis of the enrichment of classes of the LOF mutations for the hGIS2 genes revealed that deletions (includes frame shift deletions; p<0001), insertions (includes frame shift insertions; p<0001), frame shift deletions (p<0001), frame shift insertions (p<0001) and nonsense mutations (p=0003) were present at statistically significant increased levels, many, but not all of which were in *BRCA1* or *BRCA2*. Overall, 27% of the 476 ovarian cancer samples had LOF mutations in at least 1 of 44 predicted human GIS genes with 1-3 genes mutated per sample. In contrast, LOF mutations in both sets of GIS genes were significantly enriched in colorectal cancer cases (hGIS1, p<0001; hGIS2, p=0016). The frequency of LOF and predicted deleterious missense mutations for the top 20 hGIS2 genes in colorectal cancer is shown in Figure 1.6c. Deletions (includes frame shift deletions; hGIS1, p=001; hGIS2, p<0068), mononucleotide repeat frame shifts (hGIS1, p<0001; hGIS2, p<0001) and splice site mutations (hGIS1, p=0006; hGIS2, p=0001) were present at statistically significant increased levels in both hGIS1 and hGIS2 genes, and nonsense mutation were present at statistically significant increased levels in hGIS1 and at borderline significant levels in hGIS2 (hGIS1, p=0098; hGIS2, p=0636). A

proportion of colorectal cancer has MMR defects that are associated with high rates of accumulating mutations and particularly frameshift mutations (The Cancer Genome Atlas Research Network, 2012). We therefore repeated the analysis using a sample set in which the MMR defective cases had been excluded and found that deletions (includes frame shift deletions; hGIS1, p=0007; hGIS2, p<0001), nonsense mutations (hGIS1, p=0006; hGIS2, p=003), frame shift deletions (hGIS1, p=033; hGIS2, p=039), mononucleotide repeat frame shifts (hGIS1, p=0019; hGIS2, p=0031) and splice site mutations (hGIS1, p=0007; hGIS2, p<0001) were present at statistically significant increased levels in both hGIS1 and hGIS2 and frame shift insertions were present at statistically significant increased levels in only hGIS2 (p=034). This indicates that the accumulation of these classes of mutations in the colorectal cancer cases was not due to the mutator phenotype caused by MMR defects. Overall, 30% of the 537 colorectal cancer samples had loss of function mutations in at least 1 of 199 predicted human GIS genes with 1-37 genes mutated per sample. In the case of AML, there was no enrichment of LOF mutations in the GIS genes (hGIS1, p=96; hGIS2, p=99) and, as a result, individual classes of mutations were not analyzed.

All of the gene inactivation data were merged and the proportion of different classes of gene inactivation was determined (Figure 1.6d). Two groups of mutations were considered including only LOF mutations as well as LOF mutations plus those missense mutations that scored as “predicted deleterious” in at least 5 of 6 function prediction tests used. In ovarian cancer, the gene inactivation signature was dominated by cases with reduced copy number associated with reduced expression. Colorectal cancer showed a

different pattern with less overlap between the cases with mutations and the cases with reduced copy number associated with reduced expression. Overall, when only LOF mutations were considered, a minimum of approximately 93% of ovarian cancer cases and 66% of colorectal cancer cases had a signature of inactivation of one or more predicted GIS genes (Figure 1.6d), although these figures are an underestimate as not all samples were analyzed for all types of alterations. It should be noted that the colorectal cancer cases did include cases with alterations in MMR genes in which defects would be expected to cause strong MMR defects (*MSH2*, *MSH6*, *MSH1*, *PMS2*), including 46 cases when only LOF mutations were considered and 50 cases when LOF + predicted deleterious missense mutations were considered (19 of these cases had silencing of *MLH1*, 1 of which also had LOH of *MLH1*), all but 3 of which had alterations in other GIS genes. In the ovarian cancer cases, there were 23 cases of reduced copy number and reduced expression of *MLH1*, 3 cases with a LOF mutation in an MMR gene and 2 cases with a predicted deleterious missense mutation in an MMR gene; all of these cases had alterations affecting other GIS genes. This analysis indicates that potential MMR defects account for only a small fraction of the alterations affecting GIS genes.

DISCUSSION

Here we have developed methods to screen the *S. cerevisiae* systematic deletion collection to identify new GIS genes, identify genes that interact to suppress the formation of GCRs and identify candidate human genes for the analysis of cancer genomics data to identify potential GIS gene defects in human cancers. This analysis

identified 65 new *S. cerevisiae* GIS genes, including *PBY1*, *VID22*, *WSS1*, and *YDJ1* and increased the total number of known GIS genes to 183. Comparison of the GIS genes identified using different GCR assays revealed genes and pathways that only suppressed GCRs formed in specific genomic contexts. We also identified 595 enhancer mutations, 438 of which only increased the accumulation of GCRs when combined with a second mutation. Filtering the interactions between the genes using data on previously identified protein complexes and pathways robustly identified 77 complexes and pathways with many shared interactions and 91 other complexes and pathways in which interactions were not shared or were restricted to individual encoding genes. Analysis of the TCGA ovarian and colorectal cancer data (The Cancer Genome Atlas Research Network, 2011, 2012) showed that the human homologs of the *S. cerevisiae* GIS genes had a statistically significant enrichment for somatic mutations and/or changes in copy number leading to reduced expression in these 2 cancers. Overall, the majority of the cancer cases analyzed (a minimum of 93% of ovarian and 66% of colorectal cancer cases if only LOF mutations were considered) appeared to have defects affecting one or more genes that were homologs of the *S. cerevisiae* GIS genes or act in the pathways identified by the GIS genes. These results suggest that genetic or epigenetic changes giving rise to increased genome instability are likely common in some types of cancer, but that due to the large number of GIS genes, the defect signature for any single gene can be weak.

Almost half of the 183 *S. cerevisiae* GIS genes suppress the formation of GCRs detected in multiple GCR assays. The common pathways identified typically involve genes involved in DNA metabolism, including DNA replication and repair, and genes

involved in checkpoint signaling in response to DNA damage and replication errors. Some of these genes identified, such as *RAD27* and *TSA1*, likely function by directly suppressing the formation of DNA damage (Debrauwere et al., 2001; Huang and Kolodner, 2005). Other genes, such as those encoding the Mre11-Rad50-Xrs2 and Sgs1-Top3-Rmi1 complexes, likely act in the processing of the DNA damage generated by other mechanism (Mimitou and Symington, 2009), such as DNA replication errors. In contrast, there are a number of genes whose roles in suppressing genome instability is less clear, such as *PBY1*, *VID22*, *WSS1*, *YDJ1*, *SSZ1*, and *CKB2*. The fact that many GIS genes suppress GCRs detected in multiple assays that probe different genomic contexts indicates that these genes can suppress the formation of many types of GCRs (Chan and Kolodner, 2011; Myung et al., 2001a; Pennaneach and Kolodner, 2009b; Putnam et al., 2009, 2014). A notable exception to this general observation are *pif1* mutations that affect both the type and rate of GCRs formed by causing a defect in the suppression of *de novo* telomere additions that appears insensitive to genomic context (Myung et al., 2001b; Putnam et al., 2009; Schulz and Zakian, 1994). In addition to GIS genes that suppress GCRs detected in multiple assays, a number of mutations suppress GCRs detected by subsets of GCR assays (Figure 1.3). These genes include the *MSH2* and *MSH6* MMR genes, which are specific to the dGCR assay, *SSZ1* and *RAD51*, which suppress GCRs in only the tyGCR and sGCR assays, and *SAE2* and *MET18*, which are specific to the sGCR assay. In most cases, the mechanisms underlying this specificity is not yet understood; however, in the case of MMR genes, the heteroduplexes formed by non-allelic HR during the formation of duplication-mediated GCRs is likely to contain a higher density of

mispairs and hence be better recognized by MMR than heteroduplexes formed in the tyGCR and sGCR assays.

Mutations that enhance the accumulation of GCRs can in principle act in compensatory or parallel pathways or can have much more complicated relationships involving genes within pathways (Zinovyev et al., 2013). In addition, mutations can result in increased levels of DNA damage that can lead to GCRs when repair mechanisms are defective or are saturated by the increased levels of DNA damage. Many of the mutations showing genetic interactions, such as *exo1Δ*, inactivate GIS genes and also cause increased GCR rates as single mutations. Mutations like these could affect the response to normal levels of spontaneous DNA damage as well as DNA damage that is either induced in the absence of other pathways or is normally repaired in part by other pathways. In contrast, a number of enhancer mutations, such as *tel1Δ* cause no increase in GCR rates as single mutations (Myung et al., 2001a). These mutations may either cause damage that is efficiently repaired so long as the relevant repair mechanisms are functional and not overwhelmed by other sources of damage or are redundant with other pathways. Defects in the genes encoding complexes can show the same types of interactions, regardless of whether defects in all of the genes encoding a complex behave similarly (e.g., *RNH201, RNH202, RNH203* and *MMS2 UBC13*) or whether defects in only a subset of the genes encoding a complex have similar properties (e.g., *SPT3, SPT8* encoding part of SAGA).

Genomics studies have the potential to extend our understanding of the cause of genome instability in cancer. However, these studies have been limited by the need to

identify individual genes that are altered at higher than background frequencies and the lack of a comprehensive picture of human GIS genes that would support a pathway-based analysis. The studies presented here have provided a detailed view of *S. cerevisiae* GIS genes that has facilitated a pathway-based analysis of cancer genomics data through the study of the human homologs of GIS genes. The analysis presented primarily focused on ovarian and colorectal cancer, two cancers with genome instability that have been suggested to have different relative frequencies of copy number change and mutation driver alterations (Ciriello et al., 2013). We also analyzed the data available for AML, a cancer that is associated with low frequencies of mutations and low if any genome instability (Ciriello et al., 2013). In the case of the TCGA ovarian cancer data, 23% of the samples with any data had LOF mutations in GIS gene homologs with 65% of the samples with LOF mutations having LOF mutations in *BRCA1* and *BRCA2* as previously reported (The Cancer Genome Atlas Research Network, 2011); no other individual GIS gene homolog had a LOF mutation in more than 5 to 1% of the samples. In contrast, there was a high frequency of copy number alterations, including both copy number reductions and homozygous deletions, associated with reduced expression of GIS gene homologs in ovarian cancer. This included 17% of the samples that had homozygous deletions of 1 to 9 GIS gene homologs per sample, approximating the frequency of samples with LOF *BRCA1* or *BRCA2* mutations. In contrast, the colorectal cancer TCGA data showed a higher proportion of samples and GIS gene homologs with LOF mutations and a lower yet high proportion of samples and GIS gene homologs with copy number alterations associated with reduced expression. A minimum of 93% of the TCGA ovarian cancer

cases and 66% of the TCGA colorectal cancer cases had alterations affecting one or more GIS gene homologs with only 5% and 8% of the samples, respectively, having alterations in genes expected to cause a strong mismatch repair defect (*MSH2*, *MSH6*, *MLH1*, *PMS2*) and hence a mutator phenotype. Overall, these results suggest that a high prevalence of alterations in GIS genes can explain how genome stability is compromised in these two cancers. Consistent with this view, there was no evidence for significant alteration of GIS genes in AML, a cancer that is not associated with high levels of genome instability (Ciriello et al., 2013). Defects in some of the human genes identified here have been implicated in cancer (e.g., *BRCA1*, *BRCA2*, *ATM*, *BLM*, *REV3L*, *PBRM1*), and some of the genes have been associated with the suppression of genome instability (e.g., *WRN*, *BLM*, *ATM*, *ATR*, *BRCA1*, *BRCA2*) or with pathways thought to act in the suppression of genome instability (e.g., *RAD17*, *RAD50*, *XRCC6*, *TP53BP1*) (Ciccia and Elledge, 2010; Friedberg et al., 2006). Our functional studies in *S. cerevisiae* provide evidence that many of the human GIS gene homologs likely act in the suppression of genome instability in human cells and provide a restricted, prioritized list of human genes for genetics and functional validation studies.

METHODS

Query strain construction. The selectable markers used in the *MATα* query strains in systematic mating in the original SGA protocol are incompatible with the genetic markers required for GCR assays. Therefore different selectable markers were introduced into *MATα* query strains containing GCR assays. The selected markers were

as follows. First, as *CAN1* required by the GCR assay interferes with use of canavanine in combination with thialysine to kill diploid strains in the SGA protocol (Tong and Boone, 2006), we introduced a deletion of *LYP1* and the cycloheximide-resistant *cyh2-Q38K* mutation (Kàufer et al., 1983; Stöcklein et al., 1981) into our strains allowing the use of thialysine and cycloheximide to kill diploid strains in our SGA protocol. Second, we introduced a copy of *LEU2* driven by the *MFA1* promoter near the *YFR016C* gene to select for *MATa* haploid progeny. Third, we replaced the native *CAN1* gene with a selectable nourseothricin resistance gene driven by the *LEU2* promoter in the dGCR and sGCR assay strains. The *MATa* and *MATα* strains with P_{LEU2}-*NAT* were nourseothricin resistant when grown on complete synthetic media (CSM). However, the *MATa* strains were not nourseothricin resistant on YPD (1% Bacto-yeast extract, 2% Bacto-peptone, 2% dextrose) medium, which could possibly be due to increased expression of the Leu2 protein in *MATa* strains resulting in down-regulation of the *LEU2* promoter; this did not interfere with the selection scheme. The *CAN1/URA3* cassette and flanking targeting sequences was amplified from pRDK1378 and pRDKY1379 and integrated into RDKY7629 to generate the dGCR query strain RDKY7635 (*MATα hom3-10 ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 lyp1::TRP1 cyh2-Q38K iYFR016C::*P_{MFA1}*-LEU2 can1::*P_{LEU2}*-NAT yel072w::CAN1/URA3*) or the sGCR strain RDKY7964 (*MATα hom3-10 ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 lyp1::TRP1 cyh2-Q38K iYFR016C::*P_{MFA1}*-LEU2 can1::*P_{LEU2}*-NAT yel068c::CAN1/URA3*), respectively. The tyGCR assay strain was constructed by crossing RDKY6975 with RDKY6593 (Chan and Kolodner, 2011) and sporulating the resulting diploid to recover RDKY7046 (*MATα hom3-10 ura3Δ0 leu2Δ0 trp1Δ63*

his3Δ200 lyp1::TRP1 cyh2-Q38K iYFR016C::P_{MFA1}-LEU2 iYEL062W::Ty912-hphNT1 hxt13::URA3). Disruption of the 43 query genes in RDKY7635 with *HIS3* was performed using standard methods.

Bait strain collection. BY4741 derivatives that contained a deletion affecting one of the 1041 previously identified candidate GCR suppressing genes (Putnam et al., 2012) were retrieved from the *S. cerevisiae* deletion collection (Open Biosystems) yielding a starting bait strain collection. We then verified all of the deletions by PCR amplification using primers that hybridized within the inserted G418 resistance cassette and primers that hybridized to flanking sequences. Deletions that could not be verified were either replaced by crossing a verified BY4742 deletion strain with BY4741 and sporulating the resulting diploid or by PCR-mediated deletion in BY4741 when a verified BY4742 strain was unavailable. We also added a number of mutations of interest to the bait strain set that were not present in the BY4741 deletion collection, including *mec1::G418 sml1::hph*, *ddc2::G418 sml1::hph*, *rad53::G418 sml1::hph*, and *mrc1-aq.G418*. Additionally, we constructed a control strain by replacing *leu2Δ*0 present in BY4741 with the G418 resistance marker, which allows *leu2::G418*-containing progeny to be selected during systematic mating; these control strains are labeled as *leu2Δ* in the figures. The final bait strain collection consisted of 1041 single mutants (Putnam et al., 2012), an additional 16 single mutants and the *leu2::G418* control strain.

Screen for GCR suppressing genes and interacting genes. Query strains grown on YPD-agar were crossed to arrayed strains containing bait mutations on YPD-agar in quadruplicate by pinning onto a fresh YPD agar plate using a Singer RoToR robot and

grown for 1-2 days at 30˚C. The cells were then subjected to two rounds of pinning onto diploid selection media (YPD-agar containing 200 μg/mL geneticin (G418) and 100 μg/mL nourseothricin) and grown for 1-2 days at 30˚C. The cells were then pinned onto pre-sporulation media (containing 15 g Difco nutrient broth, 5 g Bacto-yeast extract, 10 g Bacto-agar, and 62.5 mL 40% glucose per 500 mL) and grown for 3 days at 30˚C. Cells from the pre-sporulation media were then pinned onto sporulation media (10 g potassium acetate, 05 g zinc acetate, 20 g Bacto-agar per liter containing a final concentration of 50 μg/mL G418 and 25 μg/mL nourseothricin) and incubated for 7 days at 30˚C. The resulting spore-containing cells were then subjected to two rounds of pinning onto diploid killing media (1.7 g yeast nitrogen base without amino acids and without ammonium sulfate, 1 g L-glutamic acid monosodium salt, 2 g CSM dropout mix without uracil, lysine, leucine, and, when appropriate, histidine, 20 g Bacto-agar, 50 mL of 40% glucose per liter containing a final concentration of 50 μg/mL thialysine, 10 μg/mL cycloheximide, 200 μg/mL G418, and 100 μg/mL nourseothricin) followed by growth for 5 days at 30˚C for the first pinning and 2 days at 30˚C for the second pinning. Cells were then subjected to two rounds of pinning and growth on haploid selection media (1.7 g yeast nitrogen base without amino acids and without ammonium sulfate, 1 g L-glutamic acid monosodium salt, 2 g CSM dropout mix without leucine, uracil, and, when appropriate, histidine, 20 g Bacto-agar, 50 mL of 40% glucose per liter containing a final concentration of 200 μg/mL G418, and 100 μg/mL nourseothricin) and grown for 2 days at 30˚C. Then the cells were pinned and grown on YPD-agar followed by storage at -85˚C.

GCR patch tests. A minimum of 3 independent isolates of each mutant progeny pool were then grown as patches on a YPD-agar plate at 30 °C for two days and replica-plated onto CSM –Arg media containing 60 mg/L canavanine and 1 g/L 5-fluoroorotic acid. The number of papilla growing on the GCR medium was scored using a semi-quantitative scoring system in which a score of '1' was the number of papilla in the *leu2Δ* control strain for the dGCR assay, scores of '2' to '5' were progressively higher amounts of papilla, and a score of '0' was a number of papilla lower than that of the dGCR control strain (Figure 1.2a). Then the scores for all independent patches analyzed for each mutant were averaged to generate a GCR strain score. Negative scores were assigned to strains that did not grow and so that these strains could be ignored during the analysis.

Determination of GCR rates. The media and protocol for strain propagation and measuring GCR rates were as described previously (Schmidt et al., 2006).

Analysis of cancer genomics data. TCGA data (The Cancer Genome Atlas Research Network, 2012, 2013), including expression z-scores, methylation and GISTIC CNV (copy number variation) data were obtained from the cBIO portal (http://www.cbioportal.org) through the CGDS-R package. Somatic mutation data was obtained from a local compilation (de Souza et al., 2014) that includes data from the TCGA, COSMIC as well as a compilation of data from the literature. As previously discussed (de Souza et al., 2014), all mutations for a given tumor were used in the S-score calculation. For all other analyses, only TCGA mutation data was used. As defined by TCGA, putative copy-number calls on samples were determined using the GISTIC algorithm (Beroukhim et al., 2007). Boxplots were generated using ggplot2, a graphics

tool for the R statistical package (http://ggplot2.org). For expression data, the Z-score metrics adopted by TCGA were used here.

Computational prediction of the functional impact of missense mutations. To identify putative deleterious missense mutations in our gene set, we used 5 different computational algorithms resulting in 6 different tests per mutation: SIFT (Kumar et al., 2009), PolyPhen-2 (Adzhubei et al., 2010), MutationTester (Schwarz et al., 2010), Fathmm (Shihab et al., 2013) and LTR (Chun and Fay, 2009). Two versions of PolyPhen-2 were used, each one trained by a different dataset (HDIV and HVAR). Each missense mutation was assigned a score, called the "Ndamage score" that was the number of prediction tests in which the mutation was scored as deleterious. To be considered "predicted deleterious", a given missense mutation had to have an Ndamage score of 5 or 6.

Simulations to determine statistical significance in cancer genomics analyses. Two types of simulations were used here. First, a gene set enrichment analysis was performed to evaluate whether the set of GIS genes were enriched with genes with extreme S-scores ($\leq$-2 or $\geq$2). Ten thousand random sets of the same size (number of genes) were selected from the pool of all human genes and for each set the number of genes with extreme S-scores was defined. A p-value for the enrichment analysis was determined by ranking the real set in the random set distribution. Second, we also evaluated whether a given set of genes was enriched for different types of mutations (or combinations of different types). To avoid any bias due to different gene lengths, we normalized the analysis for the total length of the corresponding gene set (in amino acids

of the longest coding region for each gene). The total number of amino acids for the real set was randomly selected from the total pool of human genes (ten thousands random sets). The number of mutations in the real set was then compared to all random sets and a p-value for the enrichment analysis was determined by ranking the real set within the distribution of the random sets.

ACKNOWLEDGEMENTS

This chapter is an adaptation of a manuscript that has been submitted for publication. The authorship and title are as follows: Putnam CD*, Srivatsan A*, Nene RV*, Martinez SL*, Clotfelter SP, Bell SN, Somach S, de Souza JES, Fonseca AF, de Souza SJ, Kolodner RD. An extended genetic network suppresses genome rearrangements in *S. cerevisiae*. (*Indicates first co-authorship). The work presented in this chapter was done during the first half of my graduate training. Christopher D. Putnam and Richard D. Kolodner conceived the overall experimental design and the systematic mating strategy. Anjana Srivatsan, Rahul V. Nene, Sandra L. Martinez, Sarah P. Clotfelter, and Sara N. Bell did strain construction, patching, and quantitative rate measurements. Christopher D. Putnam analyzed the resulting patch scores and rates. Sandro J. de Souza, Jorge E.S. de Souza, André F. Fonseca, Steven Somach, and Richard D. Kolodner analyzed the TCGA data. Christopher D. Putnam and Richard D. Kolodner wrote the manuscript and all other authors revised and modified the manuscript. The authors of the manuscript also thank Vincent Pennaneach and Jorrit Ensernik for assistance in construction of early versions of the query strains and Renan Valieris for

assistance in the construction of the MySQL database containing TCGA data. This work was supported by NIH grant R01-GM26017 to Richard D. Kolodner, NIH grant F30-CA177240 to Rahul V. Nene, CAPES (Brazil) grant (23038.004629/2014-19) to Sandro J. de Souza and support from the Ludwig Institute for Cancer Research to Richard D. Kolodner and Christopher D. Putnam.

Chapter 2:

Cdc73 Suppresses Genome Instability by Mediating Telomere Homeostasis and RNA:DNA Hybrid Formation

By

Rahul V. Nene[1], Christopher D. Putnam[1,2], and Richard D. Kolodner[1,3,4,5]

From

Ludwig Institute for Cancer Research[1], Departments of Medicine[2] and Cellular and Molecular Medicine[3], Moores-UCSD Cancer Center[4] and Institute of Genomic Medicine[5], University of California School of Medicine, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0669

INTRODUCTION

Genomic instability is an increasingly well-appreciated aspect of tumorigenesis. Genome rearrangements, such as translocations, copy number changes, and aneuploidy, are seen in many cancers (The Cancer Genome Atlas Research Network, 2013). A number of cancer susceptibility syndromes, such as Ataxia Telangiectasia, Lynch Syndrome, and familial breast and ovarian cancers, are caused by inherited mutations in genes involved in the DNA repair and DNA damage signaling pathways (Canman and Lim, 1998; Fishel et al., 1993; Yoshida and Miki, 2004). Numerous studies have also shown that sporadic tumors develop mutations in many of these same genes (Kolodner et al., 2011; Lengauer et al., 1998). Large scale projects such as The Cancer Genome Atlas (TCGA) hold promise to discover many more mutations that link genome instability to cancer (Chin and Gray, 2008). However, a better understanding of the genes and pathways that suppress genome instability is limited by a lack of rapid functional assays in mammalian systems. The budding yeast *Saccharomyces cerevisiae* is a model system that can be used to study the formation of gross chromosomal rearrangements (GCRs) and the genes that suppress them (Chan and Kolodner, 2011; Chen and Kolodner, 1999; Myung et al., 2001b; Putnam et al., 2009).

We have recently identified *CDC73* as a gene that suppresses genome instability in *S. cerevisiae* using in a large-scale genetic screen (Putnam *et al*, manuscript submitted; See Chapter 1). *CDC73* encodes a subunit of the Paf1 complex, which binds to RNA polymerase II and has known roles in transcription elongation and histone modification (Jaehning, 2010). Deletion of *CDC73* has previously been linked to increased rates of

chromosome loss (Yuen et al., 2007), hyper-recombination between direct repeats (Fan et al., 2001), and increased instability of yeast artificial chromosomes that could be abrogated by the overexpression of RNase H1 (Wahba et al., 2011), which acts on RNA:DNA hybrids (also called R-loops). R-loops caused by mutations in other pathways, such as the THO complex, are known to cause increases in genome instability that have been shown, in some cases, to be reversed by RNase H1 overexpression (Aguilera and García-Muse, 2012, 2013; Chávez and Aguilera, 1997; Santos-Pereira et al., 2013). A potential for increased R-loops caused by Paf1 complex defects is consistent with the role of the complex in transcription elongation and in the maturation of the 3' ends of mRNAs (Nordick et al., 2008; Sheldon et al., 2005). Deletion of *CDC73* in *S. cerevisiae* also causes a reduction in the level of the *TLC1* (Mozdy et al., 2008), which is the RNA template for the telomerase complex that maintains the ~300 bp telomeres that protect the ends of the *S. cerevisiae* chromosomes (Wellinger and Zakian, 2012). Consistent with this, deletion of *CDC73* causes a shortened telomere phenotype (Askree et al., 2004; Gatbonton et al., 2006).

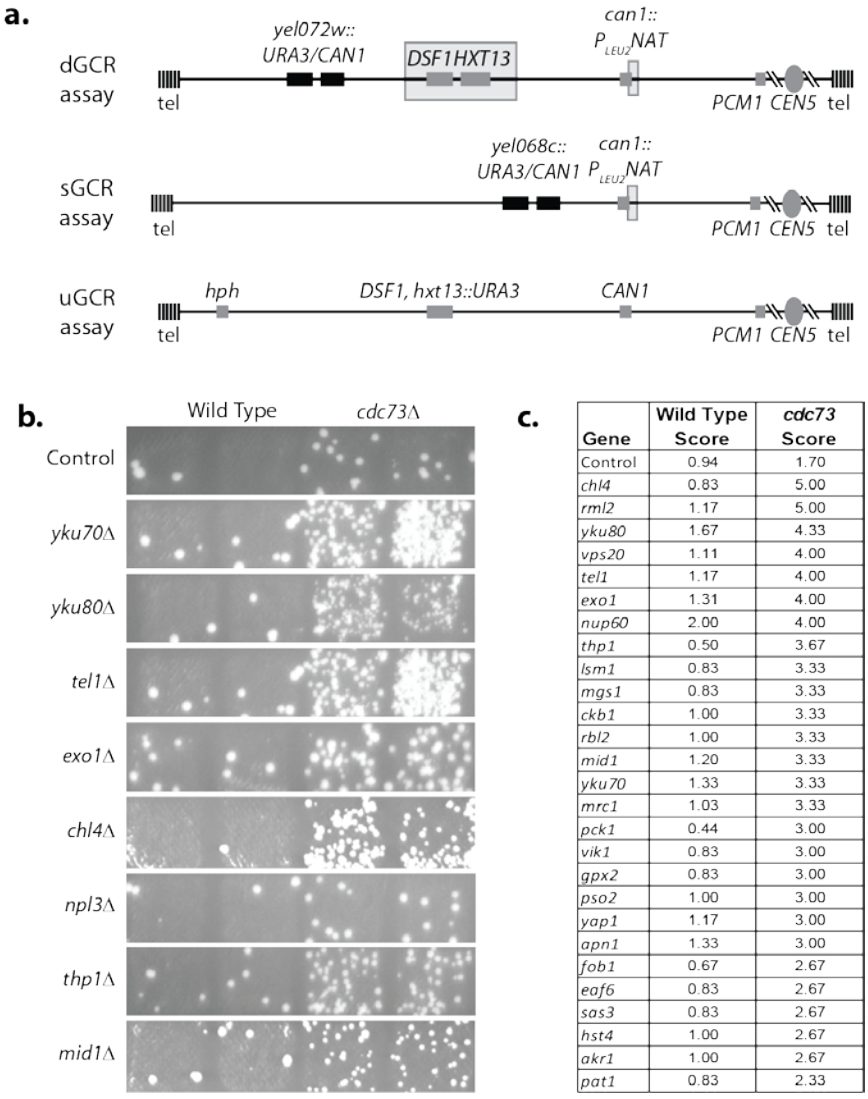
The Paf1 complex is well conserved among eukaryotes (Newey et al., 2009). Somatic *CDC73* mutations are often found in parathyroid cancer, and germline *CDC73* mutations cause hyperparathyroidism-jaw tumor syndrome (HPT-JT) (Newey et al., 2010). The human homolog has also been implicated in Wnt/β-catenin signaling and cell cycle regulation, but little else has been demonstrated (Mosimann et al., 2006; Zhang et al., 2006). Here we have investigated how *CDC73* acts in the suppression of genome instability in *S. cerevisiae*, with the goal to shed light on how its human homolog may

function as a tumor suppressor. Loss of *CDC73* shows synergistic interactions when combined with mutations causing defects in telomere maintenance, including *tel1Δ* and *yku80Δ*. In addition, rearrangements formed in *cdc73Δ* strains have reduced frequencies of rearrangements generated by *de novo* telomere additions. These strains also have substantial defects in assays that monitor different aspects of telomere structure. Remarkably, the increased instability of *cdc73Δ*, *cdc73Δ yku80Δ*, and *cdc73Δ tel1Δ* mutant strains can be partially suppressed by overexpression of the *TLC1* RNA or RNase H1, which specifically cleaves RNA from R-loops. These results demonstrate that defects in *CDC73* that lead to genome instability are due to defects in telomere maintenance and suppression of R-loop formation.

RESULTS

***cdc73Δ* causes strong synergistic increases in genome instability when combined with mutations affecting telomere homeostasis**. To identify mutations that enhance genome instability in combination with a deletion of *CDC73*, we crossed a strain containing the dGCR assay (Figure 2.1a) and a *cdc73Δ* mutation against a set of 638 strains from the *S. cerevisia*e deletion collection that were previously identified as being enriched for genome instability suppressing (GIS) genes and genes that acted cooperatively to suppress genome instability (Putnam et al., 2012; Putnam *et al*, manuscript submitted). The resulting diploids were sporulated, and haploid double mutant strains were scored by patch tests for the increased accumulation of Can^R $5FOA^R$ papillae relative to the *cdc73Δ* control (Figure 2.1b). *CDC73* was first identified as

playing a role in pheromone signaling (Reed et al., 1988), so *cdc73Δ* double mutants were also screened by flow cytometry for DNA content to exclude diploid isolates from the analysis. The screen identified 27 genes that showed a synergistic interaction with *cdc73Δ* when deleted (Figure 2.1c). Selected candidate interactors were verified by quantitative fluctuation assays (Table 2.1).



Gene	Wild Type Score	*cdc73* Score
Control	0.94	1.70
chl4	0.83	5.00
rml2	1.17	5.00
yku80	1.67	4.33
vps20	1.11	4.00
tel1	1.17	4.00
exo1	1.31	4.00
nup60	2.00	4.00
thp1	0.50	3.67
lsm1	0.83	3.33
mgs1	0.83	3.33
ckb1	1.00	3.33
rbl2	1.00	3.33
mid1	1.20	3.33
yku70	1.33	3.33
mrc1	1.03	3.33
pck1	0.44	3.00
vik1	0.83	3.00
gpx2	0.83	3.00
pso2	1.00	3.00
yap1	1.17	3.00
apn1	1.33	3.00
fob1	0.67	2.67
eaf6	0.83	2.67
sas3	0.83	2.67
hst4	1.00	2.67
akr1	1.00	2.67
pat1	0.83	2.33

Figure 2.1. Systematic analysis of *CDC73* as a Suppressor of Genome Instability.
a. The dGCR, sGCR, and uGCR assays involve selection against the *CAN1* and *URA3* genes placed on the terminal non-essential region of the left arm of chromosome V. The dGCR assay primarily forms rearrangements using the homology of the *DSF1/HXT13* segmental duplication (grey outline) with chromosomes IV, X, and XIV. For the sGCR assay, the only sequence with homology to other regions of the genome in the breakpoint region (between *CAN1/URA3* and *PCM1*) is a ~100 bp fragment of a delta sequence introduced by the *can1::*P_{LEU2}*-NAT* locus (also found in the dGCR assay). The uGCR assay contains no sequence homology within the breakpoint region. **b.** Sample patches of single and double mutants for genes with mutations that show a synergistic interaction with *cdc73Δ*. Each papilla corresponds to a GCR event and the greater the number of papillae per patch correlates to an increased GCR score. **c.** Patches were scored from 0 to 5, with 1 corresponding to baseline wild-type. Single and *cdc73Δ* double mutants were given a patch score averaged from at least 3 independent isolates.

Table 2.1. GCR rates of *cdc73Δ* single and double mutants.

	Single Mutant		*cdc73Δ* Double Mutant	
Genotype	***RDKY***	***Can^R 5FOA^R Rate†***	***RDKY***	***Can^R 5FOA^R Rate†***
Wild-type*	7635	8.59×10^{-8} (1.0)	7986	7.54×10^{-7} (8.8)
yku70	8420	3.82×10^{-7} (3.8)	8427	1.62×10^{-5} (189)
yku80	8339	3.29×10^{-7} (3.8)	8323	1.31×10^{-5} (152)
tel1	8340	3.38×10^{-7} (3.9)	8324	1.91×10^{-5} (222)
exo1	8419	2.21×10^{-7} (2.6)	8428	1.02×10^{-5} (119)
chl4	8421	2.91×10^{-7} (3.4)	8429	2.50×10^{-5} (290)
thp1	8422	2.15×10^{-7} (2.5)	8430	6.05×10^{-6} (70)
mid1	8423	7.66×10^{-8} (8)	8431	5.12×10^{-6} (60)
sir2	8172	5.05×10^{-7} (5.9)	8432	6.38×10^{-6} (74)
sir3	8173	5.40×10^{-7} (6.3)	8433	1.63×10^{-5} (190)
sir4	8424	6.49×10^{-7} (7.6)	8434	3.50×10^{-6} (41)
rad52	8425	7.00×10^{-9} (1)	8435	3.44×10^{-7} (4.0)
dnl4	8426	1.87×10^{-7} (2.2)	8436	6.96×10^{-7} (8.1)
yku80, rad52		ND	8437	5.12×10^{-7} (6.0)
yku80, dnl4		ND	8438	2.05×10^{-5} (240)
tel1, rad52		ND	8439	4.24×10^{-7} (4.9)
tel1, dnl4		ND	8440	5.15×10^{-5} (604)

* Rate data from (Putnam *et al*, manuscript submitted).

†Rate of accumulating Can^R 5FOA^R progeny. The number in parenthesis is the fold increase relative to the wild-type dGCR assay.

ND – Not Determined

Many strong *cdc73Δ* interactors were mutations that affected telomere homeostasis or the processing of telomeres with defective structures, although not all mutations implicated in causing shortened telomeres had synergistic increases in GCR formation (Askree et al., 2004; Gatbonton et al., 2006). Deletion of *YKU70* or *YKU80*, which encode subunits of the Ku complex, increased the relative rate from ~4 fold for the *yku70Δ* and *yku80Δ* strains and ~9 fold for the *cdc73Δ* strain to >150 fold for the double mutant strains. The Ku complex plays a role in telomere length maintenance and structure and nonhomologous end joining (NHEJ) (Bertuch and Lundblad, 2003; Boulton and Jackson, 1996); however, deletion of *DNL4*, which encodes the DNA ligase involved in NHEJ does not interact with *cdc73Δ* (Table 2.1). Similarly, deletion of *TEL1*, which

encodes a protein kinase involved in telomere length regulation and the DNA damage checkpoint (Greenwell et al., 1995; Morrow et al., 1995), shows a strong synergistic interaction when combined with *cdc73Δ*. Taken together, these results are consistent with *cdc73Δ* causing a partial telomere defect that synergizes with other mutations that cause telomere defects. We therefore tested the interaction of *cdc73Δ* with the deletion of other nonessential components of telomeres, such as *SIR2*, *SIR3*, and *SIR4*, which are involved in telomere maintenance (Moretti and Shore, 2001; Moretti et al., 1994), but are missing from our screen as these genes are also required for mating (Rine and Herskowitz, 1987; Wellinger and Zakian, 2012). As for the *yku70Δ*, *yku80Δ*, and *tel1Δ* mutations, strains combining *cdc73Δ* with *sir2Δ*, *sir3Δ*, or *sir4Δ* cause a synergistic increase in genome instability (Table 2.1). Interestingly, *cdc73Δ* also shows a strong synergistic interaction with a deletion of *EXO1*, which encodes a gene that acts in multiple aspects of DNA metabolism, including resection at deprotected telomeres (Fiorentini et al., 1997; Tran et al., 2004). The fact that the *cdc73Δ exo1Δ* double mutant also showed an increase in GCR rate could suggest that Exo1 resection at defective telomeres promotes genome stability via recombination between telomeres and/or that *cdc73Δ* may also influence genome stability by causing defects that are independent of any effects at telomeres. Based on the strength of the synergistic increases in GCR rates, we decided to focus primarily on the genetic interactions of *cdc73Δ* with *tel1Δ, yku80Δ*, and *exo1Δ*.

In wild-type strains, homology-driven non-reciprocal translocations between the *HXT13-DSF1* and imperfect homologies on chromosomes IV, X, and XIV dominate the rearrangements observed in the dGCR assay (Putnam et al., 2009). We analyzed the

products isolated from *cdc73Δ* strains by PCR using translocation-specific primers and found that the product distribution of the *cdc73Δ*, *cdc73Δ tel1Δ*, *cdc73Δ yku80Δ*, and *cdc73Δ exo1Δ* strains were essentially unchanged from wild-type, despite a >200-fold increase in the rate at which rearrangements accumulate (Figure 2.2). In contrast, disruption of homologous recombination by deletion of *RAD52* in *cdc73Δ* strains almost completely abrogated the formation of these GCR structures and suppressed the rates at which the rearrangements were formed (Table 2.1), which is consistent with previous findings (Putnam et al., 2009).

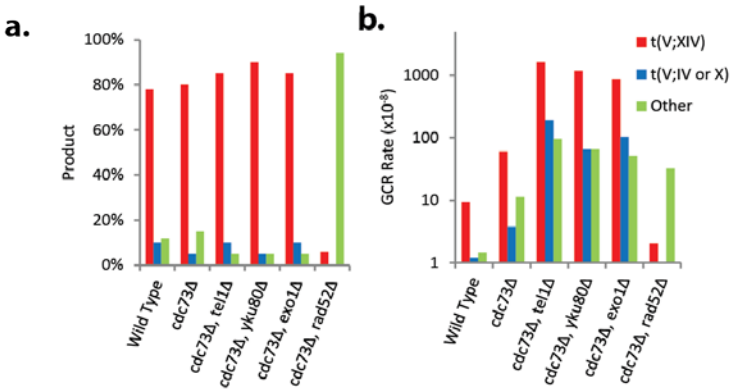


Figure 2.2. Summary of the types of GCRs formed by *cdc73Δ* Single and Double Mutants in the dGCR assay.
a. Percentage of the different types of GCRs formed in wild-type compared to *cdc73Δ* single and double mutants. The homology-driven GCRs are shown with t(V;XIV) in red and t(V;IV or X) in blue, while GCRs that do not use the segmental duplication are shown in green. **b.** Rates for each class of GCR were calculated by multiplying the fraction of each kind of rearrangement by the overall GCR rate of that genotype (from Table 2.1).

***cdc73Δ* reduces the efficiency of forming GCRs by *de novo* telomere addition.**

Because the GCR spectra for the *cdc73Δ* single and double mutants were dominated by homologous recombination in the dGCR assay, we instead introduced *cdc73Δ* into a strain with the sGCR assay on the left arm of chromosome V (Figure 2.1a). The sGCR assay cannot form GCRs mediated by recombination between segmental duplications, but it does contain a ~100 bp fragment of the *YCLWdelta5* Ty delta element upstream of the *LEU2* promoter that was used to construct the *can1::*P_{LEU2}*-NAT* locus (Putnam, *et al*, manuscript submitted). We used PCR gene disruption to reconstruct in the sGCR assay the *cdc73Δ* double mutants with *tel1Δ, yku80Δ*, and *exo1Δ* and measured GCR rates (Table 2.2). Rates in the sGCR assay are lower than in the dGCR assay, but loss of *CDC73* still causes a significant increase in the rate and it maintained the large synergistic interactions with *tel1Δ, yku80Δ*, and *exo1Δ*.

Table 2.2. Comparison of GCR rates of *cdc73Δ* mutants in the dGCR and sGCR assays.

	dGCR assay		*shGCR assay*	
Genotype	***RDKY***	***Can^R 5FOA^R Rate†***	***RDKY***	***Can^R 5FOA^R Rate$^\phi$***
Wild-type[*]	7635	8.59×10^{-8} (1.0)	7964	4.00×10^{-9} (1.0)
yku80	8339	3.29×10^{-7} (3.8)	8406	3.25×10^{-9} (8)
tel1	8340	3.38×10^{-7} (3.9)	8405	7.11×10^{-9} (1.8)
exo1	8419	2.21×10^{-7} (2.6)	8469	1.31×10^{-8} (3.3)
cdc73	7986	7.54×10^{-7} (8.8)	8407	1.68×10^{-7} (42)
yku80, cdc73	8323	1.31×10^{-5} (152)	8411	1.51×10^{-6} (377)
tel1, cdc73	8324	1.91×10^{-5} (222)	8409	8.00×10^{-7} (200)
exo1, cdc73	8428	1.02×10^{-5} (119)	8470	2.22×10^{-7} (56)
exo1, yku80	8463	1.44×10^{-7} (1.7)	8472	8.78×10^{-10} (2)
exo1, tel1	8464	6.13×10^{-7} (7.1)	8473	1.86×10^{-8} (4.6)
exo1, yku80, cdc73	8465	6.28×10^{-5} (730)	8474	1.74×10^{-7} (44)
exo1, tel1, cdc73	8466	2.05×10^{-6} (25)	8475	7.26×10^{-6} (1800)
yku80, tel1	8467	2.27×10^{-6} (27)	8408	1.86×10^{-8} (4.6)
yku80, tel1, cdc73	8468	1.64×10^{-4} (1900)	8413	7.17×10^{-8} (18)

[*] Rate data from (Putnam *et al*, manuscript submitted).
[†]Rate of accumulating Can^R 5FOA^R progeny. The number in parenthesis is the fold increase relative to the wild-type dGCR assay.
[ϕ]Rate of accumulating Can^R 5FOA^R progeny. The number in parenthesis is the fold increase relative to the wild-type sGCR assay.

We used paired-end next-generation sequencing to sequence the genomes of 1 parental strain and 11 isolates containing a Can^R 5FOA^R GCR from the wild-type, the *cdc73Δ* single mutant, and the *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutant sGCR strains. The sequencing data was analyzed to identify the structures of any rearranged chromosomes (Figure 2.3; Table 2.3). The analysis readily identified all of the engineered changes in the strain, such as the *trp1Δ63* deletion and disruption of *CDC73* by *HIS3*. In the wild-type sGCR strain, 5 of the 11 isolates formed GCRs by *de novo* telomere addition. These isolates had a characteristic pattern of copy number changes derived from the sequencing data involving loss of a non-essential portion of chrV L (telomeric to *PCM1*) and no other changes. In addition, the novel junctions between chrV L and the *de novo* telomeres could be sequenced by aligning the short reads that did not map to the

reference chromosomes but were paired with short reads mapping to chrV L adjacent to the terminal deletion. The remaining 6 of 11 GCR-containing isolates from the wild-type sGCR strain were formed by recombination mediated by the ~100 bp *YCLWdelta5* fragment at the *can1::*P_{LEU2}*-NAT* locus. These GCRs had two characteristic features identifiable in the sequencing data: (1) read pairs indicate the insertion of the P_{LEU2} cassette into the *CAN1* locus, but all sequences telomeric to that site, including the *nat* gene were missing, and (2) duplications bounded by Ty transposon-derived sequences and telomeres were found elsewhere in the genome. The duplicated regions in most isolates with GCRs were simple and started at a Ty-related sequence and ended at a telomere (for example isolates 542 and 544; Figure 2.3). Other isolates showed more complicated patterns, with multiple duplicated regions bounded by repetitive sequences, suggesting several rounds of homology-mediated break-induced replication, consistent with the formation and subsequent rearrangement of dicentric intermediates (for example isolate 547). Taken together, these data show that the sGCR assay has advantages in characterizing changes in product spectra of mutations, as the wild-type sGCR assay is balanced between generating GCRs by *de novo* telomere addition and homology-mediated rearrangements. This feature of the sGCR assay is unlike the unique sequence GCR (uGCR) assay, which is dominated by *de novo* telomere additions (Chen and Kolodner, 1999), and the dGCR assay, which is dominated by homology-mediated rearrangements (Putnam et al., 2009).

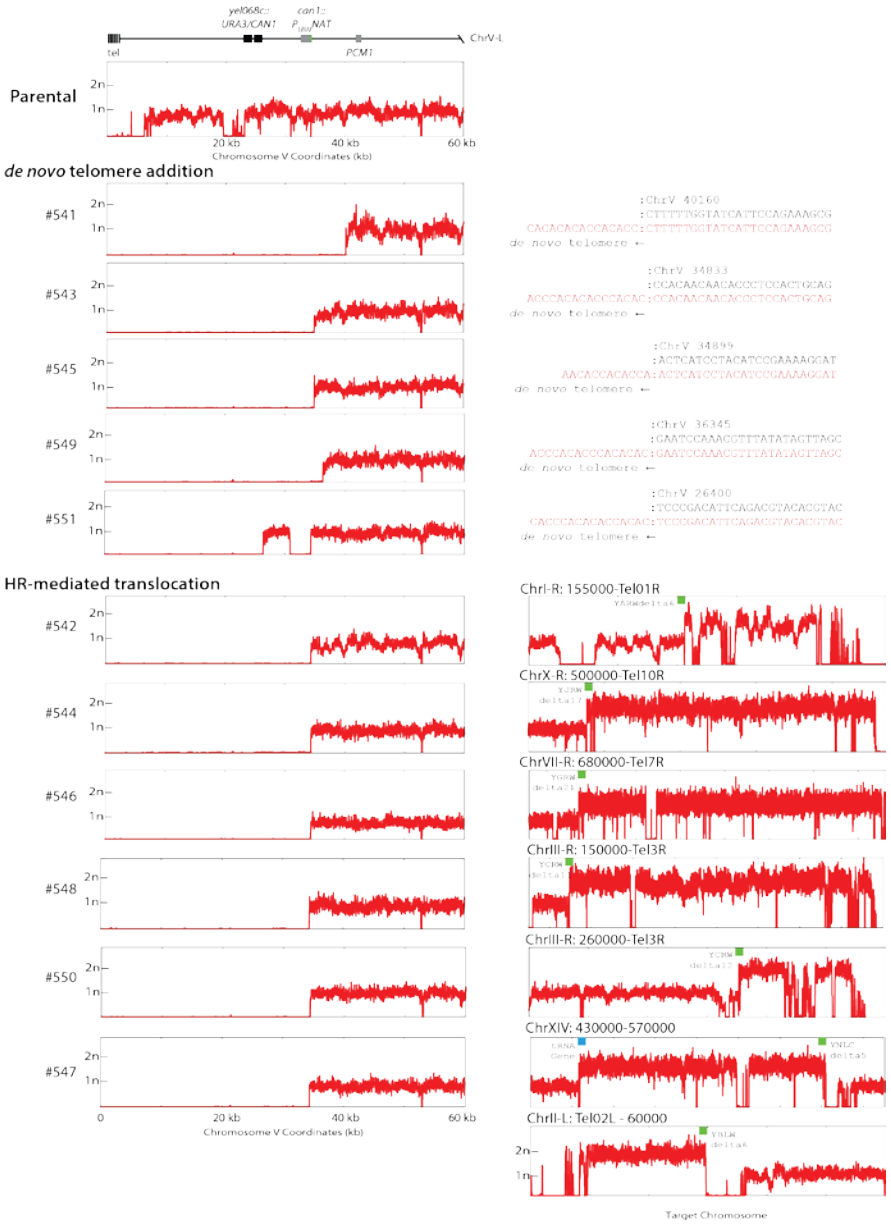


Figure 2.3. The sGCR Assay Produces GCRs by Both *de novo* Telomere Addition and Homologous Recombination.

Copy number analysis from the next generation sequencing data provides information about the site of the breakpoint on the left arm of chromosome V. *de novo* telomere additions can be identified by analyzing the sequence at the site of the breakpoint. The figure shows the sequencing reads, shown in red, which map to ChrV L up to the site of novel telomeric sequence. For HR-mediated GCRs, the target chromosome used as a template can also be identified by copy number analysis, which shows twice the number of reads mapping to that region. These GCRs were formed by delta mediated recombination between the P_{LEU2} delta sequence on ChrV L and the delta element shown on the target chromosome, both of which are illustrated in green.

Table 2.3. Junction-Defining Read Pairs and Junction-Sequencing Reads for Wild-Type sGCR Isolates

Sample	GCR-related Junction[#]	GCR Description
RDKY7964	n.a.	No GCR (starting strain)
#541	-/49	*de novo* telomere addition at the site of the breakpoint
#542	36/41	translocation to ChrIR at YARWdelta6, mediated by pLEU2 delta/delta recombination
#543	-/67	*de novo* telomere addition at the site of the breakpoint
#544	142/22	translocation to ChrXR at YJRWdelta17, mediated by pLEU2 delta/delta recombination
#545	-/112	*de novo* telomere addition at the site of the breakpoint
#546	not observed*	translocation to ChrVIIR at YGRWdelta21, mediated by pLEU2 delta/delta recombination
#547	46/38	translocation to ChrXIVLᵠ
	not observed*	translocation from YNLCdelta5 to ChrIIL at YBLWdelta6, mediated by delta/delta recombination
#548	98/57	inversion on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	not observed*	translocation from YELWdelta2 to ChrIIIR at YCRWdelta11, mediated by delta/delta recombination
#549	-/113	*de novo* telomere addition at the site of the breakpoint
#550	159/75	translocation to ChrIIIR at YCRWdelta12, mediated by pLEU2 delta/delta recombination
#551	-/93	*de novo* telomere addition at the site of the breakpoint

[#]Results are reported as the number of junction-defining read pairs before the slash and the number of junction-sequencing reads after the slash.
*Target of delta mediated translocation to a site of highly repetitive sequences, so could not be identified by Next Generation Sequencing, but was identified by copy number analysis
ᵠTranslocation was at the site of a t(RNA) gene at ~443 kb. The background strain likely has a Ty element at this site that is not present in the S288c reference genome.

Analysis of 11 GCRs formed in the *cdc73Δ* single mutant sGCR strain by next generation sequencing revealed that loss of *CDC73* altered the nature of the GCRs formed (Figure 2.4; Table 2.4). Nine of the 11 GCRs involved homology-mediated recombination; 2 of these were interchromosomal translocations and 7 were inverted duplications on ChrV L that involved homology between the P_{LEU2}-*YCLWdelta5* fragment and an oppositely oriented delta sequence on ChrV L. In all cases, these

inverted duplications did not extend to the ChrV centromere and had other duplicated regions, suggesting additional rounds of rearrangements. Once again, the duplicated regions were typically bounded by a Ty-related element and a telomere, consistent with previous observations that have implicated Ty-mediated recombination in the breakage-fusion-bridge cycles used to resolve dicentric chromosomes (Pennaneach and Kolodner, 2009a; Putnam et al., 2014). For some of these inverted duplications, we were unable to identify which telomere had been copied during GCR formation, as no duplicated regions that extended to a telomere were observed (for example isolate 305); these cases may correspond to recombination with one of the several telomere-proximal Ty-related elements that are in repetitive regions of the genome where sequencing reads do not map uniquely. Two of the 11 GCRs were formed via a non-homology mediated inverted duplication of ChrV L and a translocation to the subtelomeric region of ChrXV R. Thus, the sequenced GCR-containing isolates generated in the *cdc73Δ* single mutant sGCR strain showed a lack of GCRs formed by *de novo* telomere addition, a substantially reduced proportion of translocations formed by recombination of P_{LEU2}-*YCLWdelta5* with delta sequences on other chromosomal arms, and a dramatic increase in inverted duplication formed by recombination of P_{LEU2}-*YCLWdelta5* with other delta sequences on ChrV L.

Figure 2.4 (See Next Page). GCR Isolates in a *cdc73Δ* Background Favor Delta Mediated Homologous Recombination.
Sample GCR structures for *cdc73Δ*, *cdc73Δ tel1Δ*, and *cdc73Δ yku80Δ*. Copy number analysis from the next generation sequencing data shows that the site of the breakpoint on the left arm of chromosome V for most of the GCRs occurred at the site of the *pLEU2* delta sequence. In addition, many of the GCRs have an inverted duplication on ChrV L, mediated by recombination between the P_{LEU2} delta sequence and the *YELWdelta1/2* sequence ~25 kb centromeric to the breakpoint. The final GCR structure can also typically be identified by copy number analysis. In all 6 cases shown here, the nonreciprocal translocation was mediated by homology between the *YELWdelta5/6* sequence on ChrV L and the labeled delta sequence on the target chromosome. All sites of delta sequence are shown in green. Isolate 301 had a translocation to a subtelomeric site at ChrXV L, but could not be identified by copy number changes because this is a repetitive region of the genome where sequences do not uniquely map. However, aligning the short reads that did not map to the reference chromosomes but were paired with short reads mapping to ChrV L adjacent to the breakpoint was able to identify the junction defining reads that span both ChrV L and ChrXV L. No additional changes in copy number were identified for isolates 302, 305, and 311, but it is likely these isolates underwent additional Ty element mediated recombination for telomere capture, but these events could not be identified by copy number analysis or breakpoint sequence analysis because the events occurred at highly repetitive regions.


yel068c::
URA3/CAN1
can1::
P_{TEF2}-NAT
YELW
delta1/2
YELW
delta5/6
ChrV-L
tel
PCM1
CEN5
Parental
2n
1n
0
50 kb
100 kb
150 kb
Chromosome V Coordinates
#301
#302
#303
#304
#305
#306
#307
#308
#309
#310
#311
:ChrV 39907
:AAAAAT:GAAATTCAGATTGAAAATAA
GAAATTAAAAAAA:AAAAAT:GAAATTCAGATTGAAAATAA
GAAATTAAAAAAA:AAAAAT:
ChrXV 1091207:
No other changes observed
ChrV-R: 420000-Tel05R
ChrXVI-R: 830000-Tel16R
No other changes observed
ChrV-R: 420000-Tel05R
ChrV-R: 420000-Tel05R
ChrV-R: 420000-Tel05R
ChrI-R: 150000-Tel01R
ChrII-L: Tel02L-60000
No other changes observed
Chromosome V Coordinates
Target Chromosome

Table 2.4. Junction-Defining Read Pairs and Junction-Sequencing Reads for *cdc73Δ* sGCR Isolates

Sample	GCR-related Junction	GCR Description
RDKY8407	n.a.	No GCR (starting strain)
#301	0/128	translocation to subtelomeric sequences of TEL15R
#302	501/246	translocation to ChrXIIIR at YMRWdelta19, mediated by pLEU2 delta/delta recombination
	-	No Further Changes Seen
#303	1/3	hairpin duplication at the site of the breakpoint
	34/2	translocation from YELWdelta2 to ChrVR at YERWdelta17, mediated by delta/delta recombination
#304	264/124	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	not observed*	translocation from YELWdelta6 to ChrXVIR at YPRWdelta20, mediated by delta/delta recombination
#305	402/179	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	22/9	second inverted duplication on chrVL from YELWdelta5 to YELWdelta2, mediated by delta/delta recombination
	-	No Further Changes Seen
#306	108/42	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	15/0	translocation from YELWdelta6 to to ChrVR at YERWdelta17, mediated by delta/delta recombination
#307	174/52	inversion on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	25/5	translocation from YELWdelta2 to ChrVR at YERWdelta17, mediated by delta/delta recombination
#308	101/31	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	9/0	translocation from YELWdelta6 to ChrVR at YERWdelta17, mediated by delta/delta recombination
#309	97/13	translocation to ChrIR at YARWdelta6, mediated by pLEU2 delta/delta recombination
#310	107/37	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	not observed*	translocation from YELWdelta6 to ChrIIL at YBLWdelta6, mediated by delta/delta recombination
#311	117/41	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	-	No Further Changes Seen

[#]Results are reported as the number of junction-defining read pairs before the slash and the number of junction-sequencing reads after the slash.
*Target of delta mediated translocation to a site of highly repetitive sequences, so could not be identified by Next Generation Sequencing, but was identified by copy number analysis

Analysis of 11 GCRs formed in the *cdc73Δ tel1Δ* double mutant sGCR strain (Figure 2.5; Table 2.5) and 11 GCRs formed in *cdc73Δ yku80Δ* double mutant sGCR strain (Figure 2.6; Table 2.6) by next generation sequencing revealed that the GCR spectra were not substantially altered from the *cdc73Δ* single mutant sGCR strain, despite the large increase in GCR rates. Inverted duplications formed by recombination of the P_{LEU2}-*YCLWdelta5* fragment with other delta sequences on ChrV L were found in 7 of the 11 GCRs from the *cdc73Δ tel1Δ* double mutant strain and 9 of the 11 GCRs from the *cdc73Δ yku80Δ* double mutant strain. ChrV inversions not involving the P_{LEU2}-*YCLWdelta5* fragment, which are frequently found in *tel1Δ* single mutant strains (Putnam et al., 2014), were found in 3 and 1 of the 11 GCRs from the *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutant strains, respectively. The remaining GCRs (2 from the *cdc73Δ tel1Δ* double mutant strain and 1 from the *cdc73Δ yku80Δ* double mutant strain) were interchromosomal rearrangements between the P_{LEU2}-*YCLWdelta5* fragment and Ty-related sequences on other chromosomes.

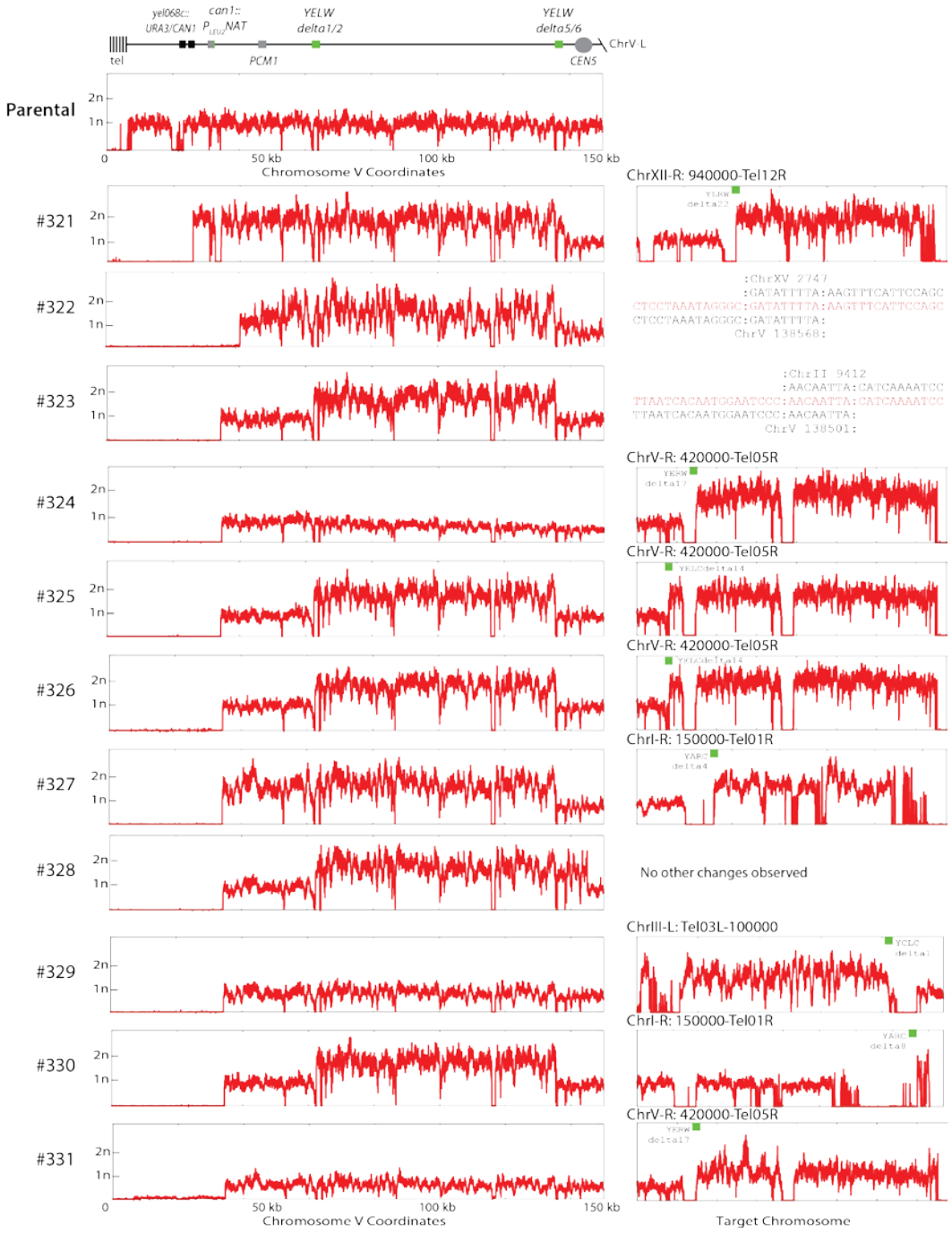


Figure 2.5. Extended Analysis of GCRs Produced in a *cdc73Δ tel1Δ* Double Mutant in the sGCR Assay.

Extension of Figure 2.4, with GCRs from a *cdc73Δ tel1Δ* double mutant. Isolates 321, 322, and 327 are consistent with a hairpin inversion, with a duplication starting exactly at the site of the breakpoint. Isolates 322 and 323 had translocations to subtelomeric sites that could not be identified by copy number changes, but could be identified by breakpoint sequence analysis.

Table 2.5. Junction-Defining Read Pairs and Junction-Sequencing Reads for *cdc73Δ tel1Δ* sGCR Isolates

Sample	GCR-related Junction	GCR Description
RDKY8409	n.a.	No GCR (starting strain)
#321	32/70	hairpin duplication at the site of the breakpoint
	not observed*	translocation from YELWdelta6 to ChrXIIR at YLRWdelta22, mediated by delta/delta recombination
#322	301/70	hairpin duplication at the site of the breakpoint
	60/2	translocation from YELWdelta6 to ChrXVL at YOLCdelta1, mediated by delta/delta recombination
#323	229/103	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	44/6	translocation from YELCdelta4 to ChrIIL at YBLWdelta3, mediated by delta/delta recombination
#324	367/177	inversion on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	41/12	translocation from YELWdelta2 to ChrVR at YERWdelta17, mediated by delta/delta recombination
#325	262/133	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	157/0	translocation from YELCdelta4 to ChrVR at YERCdelta14, mediated by delta/delta recombination
#326	423/189	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	67/16	translocation from YELCdelta4 to ChrVR at YERCdelta14, mediated by delta/delta recombination
#327	52/64	hairpin duplication at the site of the breakpoint
	not observed*	translocation from YELCdelta4 to ChrIR at YARCdelta4, mediated by delta/delta recombination
#328	230/113	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	-	No Further Changes Seen
#329	8/56	translocation to ChrIIIL at YCLCdelta1, mediated by pLEU2 delta/delta recombination
#330	208/92	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	not observed*	translocation from YELCdelta4 to ChrIR at YARCdelta8, mediated by delta/delta recombination
#331	128/37	inversion on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	12/2	translocation from YELWdelta2 to ChrVR at YERWdelta17, mediated by delta/delta recombination

[#]Results are reported as the number of junction-defining read pairs before the slash and the number of junction-sequencing reads after the slash.
*Target of delta mediated translocation to a site of highly repetitive sequences, so could not be identified by Next Generation Sequencing, but was identified by copy number analysis.

Figure 2.6. Extended Analysis of GCRs Produced in a *cdc73Δ yku80Δ* Double Mutant in the sGCR Assay.

Extension of Figure 2.4, with GCRs from a *cdc73Δ tel1Δ* double mutant. Isolate 349 is consistent with a hairpin inversion, with the duplication starting exactly at the site of the breakpoint. Isolates 345, 348 and 354 had translocations to subtelomeric sites that could not be identified by copy number changes, but could be identified by breakpoint sequence analysis. Isolate 347 has an inverted duplication on ChrV L, followed by a translocation mediated by homology between the *PAU2* gene on ChrV L and another member of the seripauperin multigene family that is located at a repetitive subtelomeric site.

Table 2.6. Junction-Defining Read Pairs and Junction-Sequencing Reads for *cdc73Δ yku80Δ* sGCR Isolates

Sample	GCR-related Junction	GCR Description
RDKY8411	n.a.	No GCR (starting strain)
#345	65/42	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	5/8	translocation from YELCdelta4 to ChrVIIIR at YHRCdelta15, mediated by delta/delta recombination
#346	117/77	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	-	No Further Changes Seen
#347	98/57	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	25/3	translocation of PAU2 to unidentified subtelomeric sequence, mediated by homeologous recombination wuth the seripauperin multigene family
#348	0/71	translocation to subtelomeric sequences of TEL15R
#349	45/80	hairpin duplication at the site of the breakpoint
	20/4	translocation from YELCdelta4 to ChrIIL at YBLWdelta6, mediated by delta/delta recombination
#350	372/163	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	-	No Further Changes Seen
#351	296/150	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	not observed*	translocation from YELCdelta4 to ChrIIL at YBLWdelta6, mediated by delta/delta recombination
#352	300/129	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	40/8	translocation from YELCdelta4 to ChrXIIR at YLRCdelta26, mediated by delta/delta recombination
#353	323/144	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	418/75	translocation from YELWdelta6 to ChrIIIR at YCRWdelta12, mediated by delta/delta recombination
#354	309/118	translocation from pLEU2 delta to ChrIIL at YBLWdelta2, mediated by pLEU2 delta/delta recombination
#355	302/145	inverted duplication on chrVL from pLEU2 delta to YELWdelta2, mediated by pLEU2 delta/delta recombination
	not observed*	translocation from YELCdelta4 to ChrIIL at YBLWdelta6, mediated by delta/delta recombination

[#]Results are reported as the number of junction-defining read pairs before the slash and the number of junction-sequencing reads after the slash.
*Target of delta mediated translocation to a site of highly repetitive sequences, so could not be identified by Next Generation Sequencing, but was identified by copy number analysis.

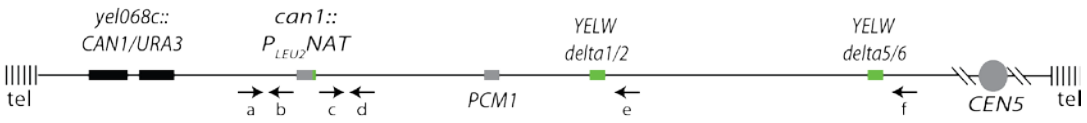


Figure 2.7. Design of primer pairs to query the GCR structure of isolates from the sGCR assay.
Figure shows the structure of ChrV L in the sGCR assay, specifically highlighting the locations of delta sequences (shown in green). Use of the P_{LEU2} delta sequence is suggested if a+b fails to give a PCR product but c+d does. If both or neither of these primer pairs gives a PCR product, that suggests the breakpoint occurred at a different site, and the GCR is likely a *de novo* telomere addition. Presence of a P_{LEU2} delta mediated inverted duplication can be detected if either d+e or d+f produces a PCR product.

Given that the preponderance of GCR isolates containing P_{LEU2}-*YCLWdelta5*-mediated translocations or inverted duplications, we designed sets of primers to facilitate the screening of additional GCR isolates without having to sequence them (Figure 2.7). The primer pairs allow us to determine whether the breakpoint was directly at the P_{LEU2}-*YCLWdelta5* locus and also whether the GCR contained a P_{LEU2}-*YCLWdelta5*-mediated inverted duplication on ChrV L. We tested the primers on isolates that were sequenced to confirm that the PCR products matched the indicated GCR structure before screening more GCR isolates (Table 2.7).

Table 2.7. PCR Analysis of sGCR Isolates.

		pLEU2 Mediated Translocation		*pLEU2* Mediated Inverted Repeat		Putative *de novo* Telomere Addition	
	sGCR Rate	**n**	**Rate of Formation***	**n**	**Rate of Formation***	**n**	**Rate of Formation***
Wild-type	4.00×10^{-09}	15	2.40×10^{-09} (1.0)	2	3.20×10^{-10} (1.0)	8	2.40×10^{-09} (1.0)
*cdc73*Δ	1.68×10^{-07}	7	5.87×10^{-08} (24)	11	9.22×10^{-08} (290)	2	1.68×10^{-08} (13)
*pif1*Δ	2.49×10^{-06}	8	7.96×10^{-07} (330)	1	9.94×10^{-08} (310)	16	1.59×10^{-06} (1200)
*cdc73*Δ, *pif1*Δ	4.36×10^{-07}	12	2.09×10^{-07} (87)	2	3.49×10^{-08} (110)	11	1.92×10^{-07} (150)

* The Rate of Formation is calculated by multiplying the fraction of each type of GCR formed for that genotype by the GCR rate of the strain. The number in parenthesis indicates the fold increase over wild-type for that GCR structure. Note: *pLEU2* Mediated Translocation refers to translocations that are not inverted repeats.

Included in this analysis is the rate of formation for each type of structure, which is determined by multiplying the GCR rate of the strain by the fraction of GCRs seen for that structure. Deletion of *PIF1* causes a substantial increase in GCR rate, with GCRs predominantly formed by *de novo* telomere addition, both of which can be abrogated by loss of telomerase components or Ku complex subunits (Myung et al., 2001b). Given that GCRs formed by *de novo* telomere addition were not observed among the Can^R $5FOA^R$ resistant isolates from the *cdc73Δ* single mutant strain and that the GCR spectrum did not substantially change in the *cdc73Δ yku80Δ* and *cdc73Δ tel1Δ* double mutant strains, we tested the effect of combining *cdc73Δ* with *pif1Δ*. The *pif1Δ* mutation caused an increased rate of forming P_{LEU2}-*YCLWdelta5* fragment-mediated GCRs, which is potentially consistent with a role of Pif1 in break-induced replication (Saini et al., 2013), but an even greater increase in the rate of *de novo* telomere additions. In contrast, the *cdc73Δ pif1Δ* double mutant strain had a reduced GCR rate that was ~6-fold lower than the *pif1Δ* single mutant strain and the class of rearrangements that was suppressed to the

greatest extent, but not eliminated, was *de novo* telomere additions. These results indicate that *cdc73Δ* causes a significant defect in *de novo* telomere addition, though not as severely as *yku70Δ*, *yku80Δ*, or the deletion of telomerase components. These data also predict that the *cdc73Δ* single mutation and particularly the *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutations may cause substantial defects in the structures of telomeres.

***cdc73Δ* causes telomere defects that synergize with deletion of *YKU80* and *TEL1*.** Cells lacking components of the Paf1 Complex have shortened telomeres, as do cells with deletions of *YKU80* or *TEL1* (Askree et al., 2004; Gatbonton et al., 2006). It has also been previously determined that the telomere shortening in *cdc73Δ* is due to decreased expression of the telomerase component *TLC1*, though overexpression of *TLC1* only partially rescues telomere length (Mozdy et al., 2008). We therefore determined the telomere lengths of freshly sporulated cells with deletions of *CDC73* combined with deletions of *YKU80* or *TEL1* (Figure 2.8a). We first confirmed the short telomere profile of each single mutant and determined that deletion of *EXO1* could partially recover telomere length, consistent with previous findings that *EXO1* is the primary nuclease that degrades deprotected telomeres (Maringele and Lydall, 2002). The *cdc73Δ tel1Δ* double mutant showed extremely shortened telomeres, with deletion of *EXO1* causing only a slight improvement. Remarkably, the *cdc73Δ yku80Δ* double mutant strain showed a pattern consistent with that of a post-senescent survivor, with highly variable amplification of subtelomeric sequences (Wellinger and Zakian, 2012), though additional deletion of *EXO1* appears to abrogate the rapid entrance of the *cdc73Δ yku80Δ* double mutant into senescence and survival. These results mirror previous

findings that a *yku80Δ est2Δ* double mutant rapidly forms survivors, and that *exo1Δ* alleviates this rapid loss of viability (Bertuch and Lundblad, 2004). In addition, we tested a *tel1Δ yku80Δ* double mutant and *cdc73Δ tel1Δ yku80Δ* triple mutant. Consistent with previous findings, the *tel1Δ yku80Δ* double mutant showed telomeres that were shorter than either single mutant (Gravel et al., 1998), suggesting a non-epistatic relationship between these genes, while a *cdc73Δ tel1Δ yku80Δ* triple mutant showed a post-senescence pattern similar to the *cdc73Δ yku80Δ* double mutant.

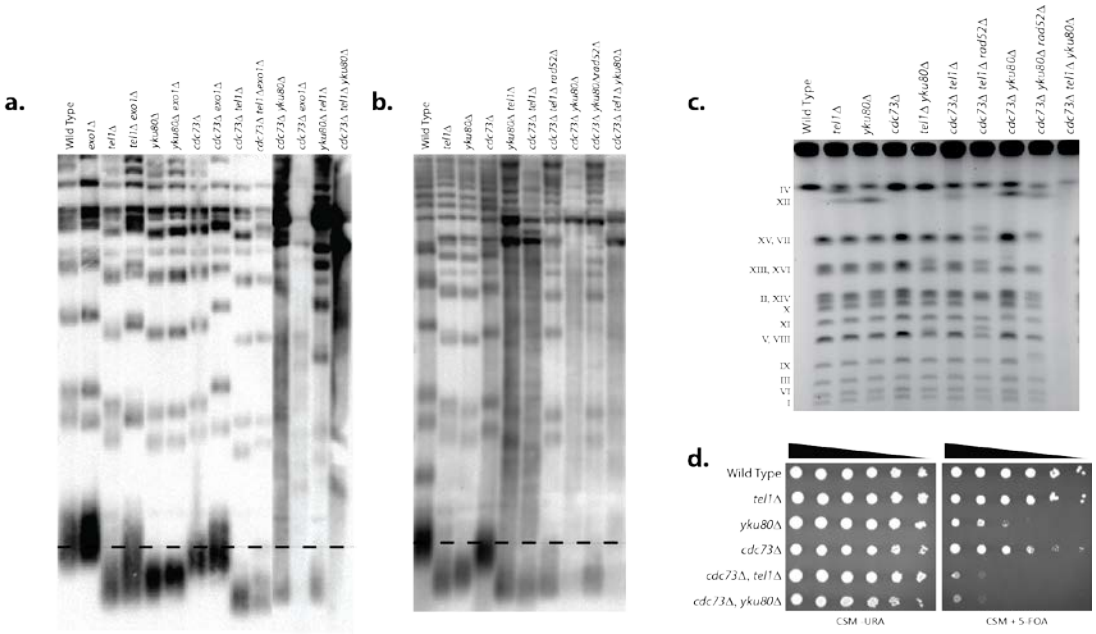


Figure 2.8. Loss of *CDC73* Results in a Telomere Defect.
a. Southern blot of *Xho*I-digested genomic DNA isolated from freshly sporulated strains of the indicated genotypes. **b.** Strains were serially propagated on non-selective media for >20 restreaks and then tested by telomere Southern blot as above. **c.** Pulse field gel electrophoresis allows for gross visualization of chromosomes on an agarose gel. The chromosome locations in wild-type are labeled as shown on the left. Fresh strains of the indicated genotype were tested. Decreased band intensity and increased smearing can be seen in strains that were shown to undergo senescence. **d.** Telomere Position Effect (TPE) assay for *cdc73Δ*, *tel1Δ*, and *yku80Δ* single and double mutants. Serial dilutions are plated on non-selective complete synthetic media lacking only uracil (SCM -URA) or selective media with 5-FOA. Loss of telomeric silencing is indicated by increased sensitivity to 5-FOA.

Since some of these strains showed hallmarks of post-senescent survivorship, we attempted to determine if the *cdc73Δ*, *tel1Δ*, *yku80Δ*, single, double and triple mutant strains would show the classic signs of crisis and escape from senescence by recovery of improved growth by serially restreaking strains generated by sporulation on non-selective medium. Consistent with previous findings, the *tel1Δ yku80Δ* double mutant strain was initially quite sick but eventually recovered a wild-growth growth rate, suggesting it does indeed go through crisis and post-senescent survival (Porter et al., 1996). Despite the fact that the *cdc73Δ* single mutant has a slow growth phenotype that is exacerbated by deletion of *TEL1* and *YKU80*, we did not observe any improvement of growth rates in any of the strains, even after >20 rounds of restreaking, which may suggest that the decreased growth rate in these strains is due defects other than just telomere erosion. We retested these serially propagated strains by telomere Southern blot (Figure 2.8b). As expected, the *tel1Δ yku80Δ* double mutant now shows a pattern consistent with post-senescent survivorship, but remarkably, so does the *cdc73Δ tel1Δ* double mutant. Post-senescent Type I survivors rely on amplification of subtelomeric Y' elements, while Type II survivors show large increases in $C_{1\text{-}3}A/TG_{1\text{-}3}$ telomeric repeats; both classes rely on *RAD52* to remain viable (Lundblad and Blackburn, 1993; Teng and Zakian, 1999). However, we were able to generate *cdc73Δ tel1Δ rad52Δ* and *cdc73Δ yku80Δ rad52Δ* triple mutants, and while these cells did have very short telomeres, they do not show a pattern consistent with post-senescent survivors, suggesting telomerase is still functional in the *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutants. However, we were unable to generate a *cdc73Δ tel1Δ yku80Δ rad52Δ* quadruple mutant by either PCR mediated gene

disruption or sporulation, suggesting that telomerase function is completely diminished in the *cdc73Δ tel1Δ yku80Δ* triple mutant.

Given the drastic telomere defects and the genome instability, we analyzed the chromosomal structure by pulse field gel electrophoresis (PFGE) of fresh strains containing combinations of the *cdc73Δ*, *tel1Δ*, and *yku80Δ* mutations (Figure 2.8c). The *cdc73Δ*, *tel1Δ*, and *yku80Δ* single mutant strains and the *cdc73Δ tel1Δ* double mutant strain have chromosomal banding patterns that are very similar to the wild-type strain. The *tel1Δ yku80Δ* double mutant shows mild abnormalities; however, the *cdc73Δ yku80Δ* double mutant has significant changes and bands with somewhat reduced intensity, while no bands are visible in the *cdc73Δ tel1Δ yku80Δ* triple mutant. Post-senescent survivors have been reported to be unable to enter PFGE gels, likely because of the highly structured recombination intermediates that are used to maintain the telomeres (Wellinger and Zakian, 2012), consistent with the facts that deletion of *RAD52* was able to recover some of the missing bands for the *cdc73Δ yku80Δ* double mutant. These large changes in chromosomal structure seen by PFGE were not observed by WGS, which could be influenced by several facts: (1) changes in telomere structure predominantly affect repetitive sequences for which reads cannot be uniquely mapped; (2) copy-neutral recombination between repetitive elements larger than the average distance between read pairs (~331 - 484 bp; Supplemental Table 1) such as recombination between Ty elements or even full length delta sequences will not be visible; and (3) heterogeneous rearrangements involving a small percentage of the total populations are likely to be

challenging to identify, even if these rearrangements were to involve uniquely mapping regions of the genome.

We also investigated whether or not combining these mutations have a synergistic effect in preventing telomeres from transcriptionally silencing adjacent genes, which has been termed the telomere position effect (TPE) (Gottschling et al., 1990). TPE was measured in strains containing a telomere-proximal *URA3* gene that causes sensitivity to 5FOA when de-repressed. Consistent with previous results, we found that deletion of *YKU80* caused significant defects in silencing the telomere-proximal *URA3* relative to wild-type cells (Figure 2.8d), (Boulton and Jackson, 1998) while deletion of *CDC73* showed a milder defect and deletion of *TEL1* had minimal effect. The *cdc73Δ yku80Δ* and *cdc73Δ tel1Δ* double mutant strains, however, had a large synergistic increase in sensitivity to 5FOA, which is indicative of a greater perturbation of telomeric structure, consistent with the results from the telomere Southern blots and the PFGE.

***CDC73* suppresses genome instability by supporting transcription elongation and promoting telomerase function.** Mutations affecting subunits of the Paf1 complex have been implicated in the formation of R-loops (Wahba et al., 2011), and *thp2* mutations cause increased R-loops at telomeres, likely involving TERRA transcripts, and give rise to Exo1-mediated telomere resection (Pfeiffer and Lingner, 2012; Pfeiffer et al., 2013). Given the telomeric dysfunctions caused by deletion of *CDC73*, we measured TERRA levels in *cdc73Δ* single and double mutants (Figure 2.9a). One set of TERRA probes monitored TERRA expressed from 6 different telomeres that contain subtelomeric Y' elements (6* Y': from 8L, 8R, 12L, 12R, 13L, and15R), and the other set of TERRA

probes were specific to TERRA expressed from the telomere at the left arm of chromosome 15, which contains only X-elements. The 6* Y' TERRA elements were largely unaffected by the deletion of *CDC73*, *TEL1*, or *YKU80* with the exception of an increase in expression in the *cdc73Δ yku80Δ* mutant strain and a further increase with the overexpression of RNase H1. In contrast, the expression of the chromosome 15 L TERRA was increased in the *yku80Δ* single mutant strain, but not the *cdc73Δ yku80Δ* double mutant strain. Remarkably, overexpression of RNase H1 increased TERRA levels in the *cdc73Δ* single mutant and the *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutant strains, raising the possibility that RNA:DNA hybrids may also negatively affect the regulation of the TERRA levels.

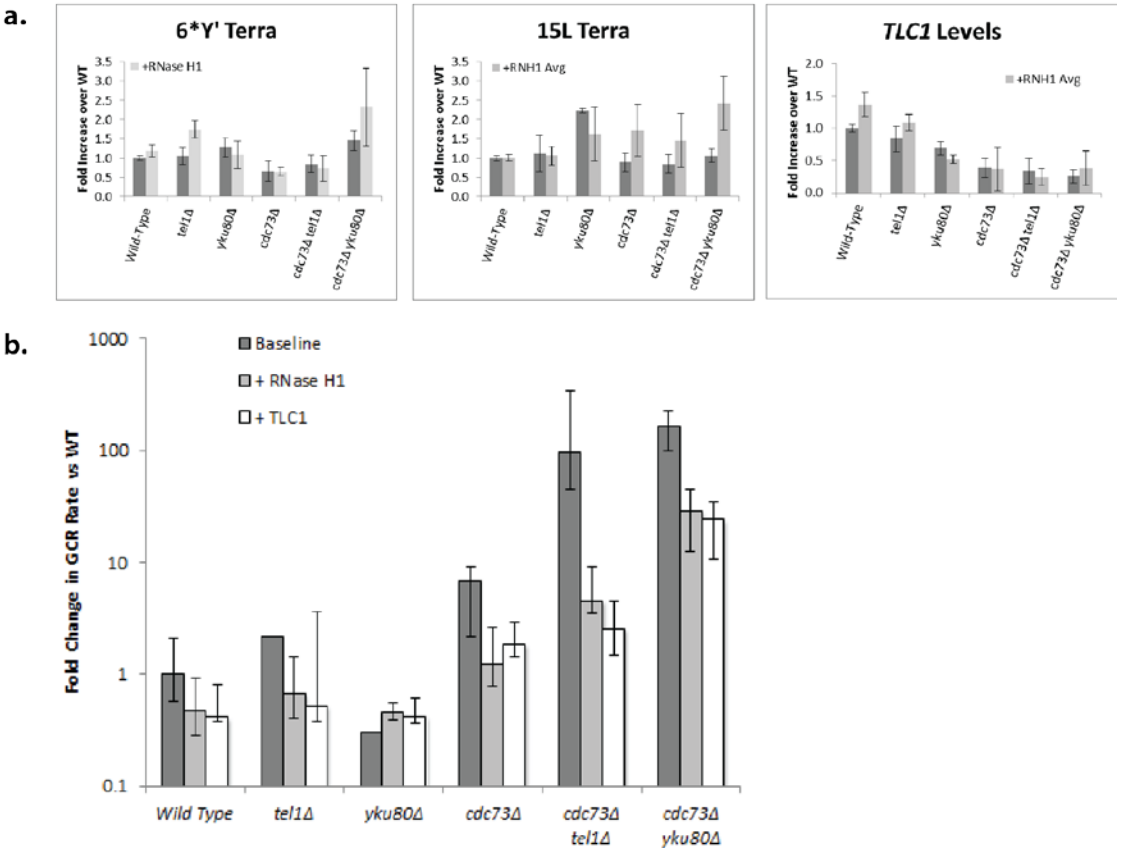


Figure 2.9. R-loop formation and Telomerase Defects Contribute to Genome Instability in a *cdc73Δ* background.

a. TERRA levels are increased in a *cdc73Δ* background upon RNase H1 overexpression. TERRA transcribed from telomeres 15L and 6*Y', and *TLC1* RNA levels were quantified by qRT-PCR. Measurements were done in triplicate and RNA levels relative to wild-type of at least three independent biological replicates were normalized against actin with standard deviations as shown. **b.** Overexpression of RNase H1 and *TLC1* suppress GCR rates in a *cdc73Δ* background. *cdc73Δ*, *tel1Δ*, and *yku80Δ* single and double mutants were reconstructed in the uGCR assay and then transformed with either pCM184 RNH1, which overexpresses RNase H1, or pVL2679, which overexpresses *TLC1*. GCR rates were measured by fluctuation method of nine independent isolates, with 95% confidence intervals as shown.

We next tested the effect of overexpressing RNase H1 and *TLC1* on genome instability in *cdc73Δ* single and double mutants. We reconstructed the single and double mutants in strains containing the uGCR assay (Figure 2.1a), as this assay does not generate GCRs that are promoted by repetitive sequences within the breakpoint region.

As with the dGCR and sGCR assays, we observe a synergistic increase in GCR rates for the *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutants relative to the respective single mutants (Figure 2.9b). Interestingly, overexpression of RNase H1 and *TLC1* caused a significant reduction in the GCR rate, but neither completely abrogated the synergistic interactions seen in the double mutants. We confirmed by qRT-PCR that the effect of RNase H1 overexpression was not due to a restoration of *TLC1* levels in the *cdc73Δ* single and double mutants (Figure 2.9a).Taken together, these data suggest that loss of *CDC73* promotes genome instability by both increased formation of R-loops and partial defects in telomerase due to reduced *TLC1* levels. The possibility that at least some of the R-loops leading to genome stability are localized to telomeric regions would be consistent with the fact that overexpression of RNase H1 partially reduces the GCR rates of the *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutant strains without restoring the levels of *TLC1*.

DISCUSSION

We have demonstrated here a thorough analysis of the genetic interactions by which *CDC73* suppresses genome instability. Using a screen to identify genes that suppress gross chromosomal rearrangements, we found the loss of *CDC73* synergizes with genes that function in telomere maintenance, and we demonstrated in three different GCR assays that *cdc73Δ* has a robust and consistent synergistic interaction with *tel1Δ* and *yku80Δ*. Using a combination of PCR analysis and next-generation sequencing, we showed that the preponderance of GCRs formed in a *cdc73Δ* background rely on

homologous recombination and not *de novo* telomere addition. This is due to the fact that *cdc73Δ* causes defects in telomere homeostasis, as demonstrated by a loss in telomeric silencing and altered chromosome structure due to shortened telomeres. Lastly, we demonstrated the synergistic increase in genome instability between *cdc73Δ* and *tel1Δ* and *yku80Δ* is caused by synergistic telomere maintenance defects, primed by both insufficient telomerase function and an increased accumulation of RNA:DNA hybrids.

By Southern blot analysis for telomere length, we discovered that *cdc73Δ tel1Δ*, *cdc73Δ yku80Δ*, and *cdc73Δ tel1Δ yku80Δ* double and triple mutants are capable of forming post-senescent survivors. It has been previously demonstrated that cells rely on *RAD52* to maintain post-senescent viability in telomerase null strains (Lundblad and Blackburn, 1993; Teng and Zakian, 1999), but we were able to generate *cdc73Δ tel1Δ rad52Δ* and *cdc73Δ yku80Δ rad52Δ* triple mutants. These cells are very sick and have drastically shortened telomeres but are still viable, indicating that telomerase is still functional. The fact that the *tel1Δ yku80Δ* double mutant enters senescence, even though *tel1Δ* and *yku80Δ* single mutants maintain stable (but shorter) telomeres suggests a non-epistatic relationship between these genes. And, in the case of the *cdc73Δ tel1Δ yku80Δ* triple mutant, even freshly sporulated isolates showed a pattern consistent with post-senescence. We were unable to generate a *cdc73Δ tel1Δ yku80Δ rad52Δ* quadruple mutant, suggesting telomerase is completely dysfunctional in this genotype. Altogether, these data imply that *cdc73Δ*, *tel1Δ*, and *yku80Δ* each represent separate hits to telomerase function. Thus, in the *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutants, cells show a preference for maintaining viable telomere structures by activating *RAD52*

dependent recombination pathways, or are forced to rely on telomerase that has drastically reduced functionality. This is consistent with the next-generation sequencing data, which a showed preference for homologous recombination over *de novo* telomere addition for GCRs formed in a *cdc73Δ* background. Furthermore, the drastically increased proportion of GCRs in a *cdc73Δ* background that contain inverted duplications on ChrV L may reflect that the telomere maintenance defect also extends to a loss of perinuclear tethering and chromosomal clustering, so that during repair of a DSB, the search for homologous sequences shows a preference for sites on the same chromosome. The additive defect of telomere maintenance in *cdc73Δ* double mutants that have high genome instability is also reflected in the fact that *cdc73Δ* shows a synergistic increase in GCR rate with mutations in the *SIR2/3/4* complex, but not the *RIF1/2* complex; loss of *SIR2/3/4* results in telomere deprotection, as evidenced by loss of telomeric silencing and telomere perinuclear localization (Moretti et al., 1994; Palladino et al., 1993), whereas loss of *RIF1/2* actually causes increased telomere length and improves telomeric silencing (Wotton and Shore, 1997).

As a member of the Paf1 Complex, Cdc73 has been shown to play a role in promoting transcription elongation and histone modification (Jaehning, 2010; Krogan et al., 2002, 2003; Nordick et al., 2008; Tomson and Arndt, 2013). To determine if the latter is responsible for Cdc73's role in suppressing genome instability, we analyzed data from the large scale screen described in the introduction (Putnam *et al*, manuscript submitted). The Set1 Complex, *SET2*, and *DOT1* encode histone methyltransferases that function downstream of the Paf1 complex, but loss of these genes do not cause a significant

increase in GCR rate as single mutants or as double mutants with *yku80Δ*. This suggests that histone modification does not play a major role in how Cdc73 suppresses genome instability. However, it is interesting to note that Paf1C-mediated telomeric silencing relies on Dot1 (Krogan et al., 2003; Ng et al., 2002), but the lack of a synergistic increase in GCR rate of a *dot1Δ yku80Δ* double mutant suggests loss of telomeric silencing may simply be an additional biomarker of altered telomere structure, rather than being causally related to increased genome instability.

Cdc73 also plays a role in promoting transcription elongation and 3' end mRNA maturation (Nordick et al., 2008; Sheldon et al., 2005), and it has been demonstrated that defects in mRNA processing, as seen in THO complex mutants, can lead to the formation of RNA:DNA hybrids, or r-loops, which promote genome instability, as they are recombinogenic and can block incoming DNA replication forks (Bermejo et al., 2012; Chan et al., 2014; Chávez and Aguilera, 1997; Gómez-González et al., 2011). THO mutants also cause the accumulation of RNA:DNA hybrids at sites of telomeric repeat-containing RNA (TERRA) transcription, resulting in Exo1-dependent resection of telomeres (Pfeiffer et al., 2013). We found that overexpression of RNase H1, an enzyme that specifically resolves RNA:DNA hybrids (Santos-Pereira et al., 2013; Wahba et al., 2011), causes a mild increase in TERRA levels in *cdc73Δ* single and double mutants, and an abrogation of the synergistic increase in GCR rates seen in the *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutants, suggesting the defects in RNA processing in a *cdc73Δ* background promote the formation of recombinogenic r-loops and may also involve the accumulation of RNA:DNA hybrids at subtelomeric sites. Azzalin and colleagues have

hypothesized that increased RNA:DNA hybrids on the leading strand of telomeres could stimulate break-induced replication by exposing C-rich ssDNA patches on the lagging strand that would provide a template for invasion by the G-overhang of another chromosome end (Arora et al., 2014). Alternatively, unresolved hybrids may cause replication fork arrest, which generates structures prone to engage in homologous recombination. Either scenario provides a possible explanation for the increased genome instability seen in *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutants and the dramatic decrease in GCR rate upon overexpression of RNase H1. We also demonstrated that *TLC1* overexpression causes a significant but not complete reduction of GCR rates in the *cdc73Δ tel1Δ* and *cdc73Δ yku80Δ* double mutants. This implies that the recombinational pathways that are activated in order to maintain telomeres in these genotypes cause sufficient genome instability to lose the *URA3*/*CAN1* cassette that is ~25 kb away from the left telomeric arm of Chr V. Overexpression of *TLC1*, and thus partial restoration of telomere length, likely abrogates this telomeric hyperrecombination, but the GCR rate is still increased ~10 fold over wild-type, suggesting the genome instability is not merely due to insufficient telomerase activity. This is consistent with previous findings from our laboratory that *est* mutants that are post-senescent and rely on recombination to maintain viable telomeres actually cause decreased GCR rates in the uGCR assay and no change in the dGCR assay (Myung et al., 2001b; Putnam et al., 2014).

The suppression of GCR rates by overexpression of RNase H1 is not simply due to a restoration of *TLC1* levels, and the fact that both RNase H1 and *TLC1*overexpression suppress GCR rates suggests *CDC73* is unique in that it affects both r-loop formation and

telomerase function. It is possible these two functions are linked, as it has been shown that RNA:DNA hybrids can form at subtelomeric sites and TERRA transcription is stimulated at shortened telomeres to aid in the recruitment of telomerase specifically to that telomere (Cusanelli et al., 2013; Pfeiffer et al., 2013). Interestingly, contrary to what has been demonstrated in *thp2Δ*, we found that overexpression of RNase H1 actually resulted in an increase in TERRA levels in a *cdc73Δ* background. This raises the possibility that in this case RNase H1 overexpression suppresses genome instability not by increasing overall telomerase levels, but by increasing TERRA levels that allow for the more efficient recruitment and activation of endogenous telomerase. In addition to the studies and data described above, it has also been shown that artificial overexpression of TERRA actually stimulates telomere shortening in *cis* (Pfeiffer and Lingner, 2012; Pfeiffer et al., 2013). The links between TERRA levels, telomere shortening and genome instability is clearly quite complex, and it will be interesting to determine how TERRA regulation affects telomerase function and exactly how telomere shortening promotes genome rearrangements. It is particularly important to distinguish the roles of overall TERRA levels versus RNA:DNA hybrid accumulation. The implications for human disease also warrant further investigation. TERRAs were first identified in human cancer cell lines (Azzalin et al., 2007) and have been shown to play a role in alternative lengthening of telomeres (ALT), a telomerase-independent pathway used by ~15% of cancers to maintain viable telomeres (Arora et al., 2014). ALT relies on homologous recombination between telomeric sequences, similar to post-senescent yeast cells. The data shown here may provide insights in to how the human homolog of *CDC73* may

function as a tumor suppressor, and it will be interesting to analyze whether loss of *hCDC73* also contributes to genome instability due to defects in telomere maintenance and increased r-loop formation.

METHODS

Construction and propagation of strains and plasmids. YPD and synthetic drop-out media for propagation of strains have been previously described (Chen and Kolodner, 1999). GCR assays were performed using derivatives of RDKY7635 (*MATalpha hom3-10 ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 lyp1::TRP1 cyh2-Q38K iYFR016::*P_{MFA1}*-LEU2 can1::*P_{LEU2}*-NAT yel072w::CAN1-URA3)*, which contains the dGCR assay, RKDY7964 (*MATalpha hom3-10 ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 lyp1::TRP1 cyh2-Q38K iYFR016::*P_{MFA1}*-LEU2 can1::*P_{LEU2}*-NAT yel068c::CAN1-URA3)*, which contains the sGCR assay, and RDKY6677 (*MATalpha ura3-52 leu2Δ1 trp1Δ63 his3Δ200 lys2ΔBgl hom3-10 ade2Δ1 ade8 can1::hisG yel068c::CAN1/URA3 iYEL072W::hph)*, which contains the uGCR assay, as previously described (Putnam et al., 2009; Putnam *et al*, manuscript submitted). Mutant derivatives of these strains were constructed using standard PCR-based gene disruption methods or mating to strains containing mutations as described (Chen and Kolodner, 1999; Janke et al., 2004). The RNase H1 overexpression plasmid pCM184RNH1 was a gift from the Andrés Aguilera (Santos-Pereira et al., 2013) and the *TLC1* overexpression plasmid pVL2679 was a gift from Victoria Lundblad. Both plasmids have TRP as the selection marker, and *cdc73Δ* single and double mutants were transformed with the plasmids and selected on –TRP

plates. For GCR rate determination, transformants were cultured overnight in –TRP liquid media.

Systematic double mutant generation. We crossed a strain containing the dGCR assay and a *cdc73Δ* mutation against 638 strains from the *S. cerevisia*e deletion collection and obtained haploid progeny by germinating spores generated from the resulting diploids, as previously described (Putnam *et al*, manuscript submitted).

DNA content measurement by flow cytometry. *cdc73Δ* double mutants generated in the dGCR assay background by a modified synthetic genetic array were also screened by flow cytometry for DNA content to exclude diploid isolates from the analysis. Briefly, double mutants in a 96 well format were cultured in liquid YPD overnight. Then, 10 µL of the overnight culture was added to 190 µL of fresh YPD and the cells were incubated in a 30^{o}C shaker for 3 hours. Cells were washed and resuspended in 60 µL of dH_2O and fixed with 140 µL of cold absolute ethanol. Cells were then sonicated and resuspended in 150 µL of 50 mM sodium citrate with 1 mg/mL Proteinase K (Sigma-Aldrich) and 25 mg/mL RNase A (Sigma-Aldrich) and incubated at 37^{o}C overnight. Then, the cells were washed and resuspended in 100 µL of 50 mM sodium citrate containing 1 uM Sytox Green (Life Technologies). The cells were then measured with a BDS LSR II flow cytometer (with a high throughput sampler) at The Scripps Research Institute flow cytometry core facility. Data was analyzed using FlowJo v10 (Zhu et al., 2012) to distinguish between haploid and diploid DNA content.

Patch tests. Double mutants generated from this cross were patched on non-selective media and then replica plated to selective plates containing the drugs

canavanine and 5-fluoroorotic acid (5FOA). Each papilla on the selective plates corresponds to a separate GCR event, and the number of papillae per patch corresponds to the GCR rate of the strain. Each patch was given a score of 0 to 5 based on the number of papillae per patch, with 1 corresponding to wild-type.

GCR rate determination. GCR rates were determined for multiple independent biological isolates using the fluctuation method as previously described (Putnam and Kolodner, 2010).

Telomere position effect assay. The telomere position effect assay was constructed in BY4742 (*MATalpha leu2Δ0 his3Δ1 ura3Δ0 met15Δ0*) by inserting the *URA3* gene onto the left arm of Chromosome VII at the site of *ADH4*. The plasmid pADH4UCA (Gottschling et al., 1990), a gift from the Zakian Lab, contains *URA3* adjacent to the centromere proximal half of *ADH4*. The plasmid was digested with *Sal*I and *Eco*RI and transformed into BY4742. Transformants were selected on –URA drop out plates, and successful transformation was confirmed by PCR, generating the strain RDKY823 Mutant derivatives of this strain were constructed using standard PCR-based gene disruption methods as above. To assay these strains, yeast were cultured overnight in liquid YPD at 30^{o}C. Then, 100 µL of the culture was used to make tenfold serial dilutions, and 1.5 µL were spotted onto complete synthetic medium (CSM), CSM lacking uracil (CSM-URA), and CSM supplemented with 1 mg/L of 5FOA (CSM+5FOA). Plates were incubated at 30^{o}C for 3 days before imaging.

Analysis of GCR Structures. The structures of GCRs generated in strains from the dGCR assay were analyzed by PCR, as previously described (Putnam et al., 2009).

Genomic DNA was prepared from individual isolates using the Purgene kit (Qiagen) and subjected to PCR analysis to categorize the GCRs. The t(V;XIV) and t(V;IV or X) translocations were identified by amplification of the junction region with a chromosome-V-specific primer centromeric to the *HXT13-DSF1* region and a chromosome XIV- or IV/X-specific primer telomeric to the *HXT13-DSF1* homologies on those chromosomes, under conditions where no product was generated with DNA from wild-type strains.

The structures of GCRs generated in strains from the sGCR assay were analyzed by whole genome paired-end sequencing, as previously described (Putnam et al., 2014). Multiplexed paired-end libraries were constructed from 5 µg of genomic DNA purified using the Purgene kit (Qiagen). The genomic DNA was sheared by sonication and end-repaired using the End-it DNA End-repair kit (Epicentre Technologies). Common adaptors from the Multiplexing Sample Preparation Oligo Kit (Illumina) were then ligated to the genomic DNA fragments, and the fragments were then subjected to 18 cycles of amplification using the Library Amplification Readymix (KAPA Biosystems). The amplified products were fractionated on an agarose gel to select 600 bp fragments, which were subsequently sequenced on an Illumina HiSeq 2000 using the Illumina GAII sequencing procedure for paired-end short read sequencing. Reads from each read pair were mapped separately by bowtie version 12.7 (Langmead et al., 2009) to a reference sequence that contained revision 64 of the *S. cerevisiae* S288c genome (Engel et al., 2014), *hisG* from *Samonella enterica*, and the *kanMX4* marker. Chromosomal rearrangements were identified after bowtie mapping by version 6 of the Pyrus suite

(http://www.sourceforge.net/p/pyrus-seq). Briefly, after PCR removal of PCR duplicates, read pairs with 2 uniquely mapping reads were used to generate 2 distributions. The number of times each base pair was read (the 'nread' distribution) was determined for identifying the sequence variants observed a significant number of times, and the number of times each base pair was spanned by a pair of reads (the 'nspan' distribution) was determined to identify the candidate chromosomal rearrangements that were supported by a significant number of read pairs. The data were then analyzed for junction-defining read pairs that indicated the presence of structural rearrangements relative to the reference genome, such as the *cdc73::HIS3* deletion or GCR-related translocations. The junction-sequencing reads were identified from read pairs in which one read could not be mapped and the other read mapped next to the junction-defining read pairs. Sequences of the junctions were generated by *de novo* alignment of the junction-sequencing reads associated with rearrangements defined by statistically significant junction-defining read pairs. The identified rearrangements included all known rearrangements in the strains that could be defined based on the average distance between the read pairs in the library.

Additional screening of GCRs from the sGCR assay was done by PCR to analyze the site of the breakpoint and the presence of an inverted repeat on ChrVL. One PCR pair amplifies directly centromeric to the *can1::*P_{LEU2}*-NAT* locus, while another pair amplifies directly telomeric, allowing a determination of whether the breakpoint occurs at the site of this construct. To determine the inverted repeat, two other primer pairs were used, whereby the first primer amplifies directly centromeric to the *can1::*P_{LEU2}*-NAT* locus while the second primer amplified centromeric to either *YELWdelta2* or *YELWdelta6*.

Telomere Southern blotting. Telomere Southern blots were performed using a modified version of a previously described protocol (Sambrook and Russell, 2006). Genomic DNA was purified from 50 mL overnight cultures using the Purgene kit (Qiagen). 5 µg was aliquoted and digested with *Xho*I (New England Biolabs) in a 50 µL reaction for 2 hr at 37^{o}C. The reaction was stopped by adding 8 µL of loading buffer, and the samples were run on a 8% agarose gel in 5X TBE for 16 hr at 50 V. The DNA in the gel was transferred to Amersham Hybond-XL membranes (GE) by neutral capillary blotting, allowed to run overnight. The DNA was crosslinked to the membrane by UV irradiation in a Stratalinker™ (Stratagene) apparatus at maximum output for 60 seconds. Biotinylated TG probes were purchased from ValueGene. Probe hybridization was performed with ULTAhyb oligo hybridization buffer (Life Technologies) at 42°C for 1 hr. The membrane was then washed extensively and developed with a chemiluminescent nucleic acid detection kit (Life Technologies) and imaged with a Bio-Rad Imager.

Pulse Field Gel Electrophoresis (PFGE). DNA plugs for PFGE were prepared as described (Gerring et al., 1991). Strains were grown to saturation in 50 mL of YPD at 30^{o}C for 3 days. Cell counts were measured by optical density at 600 nm, and 7.5 x 10^{8} cells from each strain were washed and resuspended in 200 µL of 50 mM EDTA, then mixed with 70 uL of 1 M sorbitol, 1 mM EDTA, 100 mM sodium citrate, 5% β-mercaptoethanol, 8 U/mL of zymolase. The cells were then mixed with 330 µL of liquefied 1% ultrapure agarose (Bio-Rad) to prepare multiple 80 mL plugs. The plugs were incubated in 15 mL conical tubes in 750 mL of 10 mM Tris pH 7.5, 500 mM EDTA pH 8, 1% β-mercaptoethanol for 16 hr at 37^{o}C. The plugs were then incubated in 750 mL

10 mM Tris pH 7.5, 500 mM EDTA pH 8, 1% sodium N- lauryl sarcosine, 2% sodium dodecyl sulfate containing 2 mg/ ml Proteinase K (Sigma-Aldrich) for 6 hr at 65^{o}C. Finally, the plugs were washed in 50 mM EDTA pH 8 prior to resolving the chromosomes in a 1% agarose gel run in a CHEF (clamped homogeneous electric field electrophoresis) apparatus in chilled (14^{o}C) 5x TBE (89 mM Tris-borate, pH 8.3, 25 mM EDTA). Electrophoresis was performed using a Bio-Rad CHEF-DRII apparatus at 6 V/cm, with a 60 to 120 s switch time for 24 h. The gels were stained with ethidium bromide and imaged.

RNA isolation and quantitative real-time PCR (qRT–PCR). RNA isolation and qRT-PCR for TLC1 and TERRA RNAs were performed using published techniques (Iglesias et al., 2011; Pfeiffer and Lingner, 2012). Cells were grown in rich liquid medium to an OD600 of 6 to 8. 1 mL samples were used for RNA isolation with the RNeasy kit (Qiagen), with on column DNase I treatment using the RNase-Free DNase Set (Qiagen). 1 mg RNA was reverse transcribed with the iScript cDNA Synthesis Kit (Bio-Rad), which uses random primers. cDNA was diluted 1:10 with d H_2O. qPCR was performed with 2 µL of the dilution in a final volume of 20 µL using the iTaq Universal SYBR Green Supermix (Bio-Rad) in a Bio-Rad CFX96 Touch Real-Time PCR Detection System. Reaction conditions: 95^{o}C for 10 min, 95^{o}C for 15 sec, 50^{o}C for 1 min, 40 cycles. Primer concentrations and sequences were the same as previously described (Pfeiffer and Lingner, 2012).

ACKNOWLEDGEMENTS

This chapter is an adaptation of an original document that is being prepared for publication. As of May 19th 2015, the authorship and tentative title of this manuscript are: Nene RV, Putnam CD, Kolodner RD. Cdc73 Suppresses Genome Instability by Mediating Telomere Homeostasis and RNA:DNA Hybrid Formation. The dissertation author was the primary author of this material and contributed to the conception and design of experiments, execution of experiments, data analysis, and manuscript writing. The dissertation author would like to thank Christopher S. Putnam and Richard D. Kolodner for assisting with the conception and design of the research, data analysis, manuscript writing and mentorship. The authors would also like to thank V. Zakian, V. Lundblad, and A. Aguilera for materials. Rahul Nene is supported by NIH F30 Grant CA177240-01 and the University of California, San Diego Medical Scientist Training Program T32 GM007198-4. This work was also supported by NIGMS Systems Biology Center Grant GM085764, NIH R01 Grant GM26017 and the Ludwig Institute for Cancer Research to Richard D. Kolodner and Christopher S. Putnam.

Chapter 3:

Identification of a Domain of Cdc73 That is Necessary and Sufficient for The Suppression of Genome Instability

By

Rahul V. Nene[1], Christopher D. Putnam[1,2], Christopher S. Campbell[1,2,3], Arshad Desai[1,2,3], and Richard D. Kolodner[1,3,4,5]

From

Ludwig Institute for Cancer Research[1], Departments of Medicine[2] and Cellular and Molecular Medicine[3], Moores-UCSD Cancer Center[4] and Institute of Genomic Medicine[5], University of California School of Medicine, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0669

INTRODUCTION

Defects in transcription are an increasingly well appreciated source of genome instability. The precise mechanism by which this occurs is not completely understood, but may involve collisions with the replication machinery, the formation of RNA:DNA hybrids, and/or the phenomenon of transcription associated replication (Aguilera, 2002; Aguilera and García-Muse, 2012; Kim and Jinks-Robertson, 2012). Recent studies have implicated *CDC73* as playing a role in chromosomal instability and hyperrecombination (Fan et al., 2001; Wahba et al., 2011; Yuen et al., 2007), and we also identified it in a large scale screen of genes that suppress genome instability (Putnam *et al*, manuscript submitted; See Chapter 1). *CDC73* is a member of the Paf1 complex (Paf1C), which binds to and modifies the activity of RNA polymerase II during transcription. In the budding yeast *Saccharomyces cerevisiae*, the proteins Paf1 and Cdc73 were identified as forming a complex with RNA polymerase II independently of Srbps (Shi et al., 1997; Wade et al., 1996) and further analysis led to the identification of Rtf1, Ctr9 and Leo1 as additional components (Koch et al., 1999; Krogan et al., 2002; Mueller and Jaehning, 2002; Squazzo et al., 2002). This complex has been implicated in a variety of cellular processes, including transcription elongation, 3'-end mRNA maturation, and histone modification (Jaehning, 2010; Krogan et al., 2002, 2003; Nordick et al., 2008; Tomson and Arndt, 2013). It is well conserved among eukaryotes, from budding yeast and *Drosophila* to zebrafish and humans (Newey et al., 2009).

Although the complex contains 5 subunits, only for the human homolog of *CDC73* is there strong evidence for a role in preventing tumorigenesis. The function of

CDC73 as a tumor suppressor has been linked to breast, renal, and gastric cancers, is often mutated in parathyroid cancer, and germline mutations of *CDC73* cause the cancer susceptibility syndrome hyperparathyroidism-jaw tumor syndrome (HPT-JT) (Newey et al., 2010). However, little is known about how *CDC73* functions as a tumor suppressor and how that activity may be distinct from the other members of the Paf1C. Much of what is known about DNA repair and its compensatory pathways has been elucidated in the yeast *Saccharomyces cerevisiae,* and these pathways are highly conserved throughout eukaryotes (Aggarwal and Brosh, 2012). We recently completed a comprehensive study in *S. cerevisiae* that identified the genetic interactions by which *CDC73* suppresses genome instability, and characterized the synergistic interactions it shows with defects in telomere homeostasis (Nene *et al*, manuscript in preparation; See Chapter 2). Our goal here is to determine what role the Paf1C as whole plays in suppressing genome instability. We first demonstrate to what extent loss of each member of the complex affects genome stability and how this correlates with alterations in other Paf1C functions. We then focused on Cdc73 and identified the minimal region of the protein that is necessary and sufficient for suppression of genome instability. Finally, we demonstrate how this provides insights into how the complex comes together for normal function. We hope these results will shed light on how the human homolog may play a role in tumorigenesis.

RESULTS

Genes encoding Paf1 complex subunits suppress genome instability. We have previously demonstrated that *cdc73Δ* causes a moderate increase in genome instability alone, but shows strong synergistic interactions when combined with deletions of *TEL1* and *YKU80* (Nene *et al*, manuscript in preparation). As Cdc73 is a subunit of the Paf1 complex, we tested whether or not deletion of genes encoding other Paf1 complex subunits might also play roles in suppressing genome instability (Figure 1a). To test this, we used a gross chromosomal rearrangement (GCR) assay, which functions by measuring the rate of rearrangements on the nonessential terminal arm of Chromosome V, as detected by the loss of the genetic markers *CAN1* and *URA3* (Chen and Kolodner, 1999; Putnam and Kolodner, 2010; Putnam et al., 2009). The *cdc73Δ* and *paf1Δ* single mutant strains have a substantial increase in the GCR rate, whereas the *ctr9Δ*, *rtf1Δ*, and *leo1Δ* single mutants showed a somewhat lower GCR rate than either *cdc73Δ* or *paf1Δ*, but a rate that was still significantly increased over wild-type (Table 3.1). The *cdc73Δ, paf1Δ, ctr9Δ,* and *rtf1Δ* mutations caused a synergistic increase in GCR rate when combined with the *yku80Δ* mutation; the GCR rates of the double mutants were highest for the *cdc73Δ yku80Δ* and the *paf1Δ yku80Δ* double mutant strains. Similarly, the *cdc73Δ, paf1Δ*, and *ctr9Δ* mutations caused a synergistic increase in GCR rate when combined with the *tel1Δ* mutation. Deletion of *LEO1* did not synergize with either the *yku80Δ* or *tel1Δ* mutation. Thus, all of the known genes encoding Paf1 complex subunits suppress genome instability to varying levels, and most act to suppress genome instability in strains with defects in telomere maintenance caused by deletion of *TEL1* or *YKU8*

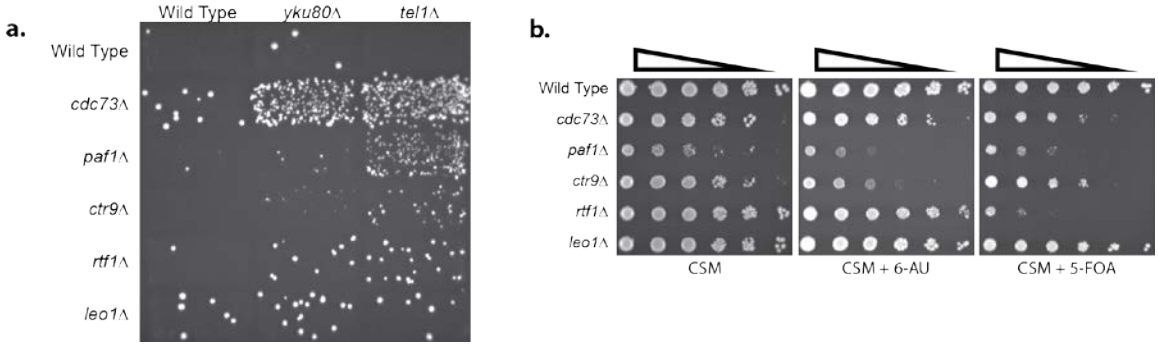


Figure 3.1. Paf1 Complex Components Play Varying Levels in the Suppression of Genome Instability, Transcription Elongation, and Telomeric Silencing.
a. Patch tests of mutations of each Paf1 Complex component as single mutants or double mutants with *yku80Δ* and *tel1Δ*. Each papilla corresponds to a GCR event and the greater the number of papillae per patch correlates to an increased GCR score. Correcting for variations in growth rate, *paf1Δ* and *ctr9Δ* show strong synergistic interactions, similar to *cdc73Δ*. **b.** Dilution analysis assessing defects in transcription elongation, as seen by sensitivity to the drug 6-azauracil, and defects in telomeric silencing using the telomere position effect assay, as seen by sensitivity to the drug 5-FOA. 10-fold dilutions of log growth cells were spotted to non-selective complete synthetic medium (CSM), CSM + 6-azauracil (50 µg/mL), and CSM + 5-FOA (1 mg/mL) and incubated at 30°C for 4 days.

Table 3.1. Paf1 Complex Components Play Varying Roles in GCR Suppression, Transcription Elongation, and Telomere Silencing.

	Single Mutant dGCR Rate Fold Increase	*yku80Δ* Double Mutant dGCR Rate Fold Increase	*tel1Δ* Double Mutant dGCR Rate Fold Increase	6-Azauracil Resistance	Telomere Silencing
Wild Type	**1.0**	**3.8**	**3.9**	+++	+++
***cdc73*Δ**	**33**	**152**	**222**	–	– –
***paf1*Δ**	**21**	**214**	**246**	– – –	– – –
***ctr9*Δ**	**8.1**	**43**	**67**	– – –	– –
***rtf1*Δ**	**4.4**	**52**	**1.6**	+	– – –
***leo1*Δ**	**3.7**	**4.4**	**2.9**	+	+

The role of Paf1 complex genes in suppressing genome instability only partially correlates with transcriptional effects. To study the roles of the genes encoding subunits of the Paf1 complex on suppressing genome instability as compared with previously identified functions of the complex, we also assayed the deletion strains for sensitivity to 6-azauracil, which depletes cellular levels of GTP and has been used to identify defects in transcriptional elongation (Tansey, 2006), and for loss of silencing of a telomere-proximal *URA3* gene (Figure 3.1b) (Gottschling et al., 1990). Loss of *PAF1* or *CTR9* caused a much stronger sensitivity to 6-azauracil than loss of *CDC73*, whereas loss of *RTF1* or *LEO1* had little or no 6-azauracil sensitivity. In contrast, loss of *PAF1* and *RTF1* caused substantial derepression of the telomeric *URA3* gene, consistent with previous observations (Krogan et al., 2003). Loss of *CDC73* and *CTR9* also caused defects in telomere silencing, but these defects were not as severe as in *paf1Δ* and *rtf1Δ* strains. *LEO1*, on the other hand, did not appear to cause any defect in telomeric silencing. Comparison of the results of deleting genes encoding subunits of the Paf1 complex on genome instability with 6-azauracil sensitivity and telomeric silencing indicate that the individual Paf1 complex subunits play roles of differing importance in each of these activities (Table 3.1). Paf1 appears to be the most important subunit for all three activities. Cdc73 plays a major role in genome stability and a minor role in resistance to 6-azauracil and telomeric silencing. Ctr9 primarily acts in resistance to 6-azauracil; Rtf1 primarily acts in telomeric silencing, and Leo1 plays a minor or no role in all of the Paf1 complex functions assayed here.

Deletion analysis of Cdc73 identified an internal region required for function. No enzymatic activity has been reported for any of the Paf1 complex subunits, which has been used to argue that the Paf1 complex primarily functions as a scaffold (Jaehning, 2010). The C-terminal domain of *S. cerevisiae* Cdc73 (residues 230-393) is the region with the greatest conservation among eukaryotes and has a GTPase-like fold that lacks GTPase activity (Amrich et al., 2012; Chen et al., 2012). The remainder of Cdc73 (residues 1-229) does not have any known structural homologs and was not predicted to adopt a known fold by the Phyre2 protein prediction server (Kelley and Sternberg, 2009). We therefore undertook a deletion analysis to characterize the regions of the protein required for Cdc73 function in suppressing genome instability, suppressing sensitivity to 6-azauracil, and mediating telomeric silencing (Figure 3.2).

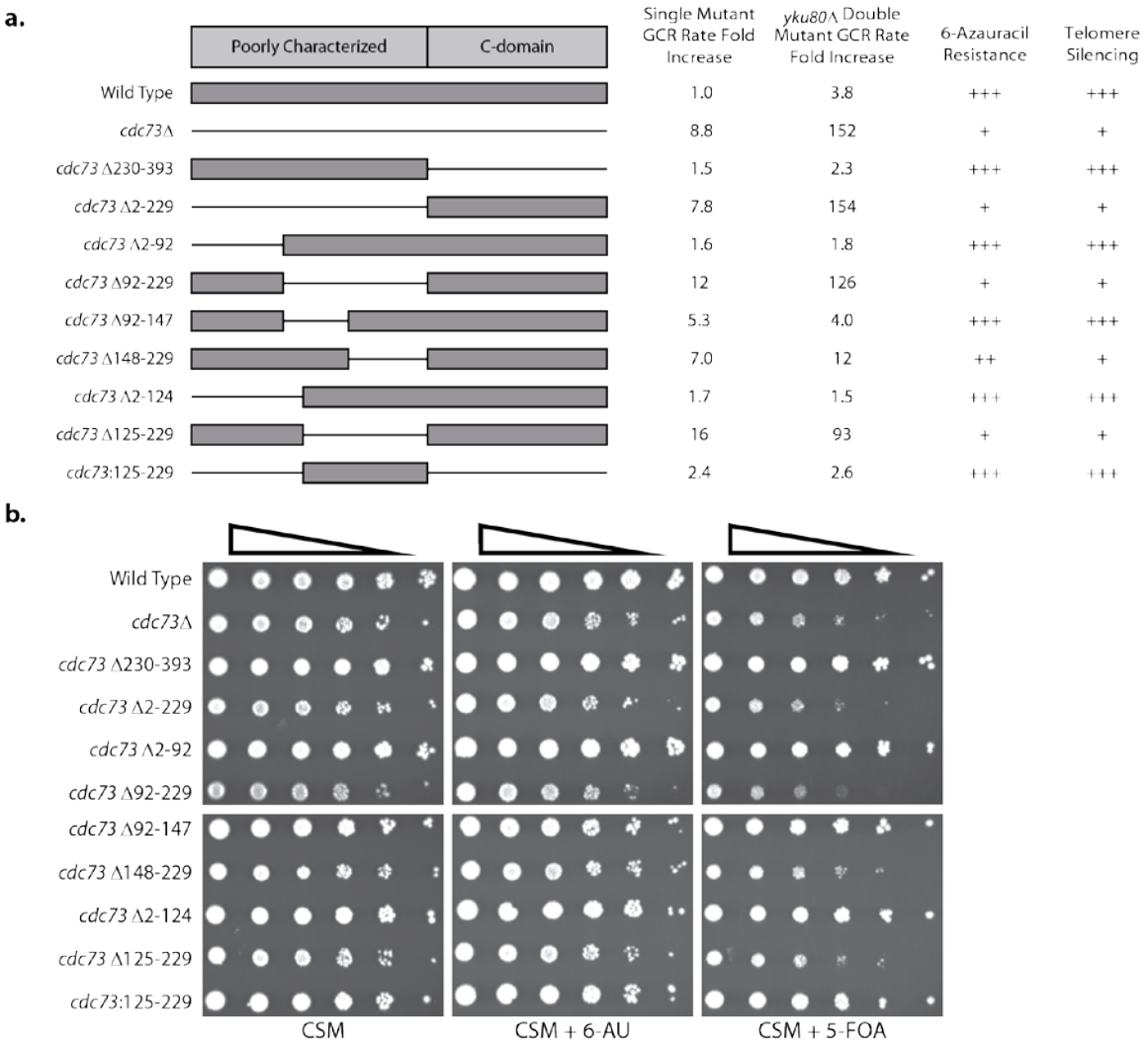


Figure 3.2. Cdc73 Residues 125-229 Are Necessary and Sufficient for Its Function.
a. The various Cdc73 mutant constructs are shown with the corresponding regions that have been deleted. Each construct was measured for GCR rates as a single mutant or double mutant with *yku80Δ*, and tested for functionality of transcription elongation, as determined by resistance to 6-azauracil, and intactness of telomeric silencing. Loss of residues 125-229 is the minimal most deletion that recapitulates a null mutation, while a construct that expresses only residues 125-229 appears to function as well as a wild-type. **b.** For each Cdc73 mutant construct, dilution analysis was used to asses defects in transcription elongation, as seen by sensitivity to the drug 6-azauracil, and defects in telomeric silencing using the telomere position effect assay, as seen by sensitivity to the drug 5-FOA. 10-fold dilutions of log growth cells were spotted to non-selective complete synthetic medium (CSM), CSM + 6-azauracil (50 µg/mL), and CSM + 5-FOA (1 mg/mL) and incubated at 30°C for 4 days.

We initially replaced the wild-type chromosomal copy of *CDC73* with *cdc73* mutations that deleted the conserved C-terminal domain (*cdc73Δ230-393*) or the remainder of the protein (*cdc73Δ2-229*). In contrast with the previous results indicating that the deletion of the Cdc73 C-terminal domain caused defects in transcription elongation (Amrich et al., 2012), we found that *cdc73Δ230-393* was similar to wild-type *CDC73* in GCR suppression, sensitivity to 6-azauracil, and telomeric silencing. Consistent with this, *cdc73Δ2-229* was similar to *cdc73Δ* in all measured assays. We therefore performed a deletion analysis of the N-terminal region based on smaller regions of conservation between *S. cerevisiae* and other fungi. We defined a minimal deletion *cdc73Δ125-229* that caused similar fold-increases in the GCR rate as a *cdc73Δ* single mutant (16 fold increase vs. 8.8 fold increase) and similar increases as a *cdc73Δ* when combined with *yku80Δ* (93 fold increase vs. 152 fold increase) (Figure 3.2a). The *cdc73Δ125-229* mutation also caused increased sensitivity to 6-azauracil and reduced telomere silencing that were similar to that caused by *cdc73Δ* (Figure 3.2b). The effect of the *cdc73Δ125-229* mutation could have been due to either the loss of a portion of Cdc73 required for Paf1 complex activity or due to the generation of a non-functional protein that simply had folding defects. We therefore constructed an allele that encoded only residues 125-229 (*cdc73:125-229*) and found that this allele was fully functional. In addition to defining a 104 residue region required for Cdc73 function, the deletion analysis also indicated that *cdc73* mutations with defects in suppressing genome instability also had defects in suppressing sensitivity to 6-azauracil and in maintaining

telomeric silencing, suggesting that all three defects caused by the *cdc73Δ* mutation may have a common mechanistic basis.

Functional Cdc73 truncations are still able to associate with the Paf1 complex. As 104 residues of Cdc73 were sufficient for mediating several functions of Cdc73, we tested to see if this region was also sufficient for Cdc73 recruitment to the Paf1 complex. Strains encoding Paf1, Rtf1, Ctr9, or Leo1 with a C-terminal 9-myc tag and full-length Cdc73, Cdc73Δ230-393, Cdc73Δ2-124, Cdc73Δ125-229, or Cdc73:125-229 with a C-terminal Venus tag (a GFP variant) were generated by mating singly tagged strains followed by sporulation and isolation of doubly tagged strains. Cell lysates were prepared from log-phase cells and immunoprecipitated with anti-myc antibodies, and the presence of Cdc73 in the immunoprecipitate was determined by Western blotting using anti-GFP antibodies. Full-length Cdc73 co-immunoprecipitated in pulldowns with myc-tagged Paf1, Rtf1, Ctr9, and Leo1 (Figure 3.3). The interaction of full-length Cdc73 with Rtf1 was much weaker than the interaction with the other Paf1 complex subunits, consistent with previous observations of the transient association of Rtf1 with the Paf1 complex in *S. cerevisiae*, Drosophila, and humans (Adelman et al., 2006; Kim et al., 2010; Nordick et al., 2008). It has previously been shown that deletion of *PAF1* or *RTF1* results in fewer molecules per cell of the other complex components (Porter et al., 2005), and we see a similar phenomenon whereby various Cdc73 truncations result in small but noticeable alterations in the abundance of the other subunits. Nonetheless, the functional Cdc73 truncations, Cdc73Δ230-393, Cdc73Δ2-124, and Cdc73:125-229, were still able to associate with Paf1, Ctr9, Leo1, and Rtf1 (Figure 3.3). Interestingly, both the

Cdc73Δ230-393 and Cdc73:125-229 truncation but not full-length Cdc73 or Cdc73Δ2-124 truncation had somewhat reduced binding to Leo1, which suggests that the conserved C-terminus of Cdc73 may play a role in stabilizing Leo1 in the Paf1 complex. In contrast, the non-functional Cdc73 truncation, Cdc73Δ125-229, had substantially reduced binding to each of the other Paf1 complex subunits; residual binding was only detected with Ctr9 and Leo1.

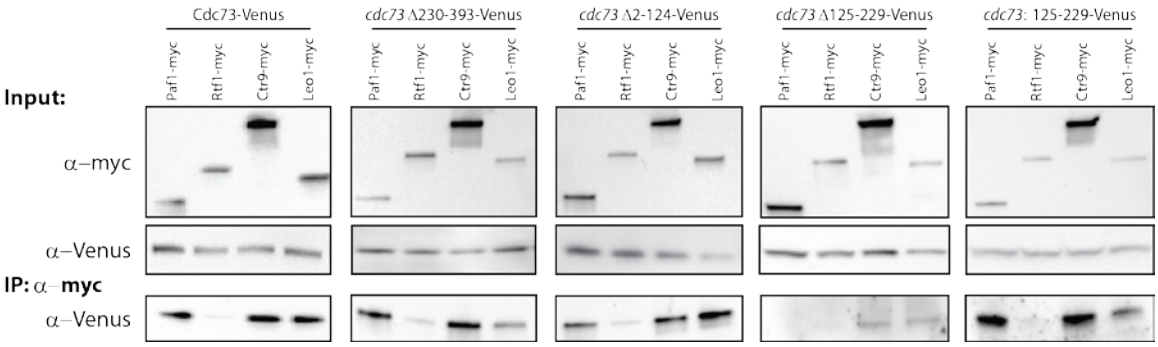


Figure 3.3. Cdc73 Residues 125-229 Are Necessary For Binding to Paf1.
Wild-type or mutant Cdc73 was tagged with C'-terminal Venus (a GFP variant) and each other Paf1 Complex member was tagged with C'-terminal myc. Whole cell lysates were prepared and 1 mg of protein was immunoprecipitated with anti-myc antibody and analyzed by Western blotting for coimmunoprecipitation using an anti-GFP antibody.

Loss of nuclear localization compromises Paf1 complex function. Subunits of the Paf1 complex affect RNA transcription and are localized to the nucleus (Huh et al., 2003; Jaehning, 2010; Mueller and Jaehning, 2002; Porter et al., 2005). We therefore monitored the cellular localization of Cdc73 truncations that were C-terminally tagged with Venus. Full-length Cdc73 and the functional truncations, including the minimal construct Cdc73:125-229, were localized to the nucleus (Figure 3.4a), and we were able

to quantify this by measuring the ratio of nuclear to cytoplasmic fluorescence (Figure 3.4b). In contrast, the non-functional Cdc73 truncation, Cdc73Δ125-229, had diffuse localization in the nucleus and cytoplasm. By quantifying the total fluorescence in the cell, we show that the expression of Cdc73Δ125-229 is only slightly lower than the other constructs, confirming this mutant protein is indeed stable, but cannot maintain nuclear localization (Figure 3.4b). Thus, residues 125-229 of Cdc73 either include a nuclear localization signal or are necessary for association with the subunits of the Paf1 complex that provide nuclear localization to Cdc73.

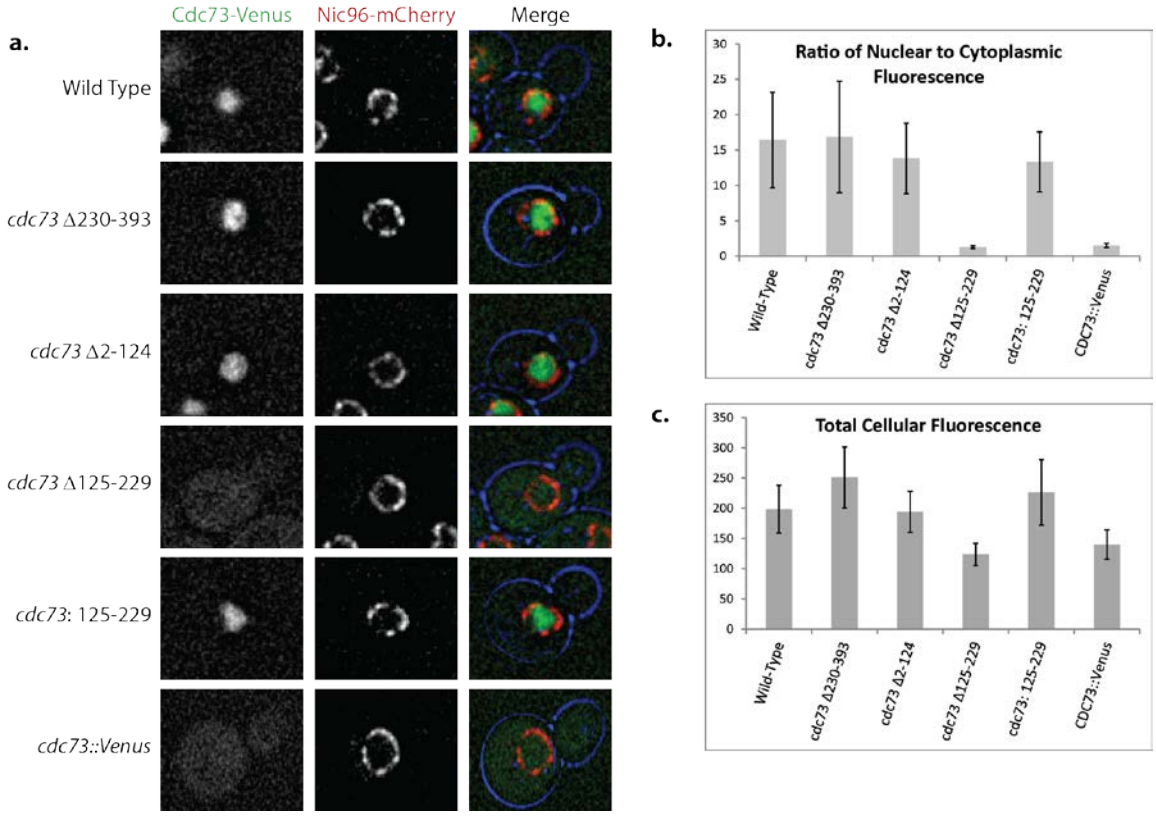


Figure 3.4. Function Cdc73 Constructs Localize to the Nucleus.
a. Wild-type or mutant Cdc73 was tagged with C'-terminal Venus and Nic96 (a member of the nuclear pore complex) was tagged with mCherry. Included as a control is *cdc73::Venus*, where the *CDC73* ORF was replaced by the sequence encoding Venus. Cells were imaged by deconvolution microscopy. **b, c.** Total cellular fluorescence and the ratio or nuclear to cytoplasmic fluorescence for each of the Cdc73 constructs was measured using ImageJ. The data represent averages of at least 20 independent biological isolates, with the standard deviations shown as such.

To determine whether Cdc73 relies on any of other subunits of the Paf1 complex to localize to the nucleus, we determined the localization of Venus-tagged wild type Cdc73 in strains where each of the other genes encoding Paf1 complex subunits was deleted (Figure 3.5a). The mutants showed an enlarged cell morphology with abnormally elongated buds, as previously described (Shi et al., 1997; Watanabe et al., 2009), but Cdc73 was still able to localize to the nucleus, indicating that it does not rely on any one

single Paf1 complex subunit for nuclear localization. We also tested whether any of the other subunits rely on Cdc73 to localize to the nucleus by tagging these subunits with Venus and determining if the localization changes when *CDC73* is deleted (Figure 5b). All four other components, Paf1, Rtf1, Ctr9, and Leo1 were observed in the nucleus in wild-type strains and maintained nuclear localization in the absence of Cdc73, suggesting there are redundancies in maintaining nuclear localization of the complex.

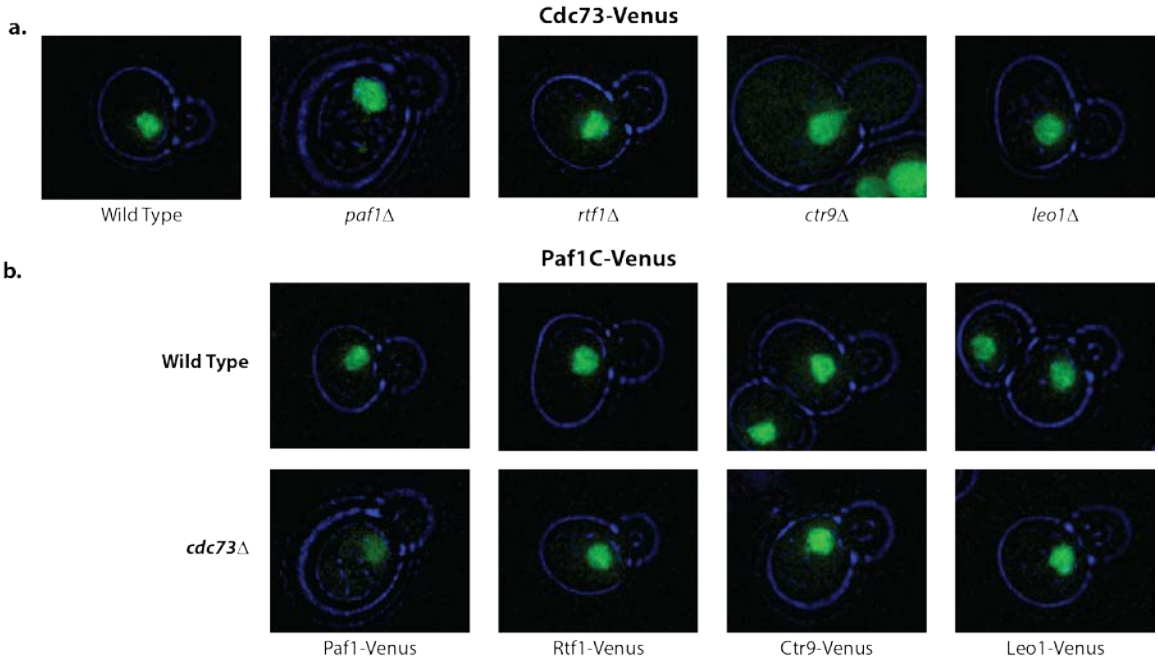


Figure 3.5. The Paf1 Complex Has Redundancies in Maintaining Nuclear Localization. **a.** Cdc73 localizes to the nucleus in the absence of each other Paf1C subunit. Wild-type Cdc73 was tagged with Venus and each other Paf1C component was deleted, then cells were imaged by deconvolution microscopy. **b.** Cdc73 is not required for the nuclear localization of each other Paf1C subunit. Analogous to the experiments in (a), each other component of Paf1C was tagged with Venus and imaged in a wild-type and *cdc73Δ* background.

DISCUSSION

Transcription, and defects therein, is becoming an increasingly well appreciated source of genome instability (Aguilera, 2002; Aguilera and García-Muse, 2012; Kim and Jinks-Robertson, 2012). The Paf1 complex associates with RNA polymerase II and plays a role in transcription elongation and histone medication, and recent studies have implicated subunits of the complex, particularly Cdc73, as playing a role in suppressing chromosomal instability. We have completed here, what is to our knowledge, the first comprehensive analysis of the extent to which each member of the Paf1 Complex plays in the suppression of genome instability and correlated this with previously described functions of the complex. We demonstrated that each subunit plays differing levels of importance to the complex's overall functions, and *PAF1* and *CDC73* are the most important in suppressing genome instability. We next did a deletion analysis of Cdc73 and found that rather than the C-terminus, which is most well-conserved among eukaryotes, a 104 amino acid region in the middle of the protein was necessary and sufficient for Cdc73 function. We found this region is necessary for nuclear localization and likely mediates direct binding to Paf1.

CDC73 is a member of the Paf1 complex and helps mediate transcription elongation, 3'-end mRNA maturation, and histone modification (Jaehning, 2010; Krogan et al., 2002, 2003; Nordick et al., 2008; Tomson and Arndt, 2013). Over the last 15 years, *CDC73* has also been implication in direct repeat hyperrecombination and chromosomal instability (Fan et al., 2001; Wahba et al., 2011; Yuen et al., 2007), and was recently identified by our lab in a large scale screen for suppressors of genome instability (Putnam

et al, manuscript submitted). We conducted a more thorough investigation of the mechanisms by with *CDC73* suppresses genome instability and found that loss of *CDC73* causes defects in telomerase regulation and increased RNA:DNA hybrids, both of which contribute to telomeric, and thus chromosomal, instability (Nene *et al*, manuscript in preparation). We found that *cdc73Δ* shows large synergistic increases in GCR rate with loss of other telomere maintenance genes, particularly *TEL1* and *YKU8* We tested here the extent to which loss of each of the other PAF1C subunits also increases GCR rates, as single mutants and double mutants with *tel1Δ* and *yku80Δ*, and correlated this to defects in other functions of the complex. We show that *PAF1* and *CDC73* play the most important role in genome instability. *PAF1* and *CTR9* play the most important role in transcription elongation, as measured by sensitivity to 6-azauracil, which is consistent with the fact that deletion of these two genes causes the most significant slow growth defect (Mueller and Jaehning, 2002). And, *PAF1* and *RTF1* play the most important role in promoting telomeric silencing. This is consistent with the fact that the telomere position effect is due to histone modification, and Paf1 and Rtf1 specifically play a significant role in the recruitment of histone modifiers (Krogan et al., 2003; Ng et al., 2002). The fact that *paf1Δ* shows the greatest defect in all three aspects suggests it is indeed central to the complex's function. However, there is not a perfect correlation between loss of Paf1C function and increased genome instability for deletion of the other components, suggesting that within the multiprotein structure of the complex, different domains have varying levels of importance for each of the complex's functions. Similar to how the Arndt laboratory identified a 90 amino acid region that is essential for Rtf1's

role in histone modification (Piro et al., 2012), we sought to determine if a similar region exists in Cdc73, which may provide insights in to the Paf1 complex is assembled and what role Cdc73 plays in contributing to the complex's function.

The precise structure and function of the complex is poorly understood. No enzymatic activity has been detected in any of the subunits, consistent with the hypothesis that the complex functions as a scaffold and aids in the recruitment of other proteins (Jaehning, 2010). The best defined domain of any subunit is the Plus3 domain of Rtf1, a 90 amino acid region that is essential for promoting histone modification (Piro et al., 2012). The only structural data for the complex is of the well-conserved C-terminus of Cdc73 (Amrich et al., 2012; Chen et al., 2012). This region was shown to play a modest role in transcription elongation and histone modification, and was independently shown to be important for the association of Paf1C with chromatin (Qiu et al., 2012). To our surprise, deletion of the C-terminus had no effect on the GCR rate, but instead deletion of the N-terminus resulted in an increase in genome instability to the same extent as a *cdc73* null mutant. We therefore performed a deletion analysis of the N-terminal region based on smaller regions of conservation among fungi, and defined a minimal deletion, *cdc73Δ125-229*, that recapitulated a *cdc73* null mutation for increase in GCR rate, 6-azauracil sensitivity, and loss of telomeric silencing. Furthermore, a minimal construct that only expressed those 104 residues, Cdc73:125-229, functioned as well as a wild type copy in all 3 assays. The data also indicated that defects in suppressing genome instability correlated well to defects in transcription elongation and loss of telomeric

silencing, suggesting that all three defects caused by the *cdc73Δ* mutation may have a common mechanistic basis.

By coimmunoprecipitation, we showed that Cdc73 function correlates with successful binding to the other complex components. Cdc73Δ230-393 and Cdc73:125-229 showed decreased binding to Leo1, suggesting the C-terminus may play a role in stabilizing Leo1 in the Paf1 complex, and may explain the mild defects previously described in a *cdc73-ΔC* mutant (Amrich et al., 2012; Qiu et al., 2012). Nonetheless, all the functional Cdc73 truncations, Cdc73Δ230-393, Cdc73Δ2-124, and Cdc73:125-229, were still able to associate with the each of the subunits of the complex. In contrast, Cdc73Δ125-229 showed significantly decreased binding efficiency to Ctr9 and Leo1, and no association was visible for Paf1 or Rtf1. It has previously been shown that deletion of *CDC73* does not affect the ability of Paf1 to associate with Ctr9, Rtf1, or Leo1 (Nordick et al., 2008), which suggests that Cdc73 directly binds to Paf1 and this is likely mediated by residues 125-229.

Because the Paf1C mediates transcription and histone modification, we sought to determine if nuclear localization can explain the functionality of the various truncation mutations. Using constructs tagged at the C-terminus with Venus, we show that the functional truncations, Cdc73Δ230-393, Cdc73Δ2-124, and Cdc73:125-229, were specifically localized to the nucleus to a similar level as in wild type. The Cdc73Δ125-229 construct had similar total cellular expression, but equal levels of the protein were found in the nucleus and cytoplasm, suggesting the 104 amino acid region either includes a nuclear localization signal or is necessary for association with the subunits of the Paf1

complex that provide nuclear localization to Cdc73. We found that wild type Cdc73 still localized to the nucleus after deletion of each of the other complex components. Analogously, we also found that deletion of *CDC73* did not affect the nuclear localization of the other subunits. This is consistent with the previous finding that deletion of *PAF1* or *RTF1* also did not affect the nuclear localization of the other subunits (Porter et al., 2005), suggesting there are redundancies that allow the Paf1C to maintain nuclear localization. We analyzed the protein sequences of each of the Paf1C subunits to test for the presence of any of the six classes of consensus nuclear localization sequences (NLSs), as defined by Kosugi and colleagues (2009). Ctr9 and Rtf1 contain bipartite NLSs, while Paf1 and Leo1 contain Class 1 NLSs; no consensus NLS was found in Cdc73. It will be interesting to verify which of these NLSs are functional, though the lack of findings for Cdc73 does not preclude it from containing its own non-consensus NLS.

Even though deletion of subunits of the complex can cause significant growth defects, the fact that there are redundancies in maintaining nuclear localization suggests how important the complex is for yeast fitness. Furthermore, the results from the IPs and the confocal microscopy together suggest that Paf1 functions at the core of the complex and Cdc73 contributes to Paf1C function by directly binding to Paf1, which is likely primarily mediated by residues 125-229. The Cdc73Δ125-229 construct showed decreased but detectable binding to Ctr9 and Leo1 and it was able to enter the nucleus, suggesting these proteins may play a role in nuclear import. However, its inability to bind Rtf1 or especially Paf1 may suggest why it cannot be retained in the nucleus. Overall, the results are consistent with studies on the human Paf1 complex, which also used pairwise

IPs to suggest that Paf1 is the central component of the complex (Kim et al., 2010). The identification of the Plus3 Domain of Rtf1 and this 104 amino acid region of Cdc73 lend further credence to the idea that the Paf1 Complex functions as a unique 3-dimensional platform with different regions playing different roles in the recruitment and activation of specific factors, and as long as the most essential domains that mediate those interactions are present, the complex is still able to retain its function. As mentioned above, the fact that Cdc73 mutants that have defects in suppressing genome instability also had defects to the same extent in transcription elongation and loss of telomeric silencing, suggest a common mechanism for the defects. This ties in well to the idea that the Paf1 Complex has a multidomain superstructure, such that Cdc73 is important for coordinating the complex's role in maintaining genome stability, but also contributes to overall stability of the complex. So, deletion of *CDC73* causes a large increase in genome instability, but also shows mild defects in transcription elongation and telomeric silencing due to the perturbation of overall Paf1C stability and function.

Interestingly, the human homolog of Cdc73 (hCdc73) has an identified nuclear localization sequence (NLS) at residues 125-139 (Hahn and Marsh, 2005). Because the N-terminus is poorly conserved, it cannot be determined exactly which residues these would correspond to in the yeast homolog (yCdc73), but does substantiate the prediction that yCdc73 contains its own NLS. The vast majority of mutations in *hCDC73* that have been identified in human cancers are frameshift or nonsense mutations in the N-terminus (Newey et al., 2010), however a couple of novel missense mutations have been identified in this region that specifically affect the nuclear or nucleolar localization of hCdc73

(Masi et al., 2014; Pazienza et al., 2013). We have shown here that loss of nuclear localization of yCdc73 results in defects in Paf1C function and, importantly, results in an increase in genome instability. It will be interesting to determine whether similar mutations that disrupt nuclear localization of hCdc73 in cancer cells also results in increased genome instability, which may provide insights into how hCdc73 functions as a tumor suppressor.

METHODS

Construction and propagation of strains and plasmids. YPD and synthetic drop-out media for propagation of strains have been previously described (Chen and Kolodner, 1999). To test for transcription elongation defects, 6-azauracil (Sigma-Aldrich) was added to synthetic complete medium at a final concentration of 50 µg/ml. GCR assays were performed using derivatives of RDKY7635 (*MATalpha hom3-10 ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 lyp1::TRP1 cyh2-Q38K iYFR016::*P_{MFA1}*-LEU2 can1::*P_{LEU2}*-NAT yel072w::CAN1-URA3),* which contains the dGCR assay. Gene deletion and tagging were performed using standard PCR-based recombination methods followed by confirmation by PCR (Janke et al., 2004). Plasmid pBS7 was used for the Venus tag with the *KanMX* selection marker, pBS34 was used for the mCherry tag, with the *KanMX* selection marker, and pYM19 was used to introduce a C-terminus 9-myc tag with the *HIS3* selection marker. Double tagged strains were generated by mating, sporulating, and selecting spores that carried both tags.

CDC73, including 998 bp upstream and 536 bp downstream, were cloned into a TOPO vector using the primers 5'-CACCGAATTGCAAGCGCTTGCAACTTGTTCT TTCTGTGC -3' and 5'-GAATTGCAAGCGCTCCCATGGAAATGAGAGAAGC-3'. Within both primers is an *Afe*I cut site, indicated by the underline. A hygromycin B resistance marker was amplified from the plasmid pFA6a-hphNT1with the primers 5'-GAATTGCAAAGCTTCGGATCCCCGGGTTAATTAA-3' and 5'-GAATTGCAAAG CTTTAGGGAGACCGGCAGATCCG-3' (*Hind*III cut site underlined) and inserted into a *Hind*III cut site located 693 bp upstream of the *CDC73* start codon, to make plasmid pRDK1706. The various Cdc73 mutant constructs were made using the GeneArt Site-Directed Mutagenesis kit (Life Technologies). The mutant constructs were integrated at the endogenous *CDC73* locus by digesting the plasmid with *Afe*I and using the digest to transform the appropriate strain background. Appropriate transformation was confirmed by PCR and sequencing.

Telomere position effect assay. The telomere position effect assay was constructed in BY4742 (*MATalpha leu2Δ0 his3Δ1 ura3Δ0 met15Δ0*) by inserting the *URA3* gene onto the left arm of Chromosome VII at the site of *ADH4*. The plasmid pADH4UCA (Gottschling et al., 1990), a gift from the Zakian Lab, contains *URA3* adjacent to the centromere proximal half of *ADH4*. The plasmid was digested with *Sal*I and *Eco*RI and transformed into BY4742. Transformants were selected on -URA drop out plates, and successful transformation was confirmed by PCR, generating the strain RDKY823 Mutant derivatives of this strain were constructed using standard PCR-based gene disruption methods as above. To assay these strains, yeast were cultured overnight

in liquid YPD at 30^{o}C. Then, 100 µL of the culture was used to make tenfold serial dilutions, and 1.5 µL were spotted onto complete synthetic medium (CSM), CSM lacking uracil (CSM-URA), and CSM supplemented with 1 mg/L of 5FOA (CSM+5FOA). Plates were incubated at 30^{o}C for 3 days before imaging.

Immunoprecipitation was done using the µMACS anti-c-myc magnetic bead IP kit from Miltenyi Biotec. Yeast strains bearing both Venus and c-myc tags in wild type and mutant backgrounds were grown to mid-log phase in 50 mL liquid YPD, harvested, resuspended in 1 mL of the supplied lysis buffer, and incubated on ice for 30 minutes. Cells were lysed with the addition of 100 µL of glass beads and vortexed four times for 1 minute with cooling. Lysates were clarified at 14,000 rpm for 10 minutes at 4^{o}C. Protein concentrations were determined using the DC Protein Assay (Bio-Rad). For the input analysis, 500 µg of protein was trichloroacetic acid (TCA) precipitated, resuspended in 100 µL of 2x SDS gel loading buffer (100 mM Tris-Cl (pH 6.8), 4% SDS, 20% glycerol, 200 mM DTT, 2% bromophenol blue) and 10 µL was used for Western Blotting. For the immunoprecipitation, 1000 µg of protein was incubated with 50 µL anti-c-myc MicroBeads (Miltenyi Biotec) for 30 minutes on ice, then passed through the µMACS separator column. The column was washed twice with 200 µL of lysis buffer, washed twice with 200 µL of wash buffer 1, then washed once with 100 µL of wash buffer 2. The column was then incubated with 20 µL of heated elution buffer for 5 minutes, before the proteins were eluted with 50 µL of heated elution buffer. Of the eluted volume, 12 µL was used for Western Blotting.

Western Blotting. Proteins were resolved on a 4-15% SDS-PAGE gel (Bio-Rad) and then transferred overnight onto nitrocellulose membrane (Bio-Rad). Venus-tagged proteins were detected with the rabbit monoclonal antibody ab290 (Abcam, 1:2000) and myc-tagged proteins were detected with 71D10 rabbit monoclonal antibody (Cell Signaling, 1:1000). Horseradish peroxidase-conjugated goat anti-rabbit secondary antibody (Jackson Laboratories, 1:5000) was used, followed by chemiluminescence detection with SuperSignal Femto Sensitivity Substrate (Life Technologies) and imaged with a Bio-Rad Imager.

Live-Cell Imaging and Image Analysis. Exponentially growing cultures were washed and resuspended in water before being placed on minimal media agar pads, covered with a coverslip, and sealed with valap (a 1:1:1 mixture of Vaseline, lanolin, and paraffin by weight). Cells were imaged on a Deltavision (Applied Precision) microscope with an Olympus 100X 1.35NA objective. Fourteen 5 µm z sections were acquired and deconvolved with softWoRx software. Experiments involving fluorescence quantification were done as described previously (Joglekar et al., 2006). Further image processing, including intensity measurements were performed using ImageJ.

ACKNOWLEDGEMENTS

This chapter is an adaptation of an original document that is being prepared for publication. As of May 19th 2015, the authorship and tentative title of this manuscript are: Nene RV, Campbell CS, Putnam CD, Desai A, Kolodner RD. Identification of a Domain of Cdc73 That is Necessary and Sufficient for the Suppression of Genome

Instability. The dissertation author was the primary author of this material and contributed to the conception and design of experiments, did the strain construction, GCR rate analysis, and coimmunoprecipitations, data analysis, and manuscript writing. Christopher S. Campbell performed the confocal microsopy and image analysis, while Arshad Desai provided additional support. The dissertation author would like to thank Christopher S. Putnam and Richard D. Kolodner for assisting with the conception and design of the research, data analysis, manuscript writing and mentorship. The authors would also like to thank V. Zakian for materials. Rahul Nene is supported by NIH F30 Grant CA177240-01 and the University of California, San Diego Medical Scientist Training Program T32 GM007198-4 This work was also supported by NIGMS Systems Biology Center Grant GM085764, NIH R01 Grant GM26017 and the Ludwig Institute for Cancer Research to Richard D. Kolodner and Christopher S. Putnam.

Conclusion

One of the main goals of this thesis was to leverage the power of genetics to gain insights in to the mechanisms that suppress genome instability and how this can go awry in the context of human cancer. In Chapter 1, we used a modified synthetic genetic array (SGA) approach to cross each of the three GCR assays against a subset of the yeast deletion collection. This led to the identification of 183 genes that suppress genome instability, 65 of which were not previously known. An additional 43 query mutations were crossed against the deletion library and double mutants were screened for synergistic increases in genome instability. This led to the identification of an additional 438 genes that suppress genome instability in the context of a deletion of another gene. Lastly, analysis of ovarian and colorectal cancers showed that the vast majority of tumors appeared to inactivate one or more genes that are homologs of the *S. cerevisiae* genes identified in this screen, suggesting these genome maintenance pathways are indeed well-conserved. We hope that a thorough analysis of our data will reveal the pathways that interact and crosstalk to suppress genome instability, and can be used to gain insights into how mutations in the human homologs of these genes contribute to tumorigenesis. Moreover, the vast amount of data here reveals avenues that warrant further investigation, including analyzing the 65 novel genes identified in the screen, one of which is *CDC73*. This gene is a member of the 5 subunit Paf1 Complex that associates with RNA polymerase II and plays a role in transcription elongation and histone modification.

Furthermore, the human homolog functions as a tumor suppressor, the mechanism of which is poorly understood. The remainder of this thesis was dedicated to using the genetic tools available in budding yeast to analyze the structure and function by which Cdc73 suppresses GCRs.

In Chapter 2, we used a similar synthetic genetic array methodology and extended the results from Chapter 1 by crossing a *cdc73* query mutation against the deletion collection and screening double mutants for increased genome instability. Loss of *cdc73* showed synergistic interactions with mutations in 27 genes, among which were genes that function in telomere homeostasis. Detailed GCR structure analysis showed that *cdc73* mutants favored homologous recombination over *de novo* telomere addition, which led to the discovery that loss of *CDC73* causes mild telomere defects that strongly synergize with loss of other telomere maintenance genes. Ultimately, we found that the large GCR rates in *cdc73* double mutants were due to both defect in telomerase and the accumulation of recombinogenic RNA:DNA hybrids.

In Chapter 3, we expanded our analysis to the rest of the Paf1 Complex members and found that each component of the complex plays varyingly important roles in the suppression of genome instability, and we correlate this to defects in other cellular functions associated with the complex. We next conducted a deletion analysis of Cdc73 and demonstrated that a ~100 amino acid region is necessary and sufficient for all the functions of Cdc73. We used coimmunoprecipitation and confocal microscopy to demonstrate that this region is necessary for the protein to localize to the nucleus, where it coordinates binding to Paf1, thereby contributing to full complex function.

These mechanistic and structural findings provide insights in to how the human homolog of *CDC73* potentially functions as a tumor suppressor by demonstrating that it functions to suppress genome instability. Furthermore, we hope these overall analyses of the genetic interactions that contribute to synergistic increases in genome instability may suggest novel targets in the development of treatments for cancer. For example, increased GCR rates may represent a sublethal hit on the essential processes of DNA replication and genome maintenance, and thus could conceivably be increased to synthetic lethality by inhibiting cooperative or compensatory pathways.

References

Adelman, K., Wei, W., Ardehali, M.B., Werner, J., Zhu, B., Reinberg, D., and Lis, J.T. (2006). Drosophila Paf1 modulates chromatin structure at actively transcribed genes. Mol. Cell. Biol. *26*, 250–260.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.

Aggarwal, M., and Brosh, R.M. (2012). Functional analyses of human DNA repair proteins important for aging and genomic stability using yeast genetics. DNA Repair (Amst). *11*, 335–348.

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

Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol. Cell *46*, 115–124.

Aguilera, A., and García-Muse, T. (2013). Causes of genome instability. Annu. Rev. Genet. *47*, 1–32.

Albuquerque, C.P., Wang, G., Lee, N.S., Kolodner, R.D., Putnam, C.D., and Zhou, H. (2013). Distinct SUMO ligases cooperate with Esc2 and Slx5 to suppress duplication-mediated genome rearrangements. PLoS Genet *9*, e1003670.

Amrich, C.G., Davis, C.P., Rogal, W.P., Shirra, M.K., Heroux, A., Gardner, R.G., Arndt, K.M., and VanDemark, A.P. (2012). Cdc73 subunit of Paf1 complex contains C-terminal Ras-like domain that promotes association of Paf1 complex with chromatin. J. Biol. Chem. *287*, 10863–10875.

Arora, R., Lee, Y., Wischnewski, H., Brun, C.M., Schwarz, T., and Azzalin, C.M. (2014). RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. Nat. Commun. *5*, 1–11.

Askree, S.H., Yehuda, T., Smolikov, S., Gurevich, R., Hawk, J., Coker, C., Krauskopf, A., Kupiec, M., and McEachern, M.J. (2004). A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. Proc. Natl. Acad. Sci. U. S. A. *101*, 8658–8663.

Azzalin, C.M., and Lingner, J. (2015). Telomere functions grounding on TERRA firma. Trends Cell Biol. *25*, 29–36.

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science *318*, 798–801.

Badin-Larcon, A.C., Boscheron, C., Soleilhac, J.M., Piel, M., Mann, C., Denarier, E., Fourest-Lieuvin, A., Lafanechere, L., Bornens, M., and Job, D. (2004). Suppression of nuclear oscillations in Saccharomyces cerevisiae expressing Glu tubulin. Proc Natl Acad Sci U S A *101*, 5577–5582.

Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., and Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. Nat. Struct. Mol. Biol. *20*, 1199–1205.

Banerjee, S., Smith, S., Oum, J.H., Liaw, H.J., Hwang, J.Y., Sikdar, N., Motegi, A., Lee, S.E., and Myung, K. (2008). Mph1p promotes gross chromosomal rearrangement through partial inhibition of homologous recombination. J Cell Biol *181*, 1083–1093.

Bermejo, R., Lai, M.S., and Foiani, M. (2012). Preventing Replication Stress to Maintain Genome Stability: Resolving Conflicts between Replication and Transcription. Mol. Cell *45*, 710–718.

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J.C., Huang, J.H., Alexander, S. (2007). Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc. Natl. Acad. Sci. U. S. A. *104*, 20007–20012.

Bertuch, A. a., and Lundblad, V. (2004). EXO1 Contributes to Telomere Maintenance in Both Telomerase-Proficient and Telomerase-Deficient Saccharomyces cerevisiae. Genetics *166*, 1651–1659.

Bertuch, A.A., and Lundblad, V. (2003). Which end: Dissecting Ku's function at telomeres and double-strand breaks. Genes Dev. *17*, 2347–2350.

Boeke, J.D., LaCroute, F., and Fink, G.R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol Gen Genet *197*, 345–346.

Bonetti, D., Anbalagan, S., Lucchini, G., Clerici, M., and Longhese, M.P. (2013). Tbf1 and Vid22 promote resection and non-homologous end joining of DNA double-strand break ends. EMBO J *32*, 275–289.

Boulton, S.J., and Jackson, S.P. (1996). Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. Nucleic Acids Res. *24*, 4639–4648.

Boulton, S.J., and Jackson, S.P. (1998). Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. EMBO J. *17*, 1819–1828.

Broach, J.R., Strathern, J.N., and Hicks, J.B. (1979). Transformation in yeast: development of a hybrid cloning vector and isolation of the CAN1 gene. Gene *8*, 121–133.

Bunting, S.F., Callen, E., Wong, N., Chen, H.T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L. (2010). 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell *141*, 243–254.

Canman, C.E., and Lim, D.S. (1998). The role of ATM in DNA damage responses and cancer. Oncogene *17*, 3301–3308.

Chan, J.E., and Kolodner, R.D. (2011). A genetic and structural study of genome rearrangements mediated by high copy repeat Ty1 elements. PLoS Genet. *7*, e1002089.

Chan, J.E., and Kolodner, R.D. (2012). Rapid analysis of Saccharomyces cerevisiae genome rearrangements by multiplex ligation-dependent probe amplification. PLoS Genet. *8*, e1002539.

Chan, Y. a., Hieter, P., and Stirling, P.C. (2014). Mechanisms of genome instability induced by RNA-processing defects. Trends Genet. *30*, 245–253.

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

Chávez, S., Beilharz, T., Rondón, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera, A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. EMBO J. *19*, 5824–5834.

Chen, C., and Kolodner, R.D. (1999). Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants. Nat Genet *23*, 81–85.

Chen, H., Shi, N., Gao, Y., Li, X., Teng, M., and Niu, L. (2012). Crystallographic analysis of the conserved C-terminal domain of transcription factor Cdc73 from Saccharomyces cerevisiae reveals a GTPase-like fold. Acta Crystallogr. D. Biol. Crystallogr. *68*, 953–959.

Chin, L., and Gray, J.W. (2008). Translating insights from the cancer genome into clinical practice. Nature *452*, 553–563.

Chun, S., and Fay, J.C. (2009). Identification of deleterious mutations within three human genomes. Genome Res. *19*, 1553–1561.

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Mol. Cell *40*, 179–204.

Ciriello, G., Miller, M.L., Aksoy, B.A., Senbabaoglu, Y., Schultz, N., and Sander, C. (2013). Emerging landscape of oncogenic signatures across human cancers. Nat Genet *45*, 1127–1133.

Cusanelli, E., Romero, C., and Chartrand, P. (2013). Telomeric Noncoding RNA TERRA Is Induced by Telomere Shortening to Nucleate Telomerase Molecules at Short Telomeres. Mol. Cell *51*, 780–791.

D'Andrea, A.D. (2010). Susceptibility pathways in Fanconi's anemia and breast cancer. N. Engl. J. Med. *362*, 1909–1919.

Debrauwere, H., Loeillet, S., Lin, W., Lopes, J., and Nicolas, A. (2001). Links between replication and recombination in Saccharomyces cerevisiae: a hypersensitive requirement for homologous recombination in the absence of Rad27 activity. Proc Natl Acad Sci U S A *98*, 8263–8269.

Engel, S.R., Dietrich, F.S., Fisk, D.G., Binkley, G., Balakrishnan, R., Costanzo, M.C., Dwight, S.S., Hitz, B.C., Karra, K., Nash, R.S. (2014). The reference genome sequence of Saccharomyces cerevisiae: then and now. G3 (Bethesda). *4*, 389–398.

Enomoto, S., Glowczewski, L., and Berman, J. (2002). MEC3, MEC1, and DDC2 are essential components of a telomere checkpoint pathway required for cell cycle arrest during senescence in Saccharomyces cerevisiae. Mol. Biol. Cell *13*, 2626–2638.

Fan, H.Y., Merker, R.J., and Klein, H.L. (2001). High-copy-number expression of Sub2p, a member of the RNA helicase superfamily, suppresses hpr1-mediated genomic instability. Mol. Cell. Biol. *21*, 5459–5470.

Fiorentini, P., Huang, K.N., Tishkoff, D.X., Kolodner, R.D., and Symington, L.S. (1997). Exonuclease I of Saccharomyces cerevisiae functions in mitotic recombination in vivo and in vitro. Mol. Cell. Biol. *17*, 2764–2773.

Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., and Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell *75*, 1027–1038.

Fourel, G., Revardel, E., Koering, C.E., and Gilson, E. (1999). Cohabitation of insulators and silencing elements in yeast subtelomeric regions. EMBO J *18*, 2522–2537.

Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., and Ellenberger, T. (2006). DNA Repair and Mutagenesis (Washington, D.C.: ASM Press).

Garvik, B., Carson, M., and Hartwell, L. (1995). Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint. Mol. Cell. Biol. *15*, 6128–6138.

Gatbonton, T., Imbesi, M., Nelson, M., Akey, J.M., Ruderfer, D.M., Kruglyak, L., Simon, J. a, and Bedalov, A. (2006). Telomere length as a quantitative trait: genome-wide survey and genetic mapping of telomere length-control genes in yeast. PLoS Genet. *2*, e35.

Gerring, S.L., Connelly, C., and Hieter, P. (1991). Positional mapping of genes by chromosome blotting and chromosome fragmentation. Methods Enzymol. *194*, 57–77.

Gómez-González, B., García-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marín, A., Foiani, M., and Aguilera, A. (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. EMBO J. *30*, 3106–3119.

Gottschling, D.E., Aparicio, O.M., Billington, B.L., and Zakian, V. a (1990). Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription. Cell *63*, 751–762.

Gravel, S., Larrivée, M., Labrecque, P., and Wellinger, R.J. (1998). Yeast Ku as a regulator of chromosomal DNA end structure. Science *280*, 741–744.

Greenwell, P.W., Kronmal, S.L., Porter, S.E., Gassenhuber, J., Obermaier, B., and Petes, T.D. (1995). TEL1, a gene involved in controlling telomere length in S. cerevisiae, is homologous to the human ataxia telangiectasia gene. Cell *82*, 823–829.

Greider, C.W., and Blackburn, E.H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature *337*, 331–337.

Hackett, J.A., Feldser, D.M., and Greider, C.W. (2001). Telomere dysfunction increases mutation rate and genomic instability. Cell *106*, 275–286.

Hahn, M. a, and Marsh, D.J. (2005). Identification of a functional bipartite nuclear localization signal in the tumor suppressor parafibromin. Oncogene *24*, 6241–6248.

Hayflick, L., and Moorhead, P. (1961). The serial cultivation of human diploid cell strains. Exp. Cell Res. *25*, 585–621.

Heron, M. (2013). Deaths: leading causes for 2010. Natl. Vital Stat. Rep. *62*, 1–97.

Horowitz, H., Thorburn, P., and Haber, J.E. (1984). Rearrangements of highly polymorphic regions near telomeres of Saccharomyces cerevisiae. Mol. Cell. Biol. *4*, 2509–2517.

Huang, M.E., and Kolodner, R.D. (2005). A biological network in Saccharomyces cerevisiae prevents the deleterious effects of endogenous oxidative DNA damage. Mol Cell *17*, 709–720.

Huang, M.E., Rio, A.G., Nicolas, A., and Kolodner, R.D. (2003). A genomewide screen in Saccharomyces cerevisiae for genes that suppress the accumulation of mutations. Proc Natl Acad Sci U S A *100*, 11529–11534.

Huh, W.-K., Falvo, J. V, Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Iglesias, N., Redon, S., Pfeiffer, V., Dees, M., Lingner, J., and Luke, B. (2011). Subtelomeric repetitive elements determine TERRA regulation by Rap1/Rif and Rap1/Sir complexes in yeast. EMBO Rep. *12*, 587–593.

Inaki, K., and Liu, E.T. (2012). Structural mutations in cancer: mechanistic and functional insights. Trends Genet *28*, 550–559.

Jaehning, J. a (2010). The Paf1 complex: platform or player in RNA polymerase II transcription? Biochim. Biophys. Acta *1799*, 379–388.

Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast *21*, 947–962.

Jimeno, S., Rondón, A.G., Luna, R., and Aguilera, A. (2002). The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. EMBO J. *21*, 3526–3535.

Jinks-Robertson, S., and Petes, T.D. (1986). Chromosomal translocations generated by high-frequency meiotic recombination between repeated yeast genes. Genetics *114*, 731–752.

Joglekar, A.P., Bouck, D.C., Molk, J.N., Bloom, K.S., and Salmon, E.D. (2006). Molecular architecture of a kinetochore-microtubule attachment site. Nat. Cell Biol. *8*, 581–585.

Kanellis, P., Gagliardi, M., Banath, J.P., Szilard, R.K., Nakada, S., Galicia, S., Sweeney, F.D., Cabelof, D.C., Olive, P.L., and Durocher, D. (2007). A screen for suppressors of gross chromosomal rearrangements identifies a conserved role for PLP in preventing DNA lesions. PLoS Genet *3*, e134.

Kàufer, N.F., Fried, H.M., Schwindinger, W.F., Jasin, M., and Warner, J.R. (1983). Cycloheximide resistance in yeast: The gene and its portein. Nucleic Acids Res. *11*, 3123–3135.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. *4*, 363–371.

Khadaroo, B., Teixeira, M.T., Luciano, P., Eckert-Boulet, N., Germann, S.M., Simon, M.N., Gallina, I., Abdallah, P., Gilson, E., Géli, V. (2009). The DNA damage response at eroded telomeres and tethering to the nuclear pore complex. Nat. Cell Biol. *11*, 980–987.

Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. Nat. Rev. Genet. *13*, 204–214.

Kim, J., Guermah, M., and Roeder, R.G. (2010). The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. Cell *140*, 491–503.

Kobayashi, H., Ohno, S., Sasaki, Y., and Matsuura, M. (2013). Hereditary breast and ovarian cancer susceptibility genes (review). Oncol Rep *30*, 1019–1029.

Koch, C., Wollmann, P., Dahl, M., Lottspeich, F., Universität, G. Der, and München, D.- (1999). A role for Ctr9p and Paf1p in the regulation of G 1 cyclin expression in yeast. *27*, 2126–2134.

Kolodner, R.D., Cleveland, D.W., and Putnam, C.D. (2011). Cancer. Aneuploidy drives a mutator phenotype in cancer. Science *333*, 942–943.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009). Six classes of nuclear localization signals specific to different binding grooves of importinα. J. Biol. Chem. *284*, 478–485.

Krogan, N.J., Kim, M., Ahn, S.H., Zhong, G., Kobor, M.S., Cagney, G., Emili, a., Shilatifard, a., Buratowski, S., and Greenblatt, J.F. (2002). RNA Polymerase II Elongation Factors of Saccharomyces cerevisiae: a Targeted Proteomics Approach. Mol. Cell. Biol. *22*, 6979–6992.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: Linking transcriptional elongation to histone methylation. Mol. Cell *11*, 721–729.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. *4*, 1073–1081.

De la Chapelle, A. (2004). Genetic predisposition to colorectal cancer. Nat Rev Cancer *4*, 769–780.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G. V, Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Lea, D.E., and Coulson, C.A. (1949). The distribution of the numbers of mutants in bacterial populations. J. Genet. *49*, 264–285.

Lemieux, J., Goodwin, P.J., Bordeleau, L.J., Lauzier, S., and Theberge, V. (2011). Quality-of-life measurement in randomized clinical trials in breast cancer: an updated systematic review (2001-2009). J Natl Cancer Inst *103*, 178–231.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature *396*, 643–649.

Loeb, L.A. (2001). A mutator phenotype in cancer. Cancer Res *61*, 3230–3239.

Loeb, L.A., and Harris, C.C. (2008). Advances in chemical carcinogenesis: a historical review and prospective. Cancer Res *68*, 6863–6872.

Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z., and Lingner, J. (2008). The Rat1p 5' to 3' Exonuclease Degrades Telomeric Repeat-Containing RNA and Promotes Telomere Elongation in Saccharomyces cerevisiae. Mol. Cell *32*, 465–477.

Lundblad, V., and Blackburn, E.H. (1993). An alternative pathway for yeast telomere maintenance rescues est1- senescence. Cell *73*, 347–360.

Madhusudan, S., and Middleton, M.R. (2005). The emerging role of DNA repair proteins as predictive, prognostic and therapeutic targets in cancer. Cancer Treat Rev *31*, 603–617.

Maicher, A., Lockhart, A., and Luke, B. (2014). Breaking new ground: Digging into TERRA function. Biochim. Biophys. Acta - Gene Regul. Mech. *1839*, 387–394.

Maringele, L., and Lydall, D. (2002). EXO1-dependent single-stranded DNA at telomeres activates subsets of DNA damage and spindle checkpoint pathways in budding yeast yku70?? mutants. Genes Dev. *16*, 1919–1933.

Masi, G., Iacobone, M., Sinigaglia, A., Mantelli, B., Pennelli, G., Castagliuolo, I., Palù, G., and Barzon, L. (2014). Characterization of a new CDC73 missense mutation that impairs Parafibromin expression and nucleolar localization. PLoS One *9*, e97994.

Mikus, M.D., and Petes, T.D. (1982). Recombination between genes located on nonhomologous chromosomes in Saccharomyces cerevisiae. Genetics *101*, 369–404.

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

Moretti, P., and Shore, D. (2001). Multiple interactions in Sir protein recruitment by Rap1p at silencers and telomeres in yeast. Mol. Cell. Biol. *21*, 8082–8094.

Moretti, P., Freeman, K., Coodly, L., and Shore, D. (1994). Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. Genes Dev. *8*, 2257–2269.

Morrow, D.M., Tagle, D.A., Shiloh, Y., Collins, F.S., and Hieter, P. (1995). TEL1, an S. cerevisiae homolog of the human gene mutated in ataxia telangiectasia, is functionally related to the yeast checkpoint gene MEC1. Cell *82*, 831–840.

Mosimann, C., Hausmann, G., and Basler, K. (2006). Parafibromin/Hyrax Activates Wnt/Wg Target Gene Transcription by Direct Association with ??-catenin/Armadillo. Cell *125*, 327–341.

Motegi, A., and Myung, K. (2007). Measuring the rate of gross chromosomal rearrangements in Saccharomyces cerevisiae: A practical approach to study genomic rearrangements observed in cancer. Methods *41*, 168–176.

Motegi, A., Kuntz, K., Majeed, A., Smith, S., and Myung, K. (2006). Regulation of gross chromosomal rearrangements by ubiquitin and SUMO ligases in Saccharomyces cerevisiae. Mol Cell Biol *26*, 1424–1433.

Mozdy, A.D., Podell, E.R., and Cech, T.R. (2008). Multiple yeast genes, including Paf1 complex genes, affect telomere length via telomerase RNA abundance. Mol. Cell. Biol. *28*, 4152–4161.

Mueller, C.L., and Jaehning, J. a. (2002). Ctr9, Rtf1, and Leo1 Are Components of the Paf1/RNA Polymerase II Complex. Mol. Cell. Biol. *22*, 1971–1980.

Mullen, J.R., Chen, C.F., and Brill, S.J. (2010). Wss1 is a SUMO-dependent isopeptidase that interacts genetically with the Slx5-Slx8 SUMO-targeted ubiquitin ligase. Mol Cell Biol *30*, 3737–3748.

Myung, K., Datta, A., and Kolodner, R.D. (2001a). Suppression of Spontaneous Chromosomal Rearrangements by S Phase Checkpoint Functions in Saccharomyces cerevisiae of Medicine. Cell *104*, 397–408.

Myung, K., Chen, C., and Kolodner, R.D. (2001b). Multiple pathways cooperate in the suppression of genome instability in Saccharomyces cerevisiae. Nature *411*, 1073–1076.

Newey, P.J., Bowl, M.R., and Thakker, R. V (2009). Parafibromin--functional insights. J. Intern. Med. *266*, 84–98.

Newey, P.J., Bowl, M.R., Cranston, T., and Thakker, R. V (2010). Cell division cycle protein 73 homolog (CDC73) mutations in the hyperparathyroidism-jaw tumor syndrome (HPT-JT) and parathyroid tumors. Hum. Mutat. *31*, 295–307.

Ng, H.H., Feng, Q., Wang, H., Erdjument-Bromage, H., Tempst, P., Zhang, Y., and Struhl, K. (2002). Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. Genes Dev. *16*, 1518–1527.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A. (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993.

Nordick, K., Hoffman, M.G., Betz, J.L., and Jaehning, J. a. (2008). Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II. Eukaryot. Cell *7*, 1158–1167.

Palladino, F., Laroche, T., Gilson, E., Axelrod, a., Pillus, L., and Gasser, S.M. (1993). SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. Cell *75*, 543–555.

Palles, C., Cazier, J.B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Salguero, I. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet *45*, 136–144.

Pazienza, V., la Torre, A., Baorda, F., Alfarano, M., Chetta, M., Muscarella, L.A., Battista, C., Copetti, M., Kotzot, D., Kapelari, K. (2013). Identification and functional characterization of three NoLS (nucleolar localisation signals) mutations of the CDC73 gene. PLoS One *8*, e82292.

Pennaneach, V., and Kolodner, R.D. (2009a). Stabilization of dicentric translocations through secondary rearrangements mediated by multiple mechanisms in S. cerevisiae. PLoS One *4*, e6389.

Pennaneach, V., and Kolodner, R.D. (2009b). Stabilization of dicentric translocations through secondary rearrangements mediated by multiple mechanisms in S. cerevisiae. PLoS One *4*, e6389.

Pfeiffer, V., and Lingner, J. (2012). TERRA promotes telomere shortening through exonuclease 1-mediated resection of chromosome ends. PLoS Genet. *8*.

Pfeiffer, V., Crittin, J., Grolimund, L., and Lingner, J. (2013). The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. EMBO J. *32*, 2861–2871.

De Piccoli, G., Cortes-Ledesma, F., Ira, G., Torres-Rosell, J., Uhle, S., Farmer, S., Hwang, J.Y., Machin, F., Ceschia, A., McAleenan, A. (2006). Smc5-Smc6 mediate DNA double-strand-break repair by promoting sister-chromatid recombination. Nat Cell Biol *8*, 1032–1034.

Piro, a. S., Mayekar, M.K., Warner, M.H., Davis, C.P., and Arndt, K.M. (2012). Small region of Rtf1 protein can substitute for complete Paf1 complex in facilitating global histone H2B ubiquitylation in yeast. Proc. Natl. Acad. Sci. *109*, 10837–10842.

Porter, S.E., Greenwell, P.W., Ritchie, K.B., and Petes, T.D. (1996). The DNA-binding protein Hdf1p (a putative Ku homologue) is required for maintaining normal telomere length in Saccharomyces cerevisiae. Nucleic Acids Res. *24*, 582–585.

Porter, S.E., Penheiter, K.L., and Jaehning, J.A. (2005). Separation of the Saccharomyces cerevisiae Paf1 complex from RNA polymerase II results in changes in its subnuclear localization. Eukaryot. Cell *4*, 209–220.

Preti, M., Ribeyre, C., Pascali, C., Bosio, M.C., Cortelazzi, B., Rougemont, J., Guarnera, E., Naef, F., Shore, D., and Dieci, G. (2010). The telomere-binding protein Tbf1 demarcates snoRNA gene promoters in Saccharomyces cerevisiae. Mol Cell *38*, 614–620.

Putnam, C.D., and Kolodner, R.D. (2010). Determination of gross chromosomal rearrangement rates. Cold Spring Harb Protoc *2010*, pdb prot5492.

Putnam, C.D., Pennaneach, V., and Kolodner, R.D. (2004a). Chromosome healing through terminal deletions generated by de novo telomere additions in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. *101*, 13262–13267.

Putnam, C.D., Pennaneach, V., and Kolodner, R.D. (2004b). Chromosome healing through terminal deletions generated by de novo telomere additions in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *101*, 13262–13267.

Putnam, C.D., Pennaneach, V., Kolodner, D., and Kolodner, R.D. (2005). Saccharomyces cerevisiae as a Model System To Define the Chromosomal Instability Phenotype Saccharomyces cerevisiae as a Model System To Define the Chromosomal Instability Phenotype †.

Putnam, C.D., Hayes, T.K., and Kolodner, R.D. (2009). Specific pathways prevent duplication-mediated genome rearrangements. Nature *460*, 984–989.

Putnam, C.D., Hayes, T.K., and Kolodner, R.D. (2010). Post-replication repair suppresses duplication-mediated genome instability. PLoS Genet. *6*, e1000933.

Putnam, C.D., Allen-Soltero, S.R., Martinez, S.L., Chan, J.E., Hayes, T.K., and Kolodner, R.D. (2012). Bioinformatic identification of genes suppressing genome instability. Proc. Natl. Acad. Sci. U. S. A. *109*, E3251–E3259.

Putnam, C.D., Pallis, K., Hayes, T.K., and Kolodner, R.D. (2014). DNA repair pathway selection caused by defects in TEL1, SAE2, and de novo telomere addition generates specific chromosomal rearrangement signatures. PLoS Genet. *10*, e1004277.

Qiu, H., Hu, C., Gaur, N. a, and Hinnebusch, A.G. (2012). Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. EMBO J. *31*, 3494–3505.

Reed, S.I., Ferguson, J., and Jahng, K.-Y. (1988). Isolation and Characterization of Two Genes Encoding Yeast Matittg Pheromone Signaling Elements: CDC72 and CDC73. Cold Spring Harb. Symp Quant Biol *53*, 621–627.

Rine, J., and Herskowitz, I. (1987). Four genes responsible for a position effect on expression from HML and HMR in Saccharomyces cerevisiae. Genetics *116*, 9–22.

Rondón, A.G., Jimeno, S., and Aguilera, A. (2010). The interface between transcription and mRNP export: From THO to THSC/TREX-2. Biochim. Biophys. Acta - Gene Regul. Mech. *1799*, 533–538.

Sahi, C., and Craig, E.A. (2007). Network of general and specialty J protein chaperones of the yeast cytosol. Proc Natl Acad Sci U S A *104*, 7163–7168.

Saini, N., Ramakrishnan, S., Elango, R., Ayyar, S., Zhang, Y., Deem, A., Ira, G., Haber, J.E., Lobachev, K.S., and Malkova, A. (2013). Migrating bubble during break-induced replication drives conservative DNA synthesis. Nature *502*, 389–392.

Sambrook, J., and Russell, D.W. (2006). Southern Blotting: Capillary Transfer of DNA to Membranes. Cold Spring Harb. Protoc. *2006*, pdb.prot4040 – pdb.prot4040.

Sandell, L.L., and Zakian, V. a. (1993). Loss of a yeast telomere: Arrest, recovery, and chromosome loss. Cell *75*, 729–739.

Santos-Pereira, J.M., Herrero, A.B., García-Rubio, M.L., Marín, A., Moreno, S., and Aguilera, A. (2013). The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. Genes Dev. *27*, 2445–2458.

Schmidt, K.H., and Kolodner, R.D. (2006). Suppression of spontaneous genome rearrangements in yeast DNA helicase mutants. Proc Natl Acad Sci U S A *103*, 18196–18201.

Schmidt, K.H., Pennaneach, V., Putnam, C.D., and Kolodner, R.D. (2006). Analysis of gross-chromosomal rearrangements in Saccharomyces cerevisiae. Methods Enzymol. *409*, 462–476.

Schulz, V.P., and Zakian, V.A. (1994). The saccharomyces PIF1 DNA helicase inhibits telomere elongation and de novo telomere formation. Cell *76*, 145–155.

Schwarz, J.M., Rödelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nat. Methods *7*, 575–576.

Sheldon, K.E., Mauger, D.M., and Arndt, K.M. (2005). A requirement for the Saccharomyces cerevisiae Paf1 complex in snoRNA 3' end formation. Mol. Cell *20*, 225–236.

Shi, X., Chang, M., Wolf, A.J., Chang, C., Frazer-abel, A.A., Wade, P.A., Burton, Z.F., and Jaehning, J.A. (1997). Cdc73p and Paf1p Are Found in a Novel RNA Polymerase II-Containing Complex Distinct from the Srbp-Containing Holoenzyme. Mol Cell Biol *17*, 1160–1169.

Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L.A., Edwards, K.J., Day, I.N.M., and Gaunt, T.R. (2013). Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. Hum. Mutat. *34*, 57–65.

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

De Souza, J.E., Fonseca, A.F., Valieris, R., Carraro, D.M., Wang, J.Y., Kolodner, R.D., and de Souza, S.J. (2014). S-score: a scoring system for the identification and prioritization of predicted cancer genes. PLoS One *9*, e94147.

Squazzo, S.L., Costa, P.J., Lindstrom, D.L., Kumer, K.E., Simic, R., Jennings, J.L., Link, A.J., Arndt, K.M., and Hartzog, G.A. (2002). The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. EMBO J. *21*, 1764–1774.

Stirling, P.C., Bloom, M.S., Solanki-Patil, T., Smith, S., Sipahimalani, P., Li, Z., Kofoed, M., Ben-Aroya, S., Myung, K., and Hieter, P. (2011). The complete spectrum of yeast chromosome instability genes identifies candidate CIN cancer genes and functional roles for ASTRA complex components. PLoS Genet *7*, e1002057.

Stöcklein, W., Piepersberg, W., and Böck, A. (1981). Amino acid replacements in ribosomal protein YL24 of Saccharomyces cerevisiae causing resistance to cycloheximide. FEBS Lett. *136*, 265–268.

Sugawara, N., and Szostak, J.W. (1983). Recombination between sequences in nonhomologous positions. Proc. Natl. Acad. Sci. U. S. A. *80*, 5675–5679.

Sweet, T.J., Boyer, B., Hu, W., Baker, K.E., and Coller, J. (2007). Microtubule disruption stimulates P-body formation. RNA *13*, 493–502.

Tansey, W.P. (2006). 6-Azauracil Sensitivity Assay for Yeast. Cold Spring Harb. Protoc. *2006*, pdb.prot4613 – pdb.prot4613.

Teng, S.C., and Zakian, V. a (1999). Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in Saccharomyces cerevisiae. Mol. Cell. Biol. *19*, 8083–8093.

The Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615.

The Cancer Genome Atlas Research Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337.

The Cancer Genome Atlas Research Network (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. *45*, 1113–1120.

Tkach, J.M., Yimit, A., Lee, A.Y., Riffle, M., Costanzo, M., Jaschob, D., Hendry, J. a, Ou, J., Moffat, J., Boone, C. (2012). Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. Nat. Cell Biol. *14*, 966–976.

Tomson, B.N., and Arndt, K.M. (2013). The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. Biochim. Biophys. Acta - Gene Regul. Mech. *1829*, 166–126.

Tong, A.H., and Boone, C. (2006). Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods Mol Biol *313*, 171–192.

Tran, P.T., Erdeniz, N., Symington, L.S., and Liskay, R.M. (2004). EXO1-A multi-tasking eukaryotic nuclease. DNA Repair *3*, 1549–1559.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L. a., and Kinzler, K.W. (2013). Cancer Genome Landscapes. Science (80-.). *339*, 1546–1558.

Wade, P. a, Werel, W., Fentzke, R.C., Thompson, N.E., Leykam, J.F., Burgess, R.R., Jaehning, J. a, and Burton, Z.F. (1996). A novel collection of accessory factors associated with yeast RNA polymerase II. Protein Expr. Purif. *8*, 85–90.

Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Mol. Cell *44*, 978–988.

Watanabe, M., Watanabe, D., Nogami, S., Morishita, S., and Ohya, Y. (2009). Comprehensive and quantitative analysis of yeast deletion mutants defective in apical and isotropic bud growth. Curr. Genet. *55*, 365–380.

Watson, J.D. (1972). Origin of concatemeric T7 DNA. Nat. New Biol. *239*, 197–201.

Wellinger, R.J., and Zakian, V.A. (2012). Everything you ever wanted to know about Saccharomyces cerevisiae telomeres: Beginning to end. Genetics *191*, 1073–1105.

Wotton, D., and Shore, D. (1997). A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in Saccharomyces cerevisiae. Genes Dev. *11*, 748–760.

Yoshida, K., and Miki, Y. (2004). Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer Sci *95*, 866–871.

Yuen, K.W.Y., Warren, C.D., Chen, O., Kwok, T., Hieter, P., and Spencer, F. a (2007). Systematic genome instability screens in yeast and their potential relevance to cancer. Proc. Natl. Acad. Sci. U. S. A. *104*, 3925–3930.

Zakian, V.A., Blanton, H.M., and Wetzel, L. (1986). Distribution of telomere-associated sequences in yeast. Basic Life Sci. *40*, 493–498.

Zhang, C., Kong, D., Tan, M.-H., Pappas, D.L., Wang, P.-F., Chen, J., Farber, L., Zhang, N., Koo, H.-M., Weinreich, M. (2006). Parafibromin inhibits cancer cell growth and causes G1 phase arrest. Biochem. Biophys. Res. Commun. *350*, 17–24.

Zhu, J., Pavelka, N., Bradford, W.D., Rancati, G., and Li, R. (2012). Karyotypic determinants of chromosome instability in Aneuploid budding yeast. PLoS Genet. *8*.

Zinovyev, A., Kuperstein, I., Barillot, E., and Heyer, W.-D. (2013). Synthetic Lethality between Gene Defects Affecting a Single Non-essential Molecular Pathway with Reversible Steps. PLoS Comput. Biol. *9*, e1003016.