Role of Transcription in Mammalian Copy Number Variant Formation

by

So Hae Park

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Human Genetics)
in the University of Michigan
2018

Doctoral Committee:

Professor Thomas W. Glover, Co-Chair
Professor Thomas E. Wilson, Co-Chair
Assistant Professor Shigeki Iwase
Professor Mats E. Ljungman
Associate Professor JoAnn M. Sekiguchi

So Hae Park

sohae@umich.edu

ORCID iD: 0000-0002-7076-9152

Dedication

I dedicate this dissertation to my parents, whose selflessness and courage opened numerous doors of opportunities for me. They are the biggest reasons why I was even able to consider graduate school as a possible option after college. They are the living, breathing examples of the American Dream, and I owe them a debt that I don't think I can ever fully pay back.

Acknowledgements

Getting a PhD is a team effort. Without these people I would not have gotten to where I am. First and foremost, I want to express my gratitude for both of my advisors, Tom and Tom, for their mentorship and patience. The last six years have been an unprecedented adventure for all three of us in navigating co-mentorship, attending multiple lab meetings, discussing different strategies, and keeping everyone in the loop. But at the end, the tug of war between different ideas and perspectives raised the quality of the research, and I am forever grateful. I also will always be impressed by their unending optimism, curiosity, and desire to still get their hands dirty (sometimes quite literally!) with research.

I want to also thank the rest of my thesis committee members, Mats, JoAnn, and Shigeki, and the past and the present members of the Wilson and Glover labs for creating a welcoming and supportive environment to discuss science. In particular, Martin has been extremely kind and patient with my never-ending questions about experiment setup, results, and interpretation and with my antics and complaints. Sountharia has taught me so many lab skills in the first three years of graduate school. During the rare times I was physically at the Wilson lab, Sham was a great person to talk to about everything, ranging from science to religion. And many thanks to my undergraduate mentor, Melanie Filiatrault, who first instilled in me the joy of independently seeking answers via scientific experiments.

For the times I was not motivated in the lab or doubted myself about my capabilities, writing helped a lot. Wudan Yan is singlehandedly responsible for getting me started in science

writing. It is amazing how someone I met during an interview weekend (for a school I ended up not attending) has such a profound influence in my life. I got my first taste in science writing at *The Michigan Daily*, and I thank for Ian, Amabel, and Katie for being such fantastic research beat editors during my three years there as a news reporter. Also, a shout-out to Lydia, Brandon, Brian, and Riyah for working with me during the time I was the summer research/hospital/campus life beat editor.

I of course need to thank everyone I had the pleasure of working with for Michigan Science Writers (MiSciWriters) — Ada, Bryan, Shweta, Alisha, Kevin, Sarah, David, Melissa, Whit, and Scott — for providing me an opportunity to explore my career interests outside academia and for letting my creative juices flow. Scott especially has been a fearless champion in students' well-being and intellectual growth, and he always made time for the students, no matter how small the problems were. I cannot thank him enough for his guidance and leadership for MiSciWriters and PIBS.

I want to thank all other people who helped me with my science communication and outreach efforts, including (but not limited to) Elyse and Katherine for the excellent RELATE workshops and being great collaborators on number of science communication projects, Joe Palca and Maddie Sofia at NPR for their tireless efforts toward making science more accessible for the public and establishing FOJBI, and Steve Benowitz for being a great mentor since the AAAS meeting.

What I did during graduate school (research or otherwise) would not have been possible without the supportive staff members. The Human Genetics Department administrative staff members were such a joy to work with, and I am convinced they are the most patient people for dealing with our problems every day. Likewise, the PIBS staff members were incredibly

supportive and caring. Brooke and I especially had many conversations about potential projects, and she has been supportive of my career-exploring and outreach efforts since day one.

My family members and friends have shown me unconditional love and support, and I thank them for my sanity. My sister Anna was always there when I needed to rant about life and when I wanted to share yet another funny meme I found on the Internet at two o'clock in the morning. The "9A" girls have taught me the value of friendship, especially one that spans across different geographical locations, goals, and values. To the "Le Bear" group — Joe, Linda, Susana, Tom, and Chris — we have made so many precious memories over the years, and I really hope one day we can open a nice restaurant/distillery that only serves the best steak and whiskey. I specifically want to thank Susana for being a reliable, sincere friend, someone who I can count on having my back when I need help or when I need to assuage my guilty pleasure of looking at cute Shiba pictures and videos. And a thousand thank you's for Erika, for the deep conversations during our Sunday brunches and all the times we watched superhero movies and your teaching me how the individual movies fit in the bigger Marvel or DC universe.

Last but not certainly not least, I thank Chris Miles for being *the* best partner and confidante for a graduate student who went through intense quarter-life crises. He and Kuma have been with me through all of my ups-and-downs, and I am truly lucky. With him by my side, I am confident that I can face whatever life throws at me next.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AID	Activation-induced cytidine deaminase
aCGH	Array comparative genomic hybridization
APH	Aphidicolin
BAC	Bacterial artificial chromosome
BIR	Break-induced replication
bp	Base pairs
BrdU	Bromodeoxyuridine
BrU	Bromouridine
Bru-seq	Bromouridine labeling and sequencing
cDNA	Complementary DNA
CFS	Common Fragile Site
CNV	Copy Number Variant (in this dissertation, specifically nonrecurrent unless stated otherwise)
CSR	Class switch recombination
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
ddPCR	Droplet digital PCR
Dox	Doxycycline
DSB	Double stranded break
dsDNA	Double-stranded DNA
FHIT	Fragile Histidine Triad
FISH	Fluorescent in-situ hybridization
FoSTeS	Fork Stalling and Template Switching
GRO-seq	Global Run-On sequencing
HR	Homologous recombination
HU	Hydroxyurea
IF	Immunofluorescence
Ig	Immunoglobulin
IR	Ionizing radiation
kb	Kilobase (1000 bp)
LOH	Loss of heterozygosity
Mb	Megabase (1 million bp)
mES	Mouse embryonic stem
MiDAS	Mitotic DNA synthesis
MMBIR	Microhomology-mediated BIR
MMEJ	Microhomology-mediated end joining
mRNA	Messenger RNA
NAHR	Non-allelic homologous recombination
Neo^R	Neomycin resistant gene

NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NS	Not significant
PCR	Polymerase Chain Reaction
Pre-RC	Pre-replication complex
Puro^{R}	Puromycin resistant gene
RNAP	RNA Polymerase
RNASEH1	Ribonuclease H1
RNH	Ribonuclease H
RT-qPCR	Real-Time quantitative PCR
SNP	Single nucleotide polymorphism
ssDNA	Single-stranded DNA
TC-NER	Transcription-coupled nucleotide excision repair
TERT	Telomerase reverse transcriptase
TLS	Translesion synthesis
TrDoFF	Transcription-Dependent Double-Fork Failure
TU	Transcription unit
WT	Wild-type
XPF	Xeroderma Pigmentosum type F
XPG	Xeroderma Pigmentosum type G

Abstract

Genome instability, defined as an increased tendency of genome alteration, is the cause of many human diseases and conditions. It is a hallmark of human cancer and plays a role in aging and the development and function of the nervous system. Genome instability can manifest in several ways, including gaps and breaks at Common Fragile Sites (CFSs) and Copy Number Variants (CNVs). CFSs are sites on human metaphase chromosomes prone to forming gaps or breaks following replication stress. CNVs are submicroscopic genomic alternations that change the copy number of the affected region, also often following replication stress. The genome regions most prone to replication stress-induced CNVs, called "hotspots," coincide with CFSs.

In spite of their implications for human health, mechanisms leading to instability at CFSs and CNV hotspots are unclear. CFSs/CNV hotspots are AT-rich and late replicating, but those properties are not sufficient for the sites' instability. DNA sequence at CFSs/CNV hotspots is shared among all cells, but instability is cell line-specific. We also found that while about 20% of the genome replicates late, hotspots only comprise 0.4% of the genome. Hence, instability at hotspots is determined by properties that vary between different cell lines and genomic regions.

Transcription is one such property. We found that CFSs/CNV hotspots are enriched in large (>500kb), transcribed genes and that given a cell line's transcription profile we can predict where CFSs/CNV hotspots will be in that cell line. I further show that abrogating expression of a large hotspot gene leads to a reduced number of aphidicolin-induced CNVs. These results established transcription of large genes as a determining factor for instability at hotspots.

We propose that a conflict between transcription of large genes and DNA replication drives hotspot instability. I tested a model in which R-loops (RNA/DNA hybrids) create a physical interference for the replication fork and cause the fork to stall and initiate genomic alteration. R-loop manipulation by altering expression of RNase H1 had no significant effect on the frequency of APH-induced instability at hotspots, implying that R-loops do not play a central role in driving APH-induced CNVs, unlike a prior study showing that R-loop manipulation changes CFS instability. However, R-loop accumulation changes the location of breakpoints of these CNVs and change the frequency of the spontaneous CNVs, suggesting that R-loops may still play a role in both APH-induced and spontaneous CNV formation.

In sum, the studies in this dissertation reveal that transcription of unusually large genes plays a pivotal role in instability at CFSs/CNV hotspots during replication stress, but not via an R-loop-associated mechanism. Nonetheless, R-loops threaten genome instability and affect CNV formation outside of hotspots. Future studies are necessary to explore other transcription-replication conflict models at CFSs/CNV hotspots and further characterize R-loop induced CNVs.

Chapter 1 — Introduction

Overview of Dissertation

Genome instability refers to an increased frequency of mutations in the genome during the cell's life cycle. Unsurprisingly, genome instability is involved in many human diseases and conditions. It is a hallmark of almost all human cancers and is thought to play a role in aging as well as development/function of the nervous system.

There are many manifestations of genome instability, as the mutations leading to genome instability can change the primary DNA sequences, grossly rearrange the chromosome structure, or lead to loss/gain of chromosome(s). This dissertation will focus on two — Common Fragile Sites (CFSs) and Copy Number Variants (CNVs).

There are also numerous known factors that contribute to genome instability, but this dissertation will mainly focus on two factors, replication and transcription, and how they potentially oppose each other to cause instability at certain places in the human genome. In particular, the dissertation will mostly focus on the antagonistic relationship between replication and transcription in the presence of replication stress. Replication stress inhibits DNA replication, slowing down or stalling DNA polymerase.

Chapter 1 will provide an introduction to CFSs and CNVs, including what is currently known in the field, what remains to be investigated, and why the studies in this dissertation are relevant. Chapter 2 will focus on my contribution to our group's recent *Genome Research* publication, in which we have demonstrated that transcription is a determining factor for the

instability at CFSs and CNV hotspots, and that we can predict where CNVs will occur based on the transcription profile. Chapter 3 expands on the findings from Chapter 2, where I show that knocking down transcription at a hotspot gene reduces the CNV frequency at the site compared to the isogenic, parental cell line. Chapter 4 explores a possible transcription-replication conflict mechanism at CFSs and CNV hotspots: Does transcription interfere with DNA replication by via R-loops? In Chapter 5, I summarize and discuss the results from Chapters 2, 3, and 4 and provide suggestions for future experiments.

Common Fragile Sites

CFSs are genomic regions that exhibit instability as gaps and breaks on metaphase chromosomes in presence of mild replication stress (Figure 1.1), which is generally defined as a condition that interferes with DNA replication and causes a slowing or stalling of DNA replication forks (Zeman and Cimprich 2013). CFSs make up a subset of fragile sites, and they are different from their counterpart, rare fragile sites, in frequency and genomic features. First, CFSs are present in all individuals whereas rare fragile sites are only present in less than 5% of the population (Zlotorynski et al. 2003). Second, all rare fragile sites are characterized by a repeat expansion, allowing us to predict which genomic regions are prone to instability (Sutherland 2003; Zlotorynski et al. 2003; Debacker and Kooy 2007). However, structural features of CFSs are not as clear, preventing researchers from predicting where CFSs will occur.

Replication stress-inducing chemical agents, such as aphidicolin (APH), a DNA polymerase alpha inhibitor, and hydroxyurea (HU), a ribonucleotide reductase inhibitor, induce CFS instability (Glover et al. 1984; Arlt, Ozdemir, Birkeland, Wilson, et al. 2011; Glover et al. 2005). CFSs already replicate late in the cell cycle, and this timing is pushed even

later, under replication stress (M. Le Beau 1998; Hellman et al. 2000; Palumbo et al. 2010). Furthermore, CFSs possess low-complexity A/T-rich sequences prone to forming non-B DNA structures that interfere with replication (H. Zhang and Freudenreich 2007; Hewett et al. 1998; Mishmar et al. 1998; Fungtammasan et al. 2012; Shah et al. 2010). Based on these observations, it is hypothesized that CFS instability depends on DNA replication error.

Since their discovery in 1984, CFS gaps and breaks were defined and mapped at low resolution on metaphase chromosomes using the accepted cytogenetic definitions of chromosome gaps and breaks, and they were mapped without the knowledge of the specific underlying DNA integrity or chromatin structure. The low-resolution nature of CFSs raises many questions, such as what do the gaps and breaks represent at the DNA sequence level?

Copy Number Variants

Copy number variants (CNVs) comprise another manifestation of genome instability. CNVs, as the name suggests, are submicroscopic alterations of the genome that change the copy number of the region. Unlike CFSs, which were mapped using low-resolution techniques, CNVs were defined and detected with new genomic technologies that enabled higher-resolution analyses, such as oligonucleotide microarrays and next generation sequencing. To date, researchers have mapped over 25,000 polymorphic CNVs, with a thousand CNVs larger than 50kb long (Mills et al. 2011). CNVs can range from 50bp to over 1Mb in size (Mills et al. 2011; Girirajan, Campbell, and Eichler 2011), suggesting that CNVs are ubiquitous and play an important role in shaping the genomic landscape.

There are two main classes of CNVs, recurrent and nonrecurrent. Both classes of CNVs arise from an improper replication mechanism. Recurrent CNVs arise from meiotic non-

homologous allelic recombination (NAHR) that uses highly repetitive non-allelic sequences in the genome as substrates for homologous recombination (Hastings et al. 2009). This results in an incorrect exchange of genetic material, leading to reciprocal gains and losses on different alleles. The breakpoint junctions, which presumably mark where the initial DNA double-stranded break has occurred are "recurrent." For recurrent CNVs, the breakpoints occur at the same genomic region for these CNVs, such as low-copy repeats or segmental duplications (Conrad, Bird, et al. 2010; Conrad, Pinto, et al. 2010; Mills et al. 2011; Sharp et al. 2005; Dittwald et al. 2013), so their locations can be predicted based on DNA sequence.

The breakpoints for nonrecurrent CNVs (often characterized with microhomologies, which are short homologous sequences) are unique for each CNV and can be highly complex (Arlt, Wilson, and Glover 2012; W. Gu, Zhang, and Lupski 2008; Hastings et al. 2009; Carvalho and Lupski 2016). Although these CNVs do tend to cluster at certain genomic regions (Arlt et al. 2012; Wilson et al. 2015), it is more difficult to predict where these CNVs will occur unlike their recurrent CNVs. In addition, the mechanisms that form nonrecurrent CNVs are not as well described as for recurrent CNVs. Like CFSs, nonrecurrent CNVs are thought to form from different DNA replication errors since they are induced with low concentrations of replication stress-inducing drugs (Arlt et al. 2009; Arlt, Ozdemir, Birkeland, Lyons, et al. 2011; Wilson et al. 2015; Durkin et al. 2008). Our observations support the idea that CNVs have a replication-error origin, as they are induced in normal human cells under mild replication stress caused by APH, HU, or even ionizing radiation (IR) (Arlt et al. 2009; Arlt, Ozdemir, Birkeland, Wilson, et al. 2011; Arlt et al. 2014).

Once DNA replication is perturbed from replication stress, the replication fork is stalled and even collapsed, where a collapsed fork is defined as a fork that is unable to synthesize DNA

(Cortez 2015). The stalled/collapsed fork then can undergo one of a few different possible processes. First, the replication fork can restart correctly. Second, the fork can restart incorrectly via mechanisms like Microhomology-Mediated Break-Induced Replication (MMBIR) or fork stalling and template switching (FoSTeS) (Hastings et al. 2009). Third, the fork collapse could lead to DSBs, and the breaks can get repaired incorrectly via mechanisms like Microhomology-Mediated End Joining (MMEJ) (Hastings et al. 2009; Arlt et al. 2012)(Figure 1.2). Of the three aforementioned scenarios, the latter two may lead to the nonrecurrent CNVs. Mechanisms leading to CNVs will be discussed in a later section.

A study from our group has reported that NHEJ is not a key pathway for CNV formation, as CNVs were also induced in mouse embryonic stem cells in the presence or absence of XRCC4 (Arlt et al. 2012), a key non-homologous end joining (NHEJ) protein. These observations support the idea that CNVs are created by DNA repair mechanisms when fork progression is impaired rather than incorrect repair, unless an alternative end-joining pathway such as MMEJ is involved.

Nonrecurrent CNVs make up 40 to 70% of normal human CNVs and 79% of pathogenic CNVs (Conrad, Bird, et al. 2010; P. M. Kim et al. 2008; Vissers et al. 2009), and some of such pathogenic CNVs are associated cancer and neurodevelopmental disorders (Glover, Wilson, and Arlt 2017; D. I. Smith et al. 2006; C. L. Smith, Bolton, and Nguyen 2010). Hence, it is imperative to study what gives rise to the nonrecurrent CNVs to learn the preventive measures and progression of those diseases and disorders.

For brevity, I will refer to the nonrecurrent CNVs simply as CNVs from this point on in this dissertation unless stated otherwise.

Sources of Replication Stress

A number of different conditions can lead to replication stress. First, there are exogenous factors. APH, the drug that is commonly used in CFS and CNV studies to induce instability, is a reversible DNA polymerase α and δ inhibitor (IKEGAMI et al. 1978; Byrnes et al. 1976). DNA polymerase α is a DNA polymerase subunit that is involved in the DNA replication initiation for leading and lagging strands. Hydroxyurea induces replication stress by inhibiting ribonucleotide reductase, an enzyme that catalyzes the formation of deoxyribonucleotides from ribonucleotides (Adams, Berryman, and Thomson 1971; Skoog and NordenskjOLD 1971). X-ray and UV radiation create DNA lesions, which can act as a physical barrier to DNA polymerase if left unrepaired (Tubbs and Nussenzweig 2017).

There are also endogenous factors leading to replication stress. Reactive oxygen species (ROS), natural byproducts in oxygen metabolism, are known to cause oxidative DNA damage, leading to DNA lesion and genome instability, as well as enhancing growth signaling, leading to replication stress (Cadet and Wagner 2013; Burhans and Weinberger 2007; Gaillard, García-Muse, and Aguilera 2015). Secondary DNA structures formed at GC- or AT-rich regions and trinucleotide repeat expansions can also physically impede DNA replication fork progression (Mirkin and Mirkin 2007; BedingerS, Munng, and Alberts 1989; Burrow et al. 2010; Dillon et al. 2013; Valton and Prioleau 2016; Thys et al. 2015). Finally, unrelieved supercoiling in the DNA can also induce replication stress by increasing the torsional stress on the DNA, leading to collapsed forks (Lisa Postow et al. 2001; L. Postow et al. 2001; Rodrigo Bermejo et al. 2009; Ray Chaudhuri et al. 2012; Jossen and Bermejo 2013; J. C. Wang 2002).

In cancer, oncogene activation can also create replication stress by causing dysregulated replication and even re-replication of the DNA (Hills and Diffley 2014; Sarni and Kerem 2017).

DNA replication is usually tightly regulated so the cell replicates its DNA at the right time and only replicates it once during S phase (Li 1995; Machida, Hamlin, and Dutta 2005; Boye, Løbner-Olesen, and Skarstad 2000; Sclafani and Holzen 2007), as deregulated replication and re-replication can lead to genome instability (Donley and Thayer 2013; Abbas, Keaton, and Dutta 2013; Truong and Wu 2011; Blumenfeld, Ben-Zimra, and Simon 2017; Di Micco et al. 2006). Oncogene activation in cancer cells, however, can induce the bypassing those regulatory mechanisms. The activation of oncogenes $HRAS^{V12}$, MYC, and cyclin E increases origin firing (Kotsantis et al. 2016; Miron et al. 2015; Sankar, Kadeppagari, and Thimmapaya 2009), which can deplete nucleotide pools or increase topological stress on the DNA and induce replication stress (Beck et al. 2012; Poli et al. 2012; Bester et al. 2011; Pfister et al. 2015; Yang et al. 2014; R. Bermejo et al. 2007).

Studying CNVs Using a Cell Culture System

Our lab has developed a cell culture model system for monitoring the formation of CNVs that resemble the nonrecurrent CNVs seen in people (Arlt et al. 2009; Durkin et al. 2008; Arlt, Ozdemir, Birkeland, Lyons, et al. 2011). In our cell culture system (Figure 1.3), we grow cells in the presence of mild replication stress for a certain time period. The low drug concentration slows down DNA synthesis but does not completely arrest it, allowing cells to still divide during the APH treatment. It is important to ensure that the cells divide at least once during the treatment since CNVs may be DNA lesions that are incorrectly repaired and manifest themselves in the subsequent cell cycle (Hastings et al. 2009).

The cells are then plated at low density so individual clonal populations (originated from single cells) are separated from one another. We pick the individual clonal populations and grow

the clones until harvesting the cells for their genomic DNA. The DNA is used for downstream analyses to detect CNVs. This cell culture experimental system allows us to more easily study the mechanism and the factors that could possibly lead to the CNVs in human population.

When replication stress-inducing agents are added to the cells, CNV frequency is elevated throughout the human genome (Figure 1.4). Frequency is increased regardless of which type of replication stress was present; for example, both APH and HU induced CNVs in the cells even though the two chemicals stress the cells in a different manner (Arlt, Ozdemir, Birkeland, Wilson, et al. 2011).

These experimentally induced CNVs share numerous defining characteristics with normal and pathogenic human CNVs in sizes and structural features, including the microhomologies at the breakpoint junctions, which are typical of almost all CNVs that arise in cancers and the majority of normal CNV variants and *de novo* CNV mutations (Korbel et al. 2007; Mills et al. 2011; Lee, Carvalho, and Lupski 2007; Campbell et al. 2008; Conrad, Bird, et al. 2010; Vissers et al. 2009; Inoue et al. 2002; White et al. 2006).

Paucity of Replication Origins at CFSs and CNV Hotspots

Even though we observed CNVs across the entire genome with replication stress, some genomic regions were clearly more susceptible to the CNVs than others (Figure 1.4), such as *LSAMP* (3q13.31), *AUTS2* (7q11.2), and *WWOX* (16q23.3). Such regions that were more susceptible to CNVs (specifically with more than five CNVs based on our dataset), were called "hotspots." Two recent studies from our group have highlighted how CNV hotspots are located in the same genomic regions as CFSs (Wilson et al. 2015; Arlt, Ozdemir, Birkeland, Wilson, et al. 2011). Both studies showed that almost all CNV hotspots coincide with previously identified

CFSs (Figure 1.5), suggesting that there is a link between the mechanisms that contribute to the two forms of instabilities. In fact, CFS gaps/breaks and CNVs are thought to be two different manifestations from one shared property, which is late replication.

Both CFSs and CNV hotspots replicate late in the S phase, to a point that the part of the site remains un-replicated up to late G2 phase (Wilson et al. 2015; Minocherhomji et al. 2015; L Wang et al. 1999; M. Le Beau 1998; Palakodeti et al. 2004; Pelliccia et al. 2008). Because of this property, these sites are vulnerable to an incomplete replication to begin with, and replication stress makes CFSs and CNV hotspots even more vulnerable to incomplete replication and instability.

What makes CFSs and CNV hotspots replicate so late? These regions are known to exhibit replication origin scarcity, especially in the middle of the site. Using molecular combing approaches, Letessier et al showed that the CFSs (FRA3B and 3q13) have a paucity of active replication origins in the middle of the sites, rendering the center of the region more susceptible to instability. The authors also observed that CFS instability is cell-type specific. The study reported the only the cells that exhibited instability at certain CFSs showed paucity of active origins at the center of the sites, whereas cells that did not exhibit instability at the same CFSs did not show the same paucity of active origins.

Dormant Origins to the Rescue

The flanking portions of CFSs and CNV hotspots tend to replicate earlier than the middle of the sites, meaning that the replication forks have to travel from edges toward the middle (Wilson et al. 2015; Letessier et al. 2011; Mi. Debatisse, Achkar, and Dutrillaux 2006). With replication stress, the two forks traveling toward the middle of the sites may stall and collapse,

rendering the middle sensitive to incomplete replication (Wilson et al. 2015). Dormant origins may provide the cell's line of defense for these situations. These origins are unused under normal conditions, such as the origins at the center of CFSs and CNV hotspots (Ge, Jackson, and Blow 2007; Blow, Ge, and Jackson 2011; Alver, Chadha, and Blow 2014; D. McIntosh and Blow 2012; Blow and Ge 2009). They are hypothesized to only fire to complete replication upon replication stress, thereby alleviating the effects of stalled or collapsed forks, especially when two neighboring forks stall (Figure 1.6).

It is unclear whether the dormant origins are structurally different from the earlier-firing origins. Dormant origins are thought to be extra replication origin proteins that were loaded onto the DNA during G1 (Ge, Jackson, and Blow 2007; Blow and Ge 2009; Woodward et al. 2006; Debbie McIntosh and Blow 2012; Blow, Ge, and Jackson 2011). In G1 phase of the eukaryotic cell cycle, MCM2-7 helicase hexamer complex is loaded onto a protein complex comprised of ORC1-6, CDC6, and CDT1 on the DNA to form pre-replication complex (pre-RC) to mark the replication origins in preparation for S phase during a process called licensing (Lei and Tye 2001; Nishitani and Lygerou 2018; Tsakraklides and Bell 2010). Not all the licensed origins are activated under the normal condition, as there is up to a 20-fold excess of MCM2-7 helicase loaded onto the DNA compared to the number of origins that actually fire during S phase (Ibarra, Schwob, and Méndez 2008; Ge, Jackson, and Blow 2007; Hyrien 2016; Woodward et al. 2006; Donovan et al. 1997). The limiting factor for origin activation is hypothesized to be CDC45, a protein that binds to pre-RC after the licensing in G1 to activate the replication origin during the S phase (Wong et al. 2011; Köhler et al. 2016; Pollok et al. 2007; Mantiero et al. 2011; Tanaka et al. 2011; Edwards et al. 2002), not the number of pre-RCs bound to the DNA.

Unsurprisingly, mutations in the pre-RC proteins leads to increased genome instability, diseases, and sometimes even death of the organism. Shima et al. have demonstrated that *Mcm4* null mice are embryonically lethal, but the Chao3 mutation in mouse *Mcm4*, which changes the amino acid phenylalanine to isoleucine at residue 345, results in a hypomorphic mutation. While *Mcm4* null mice died before birth, more than 80% of $Mcm4^{Chaos3/Chaos3}$ females died from mammary adenocarcinomas within 12 weeks of age. Mouse $Mcm4^{Chaos3/Chaos3}$ embryonic fibroblasts were more prone to chromosome breaks after aphidicolin treatment. Similarly, missense mutations in the ORC1, ORC4, ORC6, CDT1, and CDC6 protein cause Meier-Gorlin syndrome, which is a form of primordial dwarfism (Bicknell et al. 2011; de Munnik et al. 2012).

Different Manifestations of Stalled and Collapsed Forks

Studies in human cells have shown that replication stress induces the activation of dormant origins to compensate for the slow or stalled fork progression, and the lack of the dormant origin activation can result in higher genomic instability, such as the case of the origin paucity at CFSs. The late replication and origin paucity at CFSs and CNV hotspots can slow down and stall replication forks following replication stress, and the sites can then manifest into different forms of genome instability. The different consequences are detailed in the following sections (Figure 1.7).

CFS Gaps and Breaks

The defining consequences of incomplete replication at CFSs are the site-specific gaps and breaks on metaphase chromosome, which were previously thought to represent unreplicated

DNA. However, Minocherhomji et al. recently detailed the events that occur at CFSs during the G2 and M phases of the cell cycle and uncovered a novel mechanism of DNA replication error resolution. The study reported that after APH treatment, the passage of incompletely replicated DNA at CFSs into mitotic chromosomes serves as the trigger to activate a novel and distinct M phase DNA replication pathway called MiDAS (Mitotic DNA Synthesis).

The authors identified a number of components of the MiDAS pathway, including the non-catalytic subunit of DNA Polymerase δ, POLD3. POLD3 is involved in break-induced replication (BIR) (Tumini et al. 2016), which is a DNA repair pathway similar to homologous recombination (HR) that involves strand invasion but only involves a break at one end instead of two (Malkova and Ira 2013; Anand, Lovett, and Haber 2013). Similarly, another study from the same group found that Rad52, another factor involved in BIR (Sotiriou et al. 2016), is important to MiDAS (Bhowmick, Minocherhomji, and Hickson 2016). These studies imply that BIR is involved in MiDAS.

Ultrafine Anaphase Bridges

If DNA replication is still not completed in early mitosis, the persistent unreplicated DNA at CFSs can lead to the formation of ultrafine anaphase bridges (UFBs), which look like threads of DNA that links CFS loci on the separating sister chromatids. UFBs are different from other anaphase bridges because they lack histones and cannot be visualized with conventional DNA dyes, such as DAPI (K. Chan, North, and Hickson 2007). Instead, UFBs are visualized by probing for proteins such as PLK1-interacting checkpoint helicase (PICH or ERCC6L), FANCD2, FANCI, and BLM69–71, which are important for the resolution of UFBs (Baumann et al. 2007; Biebricher et al. 2013; K. L. Chan and Hickson 2011).

Although UFBs can be observed with low-dose APH, their frequency is greatly increased after the depletion of MUS81, SMC2, or other proteins that are involved in MiDAS (Minocherhomji et al. 2015). Also, large p53-binding protein 1 (53BP1 or TP53BP1) foci co-localize with CFSs in the subsequent G1 phase, demonstrating that UFBs at CFSs represent DNA that fails to complete replication in M phase (Minocherhomji et al. 2015). The fate and biological importance of UFBs are not completely understood, but they are hypothesized to have a role in genome instability, including generating DNA breaks, genome rearrangements, and chromosomal non-disjunction (Minocherhomji et al. 2015; Baumann et al. 2007).

Incorrect Fork Start and DNA Repair Resulting in CNVs

CNVs are thought to manifest when the stalled replication fork either starts incorrectly or repairs the resultant DSB incorrectly. One prevailing model, FoSTeS, proposes that the 3' end of the stalled replication forks switch DNA templates in other replication forks using sequence microhomology in the templates (Lee, Carvalho, and Lupski 2007). The idea behind this model was initially observed in *E. coli lac* genes, where the data suggested that template switching, which initially was thought to only occur in a single replication fork, could occur between different, nearby forks (Slack et al. 2006). Like the human CNVs, the *E. coli* genome rearrangements were marked by microhomologies at the breakpoints.

Another prevailing model is MMBIR, which differs from the BIR discussed above in that the role of Rad51 is not as important (Hastings et al. 2009). In Rad51-dependent BIR, Rad51 coats the resected ssDNA and aids the invasion of the 3' end of the collapsed fork into a template to form a D-loop, where DNA replication eventually gets re-initiated (Pâques and Haber 1999; Bianco, Tracy, and Kowalczykowski 1998). The replication stress that induces CNV formation

may downregulate Rad51, but there have been studies reporting that BIR could occur with low Rad51 levels (Davis and Symington 2004; VanHulle et al. 2007), suggesting that BIR is still a possible mechanism for CNV formation. Also, MMBIR, as the name suggests, results in regions of microhomology at the breakpoint junctions. This also sets MMBIR different from the traditional BIR, as Rad51-mediated BIR requires longer regions of homology than 2-15bp, possibly around 100bp (Ira and Haber 2002; Mehta, Beach, and Haber 2017).

Finally, the MMEJ model postulates that stalled replication forks collapse and create DSBs, which are repaired via MMEJ to create CNVs with microhomologies at the breakpoint junctions. MMEJ, sometimes called alternative NHEJ or back-up NHEJ, is unlike other end joining mechanisms like the classical NHEJ in that the process aligns the broken ends using microhomologies before joining the two ends (McVey and Lee 2008; H. Wang and Xu 2017). MMEJ was first observed in yeast, in which classical NHEJ-defective mutants still showed a low level of end joining marked by 8-10bp of microhomology (Ma et al. 2003). Like FoSTeS and MMBIR, MMEJ is error prone; the ends produced by DSBs are processed and resected before short stretches of microhomology are recognized for end joining (Sfeir and Symington 2015).

Importance of Transcription in CFSs and CNV Hotspots

Although late replication and origin scarcity are defining features for CFSs and CNV hotspots, it is not sufficient to cause instability at the sites since approximately 20% of the entire genome replicates late, but CNV hotspots only comprise 0.4% of the genome (Wilson et al. 2015). Likewise, the AT-richness at CFSs is not sufficient for instability either since CFS instability is cell-type specific (Letessier et al. 2011; Le Tallec et al. 2013; Murano, Kuwano, and

Kajii 1989), meaning that shared characteristics like DNA sequence cannot be the sole determining factor.

CFS and CNV hotspots are also preferentially located in coding regions of very large genes (D. I. Smith et al. 2006; Mai et al. 2007; M. Debatisse et al. 2012; Wilson et al. 2015), thus a possible determining factor is transcription. Several recent studies have reported that it is the transcription of these large genes that is crucial for CFS instability. Helmrich et al. focused on five large genes underlying CFSs (*FHIT*, *IMMP2L*, *WWOX, CNTNAP2,* and *DMD*) and demonstrated that the CFSs at the three transcribed genes (*FHIT*, *IMMP2L*, and *WWOX*) exhibited gaps and breaks after the APH treatment, but the other two un-transcribed sites did not (Helmrich, Ballarino, and Tora 2011). Wei et al. mapped 27 recurrent DSB clusters (RDCs), which were more prone to translocations after replication stress. All but one RDCs mapped to large genes (>400kb), and all but one RDCs were transcribed in the used cell line (P.-C. Wei et al. 2016).

Transcription-Replication Collisions

Currently, it is unclear how transcription is contributing to the instability at CFSs and CNV hotspots. We draw on what we know about these sites and propose that the instability at CFSs is due to a conflict between transcription and replication. Since transcription and replication occur on the same DNA template, if the two processes occur at the same time at the same region, conflicts are inevitable. This conflict between replication and transcription has been extensively studied in many different organisms — ranging from bacteria to humans — as the two processes are present and necessary in all forms of life.

Bacteria

Many studies have suggested that transcription is a natural impediment to replication as the two machineries can head-butt or the replication machinery (replisome) can collide with the transcription machinery from behind (Figure 1.7). Head-on collision occurs when the two machineries move toward each other, either causing a direct physical contact between the two machineries or halting two machineries due to the overwound DNA strands between them (Mirkin and Mirkin 2005b; Deshpande and Newlon 1996; Prado and Aguilera 2005).

Co-directional collision occurs when the two machineries move in the same direction. Because replisomes move about 12-fold faster than RNA polymerase in bacteria, replisomes can collide with the transcription machinery from behind (Mirkin and Mirkin 2005b; Hwang and Kornberg 1992; Breier, Weier, and Cozzarelli 2005). The collision leads to a stalled replication fork and possibly causes some members of the replisomes to dissociate (Goranov et al. 2009; Su'etsugu and Errington 2011). The failure to resolve these conflicts induced by the collision can lead to genome instability in bacteria (Vilette, Ehrlich, and Michel 1995; Gan et al. 2011).

The bacterial genome has evolved so the majority of the transcribed genes are oriented in the same direction as the replisome movement, minimizing head-on collisions (Brewer 1988). Although co-directional collisions will occur, studies show that co-directional collisions are not as detrimental to the replication fork movement as head-on collisions (Mirkin and Mirkin 2005b; Vilette, Ehrlich, and Michel 1995). Hence, orienting most of the transcribed genes co-directional to the replisome movement would not eliminate all collisions but minimize the more harmful consequences.

Yeast

Unlike bacteria, eukaryotes have limited their DNA replication in the S phase of the cell cycle, which allows the cells to transcribe during the non-S phases to avoid conflicts between replisomes and transcription machinery. Transcription does occur during S phase, but generally the domains undergoing transcription are spatially separated from the domains going through replication (Wansink et al. 1994; X. Wei et al. 1998).

Furthermore, eukaryotic replisomes and transcription machineries travel in similar speeds (Hiratani et al. 2008; Darzacq et al. 2007; Pérez-Ortín, Alepuz, and Moreno 2007; Singh and Padgett 2009; Veloso et al. 2014). These observations from previous studies imply that co-directional collisions between replisomes and transcription machinery are less likely in eukaryotes, although such collisions were observed in budding yeast using *in vivo* recombination constructs (Prado and Aguilera 2005).

Head-on collisions in budding yeast were also shown to be more detrimental than co-directional collisions, consistent with the observations in bacteria (Prado and Aguilera 2005). Unlike the bacterial genome, the co-orientation bias of transcription and replication is not obvious in the yeast genome, but this may be explained by the compact nature of the genome (Hamperl and Cimprich 2016). The compact-ness makes it more likely that the replication and transcription machineries encounter each other in either direction, which does not create an evolutionary need for the genome to evolve to create the co-orientation bias (McGuffee, Smith, and Whitehouse 2013). However, the yeast genome does seem to have mechanisms to minimize head-on collisions in other ways, as the highly transcribed ribosomal DNA transcription unit in yeast contains a replication fork barrier at the 3' end to prevent replication fork progression toward the transcription machinery (Brewer and Fangman 1988).

Humans

Potential conflict between replisome and transcription machinery is not as well characterized in humans as in other organisms, but studies do imply that human cells also experience similar conflicts (Hamperl et al. 2017; Helmrich, Ballarino, and Tora 2011). In human cells treated with replication stress-inducing drugs, replication fork stalling is thought to preferentially occur at transcribed regions, suggesting that replisomes and transcription machinery have conflicts (Tuduri et al. 2009; Helmrich, Ballarino, and Tora 2011).

In dividing cells, the large genes in the human genome are expected to take more than one complete cell cycle to be fully transcribed based on their length, the speed of the transcriptional machinery, and the average time it takes for a cell to complete a cell cycle (Helmrich, Ballarino, and Tora 2011). Consistent with the theme seen in other organisms, the human genome is organized such that replication and transcription tend to be oriented in the same direction (Petryk et al. 2016). The co-directional collisions are not as detrimental to the replication fork movement as head-on collisions in human cells as well (Hamperl et al. 2017). This observation supports the idea that head-on collisions are more detrimental for the genome than the co-directional collisions across species.

Role of R-Loops in Transcription-Replication Conflict

R-loops are RNA/DNA hybrids that rarely form as a result of transcription and are thought to form via "thread-back" model, when the nascent RNA invades the duplex DNA once it leaves RNAP (RNA polymerase). R-loops are natural intermediates during *E. coli* plasmid replication (Itoh and Tomizawa 1980), mitochondrial DNA replication (Xu and Clayton 1996;

Pohjoismäki et al. 2010), and immunoglobulin class switching (Reaban and Griffin 1990). In addition to these *in vivo* observations, R-loops were initially thought to be rare byproducts of transcription, but R-loops are now thought to play an important role in regulating gene expression. R-loops tend to cluster near the GC-rich 5' and 3' ends of transcribed genes and are associated with relevant histone marks to activate transcription or facilitate termination (Ginno et al. 2013).

How R-loops regulate transcription is unknown, but recent studies have proposed that R-loops induce epigenetic changes to regulate gene expression. Skourti-Stathaki et al. (2014) demonstrated that R-loops and H3K9me2 (a repressive mark) co-localize at the termination regions of coding genes, suggesting a potential functional link between the two (Skourti-Stathaki, Kamieniarz-Gdula, and Proudfoot 2014); however, it is not clear whether R-loops recruit the epigenetic signature or the other way around. Another study has proposed that R-loop accumulation at the 5' UTR either recruits H3K4me3 (an active mark) or protects the promoter from DNA methyltransferases to activate transcription, suggesting that R-loops may recruit the histone marks (Ginno et al. 2012).

Although R-loops have regulatory roles, they are also a threat to genome integrity. Groh et al. showed that R-loops form in the repeat expansion of *FXN* (GAA repeats) and *FMR1* (CGG repeats) genes, which are associated with Friedreich ataxia (FRDA) and Fragile X syndrome (Campuzano et al. 1996; Kremer et al. 1991; S. Yu et al. 1991; Verkerk et al. 1991; Oberlé et al. 1991), respectively. The R-loop accumulation led to H3K9me2 upregulation at least in *FXN*, silencing the gene and contributing to the pathology of FRDA (Groh et al. 2014).

In addition, the structure itself can be vulnerable to DNA damage, leading to mutations. The ssDNA in R-loops is vulnerable to various enzymes like AID, an enzyme that initiates the

class switch recombination (Muramatsu et al. 2000; Revy et al. 2000; Petersen-Mahrt, Harris, and Neuberger 2002; Chaudhuri et al. 2003; Dickerson et al. 2003; Pham et al. 2003), and XPF/XPG, two enzymes involved in the nucleotide excision repair pathway (NER) (Zotter et al. 2006; Houtsmuller et al. 1999; Mone et al. 2004). The breaks in the ssDNA may result in DSBs, which could incorrectly resolve into genome rearrangements like CFS gaps/breaks and CNVs.

Additionally, R-loops are hypothesized to stabilize the RNAP during head-on replication-transcription collisions since RNA/DNA hybrids are more stable than DNA/DNA or RNA/RNA duplexes (Roberts and Crothers 1992; Lesnik and Freier 1995; Chien and Davidson 1978). Thus when there is a conflict between the replisome and transcription machinery (especially during a head-on collision), the transcription machinery cannot be ejected from the DNA because of the stable R-loop (Figure 1.8), thereby dissociating the replisome. Helmrich et al (2011) showed a link between CFS instability and R-loops; the authors detected R-loops in the middle of FRA3B/*FHIT*, unlike in the middle of Cyclin B1, a non-CFS/hotspot gene. The R-loop accumulation was exacerbated upon APH addition and even further upon knockdown of *RNASEH1*, an enzyme that resolves R-loops. *RNASEH1* knockdown increased the instability at three CFSs (FRA3B, FRA16D, and FRA7K) and vice versa. These data imply that R-loops form in middle of the CFS genes with replication stress, and R-loop accumulation increases CFS instability.

CFS and CNV Hotspot Genes Are Clinically Relevant

Based on the existing models, we propose that transcription poses a risk for instability in replicating cells, which could lead to various human diseases and disorders. A number of genes underlying CFSs and CNV hotspots in cultured cells correspond to clinically relevant human

genes. Curiously, many of the largest human genes are important for neuronal development and function (Wilson et al. 2015; P.-C. Wei et al. 2016). Wei et al. has shown that many DSB-mediated translocation and other genome rearrangement hotspots are in genes with neuron-specific functions. Consistent with this observation, deletion CNVs within *AUTS2, IMMP2L, NRXN1*, and *CNTNAP2* are linked to autism spectrum disorder, intellectual disability, and other psychiatric disorders (Mefford et al. 2010; Beunders et al. 2013; Nagamani et al. 2013; Maestrini et al. 2010; Gregor et al. 2011; Tucker et al. 2014; Friedman et al. 2008; Elia et al. 2010; Mikhail et al. 2011; Veerappa et al. 2013).

Many previous studies also have reported deletions of CFSs and CNV hotspots, such as FRA3B and FRA16D (underlying genes are *FHIT* and *WWOX*, respectively), in cancer cell lines and primary cancer samples (Saldivar, Shibata, and Huebner 2010; Iliopoulos et al. 2006; M. M. Le Beau et al. 1998; Kuroki et al. 2002; Paige et al. 2001; Siprashvili et al. 1997), leading to thinking that many of the underlying large genes are tumor suppressor genes. The induced deletion CNVs are also found in numerous human cancers including gastrointestinal tract, cervical, lung, and breast, and according to a meta-analysis of The Cancer Genome Atlas database, 22 out of the 70 focal deletion regions are located in the large genes (Glover, Wilson, and Arlt 2017).

In addition, CFSs and CNV hotspots can facilitate cancer development by providing preferential sites for viral DNA integration. FRA3B, one of the most extensively studied CFSs, contains a human papilloma virus (HPV-16) integration site in human cervical carcinoma (Wilke et al. 1996). In addition, Hu et al. has found that HPV integration site junctions were characterized with microhomology, which is also present in CNV breakpoint junctions; the

authors proposed that viral integration uses the same FoSTeS/MMBIR pathways that are thought to drive CNV formation (Hu et al. 2015).

The distribution of the induced CNVs is strikingly similar to that observed in cancers and in some individuals with autism and intellectual disability (Glover, Wilson, and Arlt 2017; Beunders et al. 2013), suggesting some type of replication stress is driving instability. As discussed in an earlier section, in cancer cells, a possible source of *in vivo* replication stress is oncogene activation. Oncogene activation can lead to deregulated DNA replication and accumulation of R-loops and reactive oxygen species (Gaillard, García-Muse, and Aguilera 2015). Previous studies have shown that aberrant activation of cyclin E or $HRAS^{V12}$ increases instability at some known CFSs and CNV hotspots but also at a number of unique sites (Miron et al. 2015) and that cyclin E overexpression can lead to genomic instability (Jones et al. 2013; Ekholm-Reed et al. 2004).

Other Fragile Sites and CNVs

The models and properties presented so far are the most relevant for a subset of the most unstable CFSs and CNV hotspots. CNVs in hotspots comprise a large portion of all detected CNVs (more than 40%), but the majority of CNVs are still in non-hotspot regions (Wilson et al. 2015). It is not clear how these non-hotspot CNVs form, although it is possible that the same mechanism for hotspot can apply to these CNVs since regions with one to five CNVs are still significantly enriched in transcribed large genes (Wilson et al. 2015).

As stated above, there are other types of fragile sites and CNVs that were not as extensively covered in this chapter. Based on what is known so far, the mechanisms for rare fragile sites and recurrent CNVs are different from the mechanisms for CFSs and nonrecurrent

CNVs. Rare fragile sites, which are characterized by repeat expansions, are thought to occur based on the DNA replication perturbations from DNA secondary structures (Zlotorynski et al. 2003; Sutherland 2003; Schwartz, Zlotorynski, and Kerem 2006). Recurrent CNVs, characterized by low-copy repeats, are thought to arise via NAHR (Hastings et al. 2009).

Furthermore, Barlow et al. recently identified a novel class of fragile sites called Early-Replicating Fragile Sites (ERFSs), which replicate early in the cell cycle, are preferentially located at repetitive elements and at CpG dinucleotides (i.e. promoters), and not enriched in large genes unlike CFSs. However, both ERFSs and CFSs are increased in instability upon replication stress, ATR knockdown, and oncogenic stress. In addition, ERFSs and CFSs are both associated with transcription, as more than 86% of ERFS-associated genes are among the highest transcribed genes (Barlow et al. 2013). Although the exact mechanism for ERFS is yet to be elucidated, Barlow and colleagues proposed that the persistent damage at ERFSs could carry into the next G1 stage of the cell cycle, becoming a preferential site for translocation with sites of AID damage.

Rationale for Dissertation

Maintaining genome stability is crucial for an organism's viability. The two forms of genome instability discussed above, CFSs and CNVs, are involved with many human diseases and conditions. Yet, the mechanisms that lead to the instability at CFSs and CNV hotspots are currently unclear. Recent efforts have characterized CFSs and CNV hotspots in a genome-wide fashion in order to elucidate what makes them relatively unstable compared to other parts of the genome. Understanding their unique features will help generate more hypotheses about how CFSs and CNV hotspots become unstable.

Investigating the relationship between replication and transcription at these sites may shed light to the mechanism. Proper and complete replication is crucial for maintaining stability, but the transcription of large genes at these sites can potentially interfere with the replication process. Future experiments should focus on uncovering the relationship between replication and transcription at CFSs and CNV hotspots. Uncovering the relationship will not only grant us an insight into what leads to instability at these sites, but what is required to maintain the genome integrity to prevent human diseases and conditions that are induced by the instability.

Figures

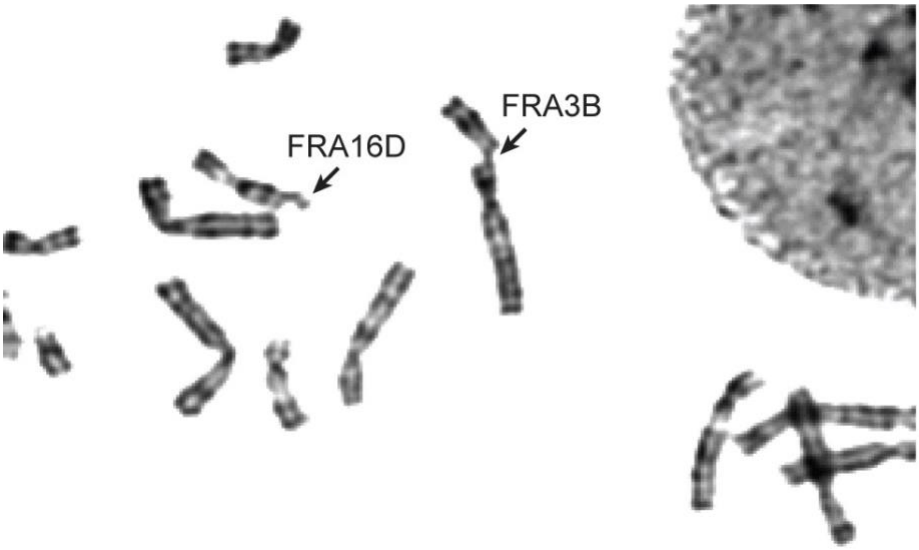


Figure 1.1. CFSs are sites that exhibit instability as gaps and breaks on metaphase chromosomes. FRA16D and FRA3B are two well-characterized CFSs on human chromosomes 16 and 3, respectively. Adapted from Durkin and Glover (2008).

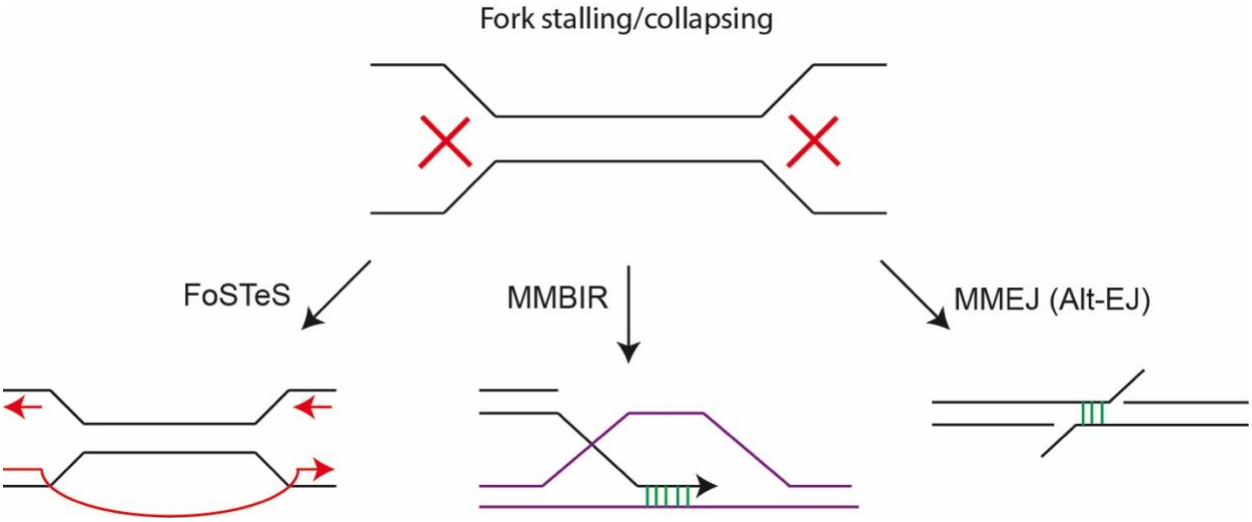


Figure 1.2. Possible mechanisms for nonrecurrent CNVs. Upon replication stress, replication fork may stall and even collapse, creating DSB(s). When the collapsed fork does not restart correctly via FoSTeS or MMBIR, or if the DSB is repaired incorrectly through MMEJ, nonrecurrent CNVs may result.

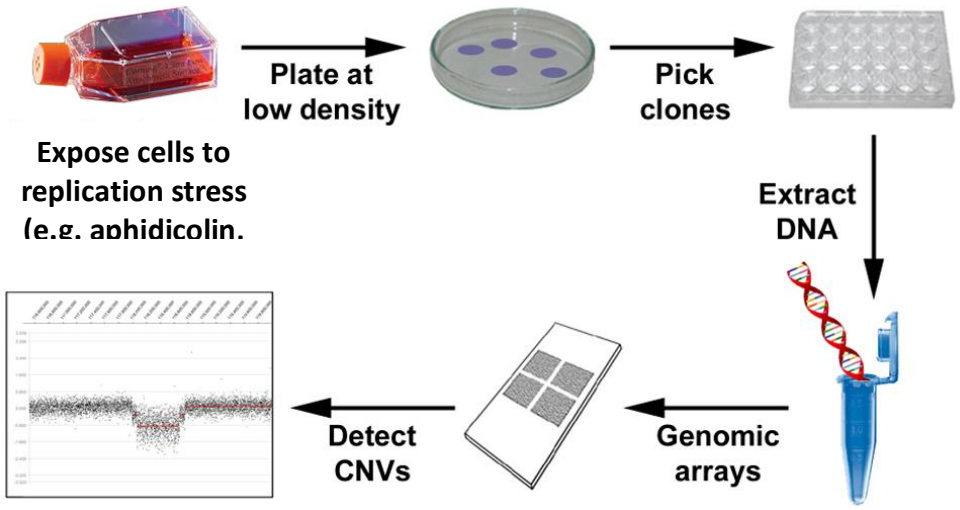


Figure 1.3. Cell culture system can be used to study nonrecurrent CNVs. Cells are first exposed to mild replication stress for a defined amount of time. Afterward, the cells are plated at low density to isolate clonal populations. Clonal populations are picked, and the genomic DNA is extracted for downstream applications such as microarrays.

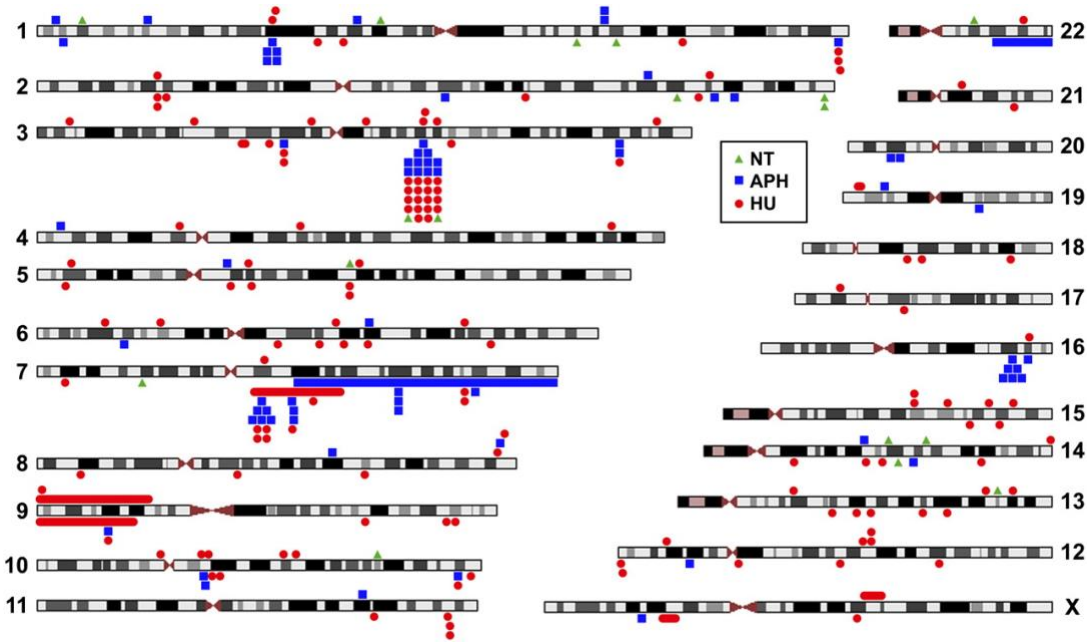


Figure 1.4. Replication stress induces CNVs across the human genome. Each shape represents one observed incidence of a CNV. Green triangles represent spontaneous CNVs that were observed with no replication stress. Blue squares represent CNVs induced by APH. Red circles represent CNVs induced by HU. Duplication CNVs are on top of the chromosome ideograms, and deletion CNVs are on the bottom. Some genomic regions were more prone to CNVs than others. Adapted from Arlt et al (2011).

A

band	gene	CNVs	breaks	known CFS
1p31.1	*NEGR1*	7	28	FRA1C or 1L*
1q44	*KIF26B,SMYD3*	5	1	yes
3p12.3	*ROBO2*	5	12	yes
3q13.31	*LSAMP*	41	14	yes*
7q11.22	*AUTS2*	21	25	FRA7A**
7q21.1	*MAGI2*	5	13	FRA7E**
9p21.3	CDKN2A etc.	6	0	-
10q21.1	*PRKG1*	7	1	FRA10C or 10G
16q23.1-2	*WWOX*	9	10	FRA16D*

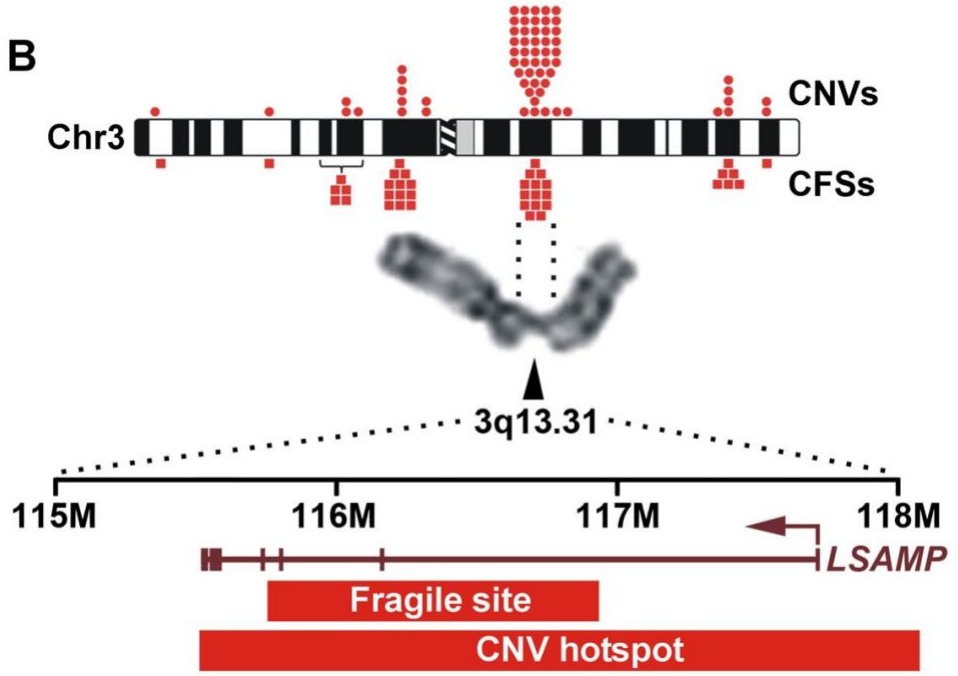


Figure 1.5. CFSs and CNV hotspots are the same loci. (A) Out of nine hotspots in human fibroblast cell line 090, only one hotspot (9p21.3) did not show any gaps or breaks. (B) CNVs on Chromosome 3 are denoted as red circles on top of the chromosome ideogram, CFSs as red squared on the bottom. The metaphase chromosome shows a break at 3q13.31, where the *LSAMP* gene is located. Adapted from Wilson et al (2015).

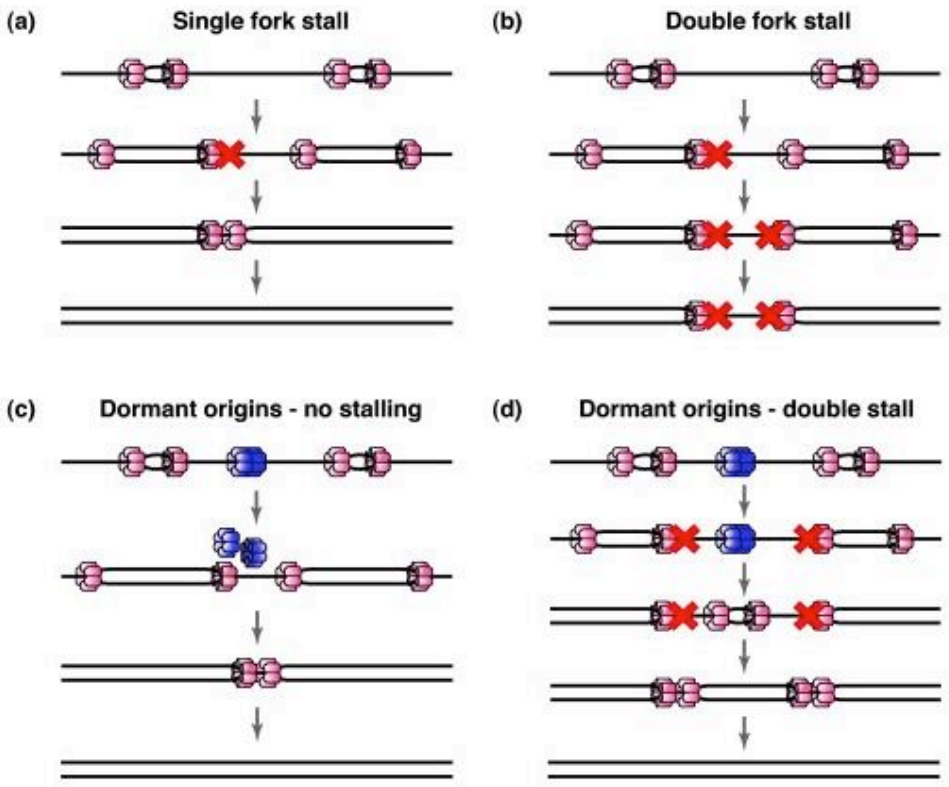


Figure 1.6. Dormant origins fire to ensure complete replication when two neighboring replication forks stall. (A) When one fork stalls, a neighboring fork travels further than usual to converge with the stalled fork. (B) When two neighboring forks stall, the DNA between the forks remains unreplicated. (C) When there are no stalled forks, the flanking origins fire and the forks converge before the dormant origin fires. (D) When two neighboring forks stall and there is a dormant origin between the forks, it fires to complete replicating the region that might otherwise be unreplicated, as in (B). Adapted from Blow, Ge, and Jackson (2011).

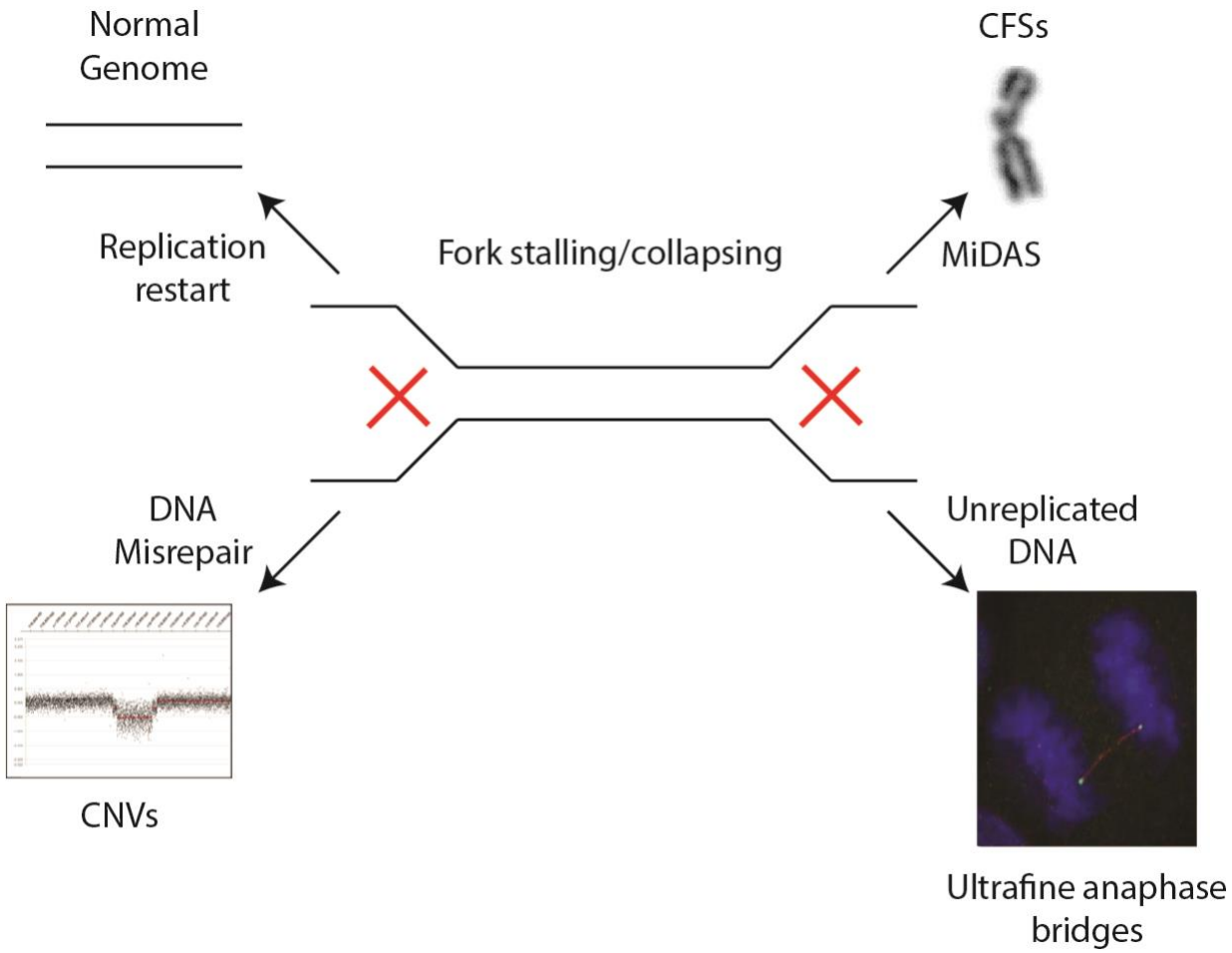


Figure 1.7. Stalled or collapsed replication forks can be resolved in several ways. Top left: A successful restart and completion of replication leads to an intact, normal genome. Top right: If unreplicated regions persist through late S phase, replication may be completed as late as M phase via mitotic DNA synthesis (MiDAS), resulting in a CFS gap or break. Bottom right: If the unreplicated DNA is not resolved and persists to anaphase, ultrafine anaphase bridges can form at these sites. Bottom left: If the stalled or collapsed forks are repaired or restarted through an error-prone mechanism, genome rearrangements like CNVs can occur.

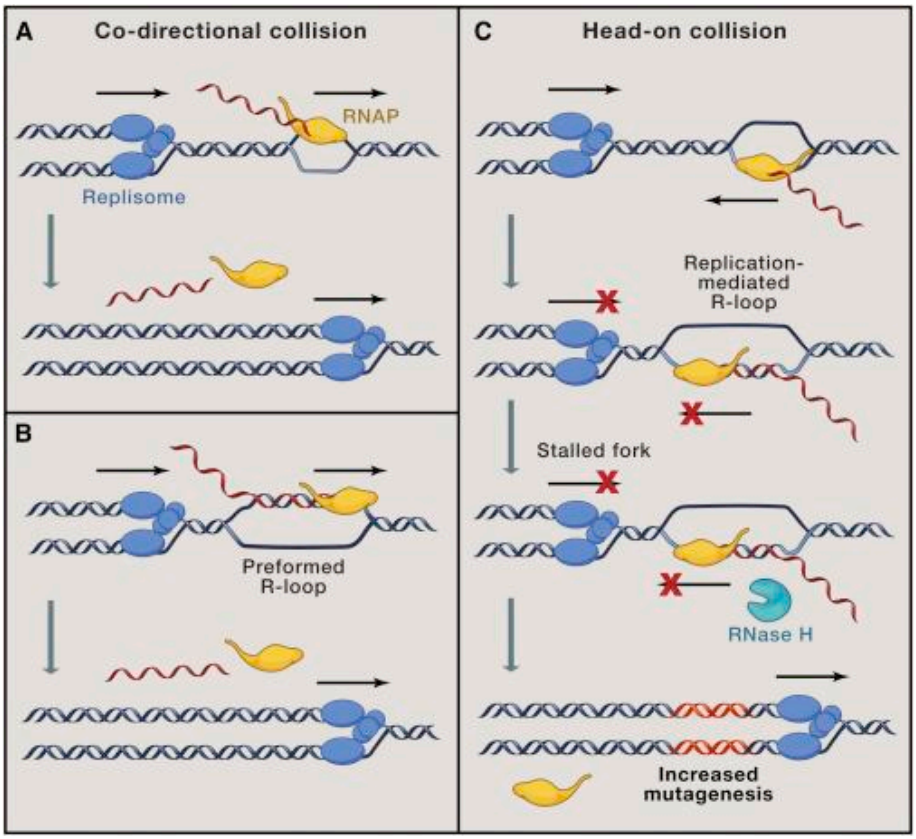


Figure 1.8. R-loops stabilize RNAP during transcription-replication machinery collisions. (A) Co-directional collision, which occurs when the replisome collides with the RNAP from behind, ejects the protein from the DNA template. (B) Similarly, even when there's an R-loop behind the RNAP, the replisome can still eject the protein. (C) In a head-on collision, neither RNAP nor replisome can proceed, which stalls both enzymes before RNase H resolves the R-loop and let the replisome eject RNAP and proceed. Adapted from Lin and Pasero (2017).

Tables

gene[a]	size (Mb)[b]	*de novo* CNVs in this study	*de novo* CNVs in human genomic disease and cancer[c]	references
matching long transcription unit in 090 or HF1 fibroblasts[d]				
LSAMP	2.2	41	osteosarcoma	(Pasic et al. 2010)
MACROD2	2.1	3	major depressive disorder, multiple cancers	(Perlis et al. 2012; Zack et al. 2013)
PDE4D	1.6	2	esophageal adenocarcinoma	(Nancarrow et al. 2008)
AUTS2	1.2	19	ASD, developmental delay	(Mefford et al. 2010; Beunders et al. 2013; Nagamani et al. 2013)
WWOX	1.1	10	multiple cancers	(Beroukhim et al. 2010; Zack et al. 2013)
IMMP2L	0.9	5	Alzheimer's disease, ASD	(Maestrini et al. 2010; Swaminathan et al. 2012)
NEGR1	0.9	11	childhood obesity, cancers, dyslexia	(Jarick et al. 2011; Veerappa et al. 2013; Zack et al. 2013)
MAGI1	0.7	2	bipolar affective disorder, schizophrenia	(Karlsson et al. 2012)
matching long transcription unit in mES cells only [e]				
PTPRD	2.3	4	ADHD, hepatocellular carcinoma, esophageal adenocarcinoma	(Beroukhim et al. 2010; Elia et al. 2010; Nalesnik et al. 2012)
FHIT	1.5	5	ASD, multiple cancers	(Beroukhim et al. 2010; Girirajan et al. 2013; Zack et al. 2013)
NRXN1	1.1	7	ID, ASD, schizophrenia	(Gregor et al. 2011; Tucker et al. 2013)
NLGN1	0.9	3	ASD, ID	(Glessner et al. 2009)
not transcribed in cell lines in this study[d]				
CNTNAP2	2.3	0	schizophrenia, autism, epilepsy, ADHD, dyslexia, multiple cancers	(Friedman et al. 2008; Elia et al. 2010; Mikhail et al. 2011; Veerappa et al. 2013; Zack et al. 2013)
DMD	2.2	0	Duchene muscular dystrophy, multiple cancers	(White and den Dunnen 2006; Beroukhim et al. 2010; Mitsui et al. 2010; Zack et al. 2013)
DLG2	2.2	0	ID, multiple congenital abnormalities	(Vulto-van Silfhout et al. 2013)
NRXN3	1.6	0	ASD	(Vaags et al. 2012)
PARK2	1.4	0	Parkinsonism, ASD, multiple cancers	(Glessner et al. 2009; Mitsui et al. 2010; Wang et al. 2013; Zack et al. 2013)
GRID1	0.8	0	ASD	(Glessner et al. 2009)

Table 1.1. Large genes in CNV hotspots are associated with various human diseases. (a) The table provides examples of genes with human disease-associated CNVs and is not exhaustive. (b) Gene sizes are from the Ensembl annotation. (c) Abbreviations: ASD, autism spectrum disorder; ADHD, attention deficit hyperactivity disorder; ID, intellectual disability. (d) CNV counts are from combined 090 and HF1 fibroblast data. e: CNV counts are from mES cell data. Adapted from Wilson et al (2015).

Chapter 2 — Genome Instability Can Be Predicted from Nascent Transcription Profile

Summary

Instability at CNVs and CFSs are distinct forms of structural chromosome instability followed by replication inhibition. Although they share a common induction mechanism, it is not known how CNVs and CFSs are related or why some genomic loci are much more prone to their occurrence. In Wilson et al., we compared large sets of *de novo* CNVs and CFS gaps/breaks to each other and to overlapping genomic features. We found that CNV hotpots and CFSs occurred at the same human loci in the same cultured cell line. Bru-seq, a nascent RNA sequencing method, further showed that the CNV hotpots (>5 CNVs in a region) and corresponding CFSs were enriched in large, transcribed genes in both human and mouse cells. I show in this chapter that nascent transcription profile of a particular cell line can be used to predict the sites of instability in the same cell line, in which the transcription of the large underlying gene was the determining factor for CFS and CNV hotspot instability. Unlike most transcribed genes, the hotspot genes replicate late and organize deletion and duplication CNVs into their transcribed and flanking regions, respectively, implying that the conflict between transcription and replication might be responsible for the mechanism leading to instability. Altogether, the results indicate that the transcription of large genes drives the locus- and cell-type-specific genomic instability upon replication stress, resulting in both CNVs and CFSs.

Introduction

Recently, our group used large datasets of experimentally induced CNVs to characterize the genomic features contributing to CNV and CFS formation and their distribution profiles (Arlt, Ozdemir, Birkeland, Wilson, et al. 2011; Wilson et al. 2015). We compared the genomic locations between CNV hotspots (defined as regions with five or more CNVs) and CFSs in two cell lines, TERT-immortalized human fibroblasts called 090 and mouse embryonic stem cells. We observed that both forms of genome instability occur at the same loci in a given cell line.

The CNV hotspots and CFSs tend to replicate very late in the cell cycle, as about half of all hotspots replicated in G2 (Figure 2.1); however, late replication alone does not explain why these regions were especially susceptible for CNV formation or CFS expression. About 20% of the entire human fibroblast genome replicates in G2, but not all of these regions are hotspots; in fact, hotspots/CFSs comprise only about 0.4% of the entire human genome. In addition, not all cell types show instability at the same hotspots and CFSs (Letessier et al. 2011), implying that there is a cell-type-specific factor rather than commonly shared characteristics (i.e. DNA sequence) that contributes to the instability at these sites. For example, DNA sequence alone cannot be a determining factor for CNVs and CFSs since all cell types share the same DNA sequence.

Since transcription profiles can vary widely across different cell types, we reasoned that transcription could be the determining factor — in particular, transcription of large genes. It was known that CFSs overlapped with large genes (Mai et al. 2007; D. I. Smith et al. 2006; M. Debatisse et al. 2012), so predictably, the hotspots also overlapped with large genes. But in addition to correlating hotspots to large genes, we combined our previous CNV data and nascent

transcription data (Bru-seq) on 090 cells and demonstrated a strong association between large, active transcription unit (TU) and CNV hotspots.

Bru-seq is a novel technique that maps nascent RNAs using bromouridine (BrU) tagging (Paulsen et al. 2014). The BrU-labeled RNA is pulled down with anti-BrdU antibodies and then is used for construction of cDNA libraries for high throughput sequencing (Figure 2.2). Unlike the traditional RNA-seq, Bru-seq sequences nascent RNAs, not mature RNAs — doing so reveals the entire span of the TU, not just the exons (Paulsen et al. 2014). In addition, Bru-seq differs from other nascent RNA sequencing techniques, such as GRO-seq (Global Run-On Sequencing), such that the BrU labeling occurs *in vivo*, bypassing the need to isolate the nuclei for the nuclear run-on step (Core, Waterfall, and Lis 2008; Paulsen et al. 2013). In addition, other nascent RNA sequencing techniques cannot be used to sequence nascent transcripts in altered environments (e.g. environmental stimuli or stress) since the techniques work well at optimized experimental conditions.

The CNV hotspots showed almost an 80% overlap with a large active TU (larger than 500kb) on average (Figure 2.3). Conversely, 11 out of 12 large active TU (larger than 1 Mb) overlapped with a CNV (Figures 2.4). Together, the two correlations show that the active TU length is a robust cell-type-specific predictor of locus instability under replication stress — as size of TU increases, so does the instability of the corresponding region.

We hypothesized that due to the strong association between large active TU and large genes, that we could choose a hotspot gene based on the Bru-seq profile. Here, I tested whether I could predict where the deletion CNVs and CFSs will occur solely based on the nascent transcription profile in a cell line with no prior CFS or CNV data.

Results

To determine whether our observations establish a predictive model for CNV and CFS formation under replication stress, we performed a prospective study of a distinct human fibroblast cell line called HF1 for which we have reported detailed Bru-seq descriptions, but no prior CNV or CFS data. Like 090, HF1 is a TERT-immortalized normal human fibroblast line. However, Bru-seq revealed differences in their transcriptional profiles (Figure 2.5), including their active TUs larger than 1Mb (Table 2.1). The difference in their transcriptional profile presumably stems from the fact that 090 and HF1 are both human fibroblasts, but they have a different origin. 090 is a normal skin fibroblast cell line from a 37-year-old female (Arlt et al. 2009) whereas HF1 is a normal foreskin fibroblast cell line from an infant (Paulsen et al. 2013).

Some differences between the transcriptional profiles of the two fibroblasts were on-or-off differences, but others were one line expressing a much longer isoform of a gene. Based on these differences, we predicted that HF1 would fail to show CNVs at the most intense hotspot in 090, *LSAMP* (3q13.31), since HF1 did not express the gene. Conversely, we predicted that HF1 would show CNVs at its much longer 1.6-Mb *DAB1* isoform (1p32.1), unlike 090, which expresses only a shorter isoform (<500kb).

I performed SNP microarray analysis on 14 HF1 clones treated with 0.4 μM APH to detect CNVs (Table 2.2). For genes whose long isoforms were expressed in both HF1 and 090, 15 *de novo* CNVs were seen in 14 HF1 clones, and 20 CNVs in 171 treated 090 cell clones. In contrast, we detected only one CNV in the seven genes for which HF1 had a much shorter or no TU, compared to 65 CNVs in 090 clones ($P = 1.4 \times 10^{-7}$ by Fisher's exact test). As we predicted, we saw no CNVs at *LSAMP* in any of the 14 HF1 clones and no CNVs at *DAB1* in any of the 171 090 clones.

To increase statistical power at *LSAMP* and *DAB1*, I scored an additional 36 HF1 cell clones for loss-of-heterozygosity (LOH) at one informative SNP in the middle of the *LSAMP* gene and six informative SNPs in along the length of the *DAB1* gene (Figure 2.6A and B). The LOH indicates a deletion CNV at one of the two alleles. Since the SNPs were informative, I was able to utilize Restriction Fragment Length Polymorphism (RFLP) to detect the LOH in each clone. The presence of only restriction-enzyme cleavable or un-cleavable products indicated LOH and the presence of a deletion CNV (Figure 2.6C), subsequently confirmed by Sanger sequencing (Figure 2.6D).

Consistent with our hypothesis, I detected no LOH events indicative of deletion CNVs in the un-transcribed HF1 *LSAMP* gene, for a total of zero out of 100 HF1 *LSAMP* alleles with CNVs ($P = 0.018$) (Figure 2.6A and E) In contrast, we detected two HF1 CNVs at *DAB1* ($P = 0.051$) in 50 HF1 cell clones (Figure 2.6B and E). Thus, we were able to use the Bru-seq data to correctly predict that *DAB1* has a high risk for CNVs in HF1, a cell line with no prior CNV data.

Moreover, 090 fibroblasts had CNV hotspots and large TUs at both *AUTS2* (7q11.2) and the nearby *MAGI2* (7q21.11) genes, while HF1 showed CNVs and a large TU only at *MAGI2* (Table 2.2 and Figure 2.6F). We therefore predicted that both of these 7q loci would be fragile in 090, but that only *MAGI2* would be fragile in HF1. This prediction seemed to be confirmed on G-banded chromosomes (Figure 2.6G), but these metaphases lacked the resolution required to define the precise band at which all 7q11–q21 breaks occurred. We therefore performed metaphase FISH using BAC probes specific to *AUTS2* and *MAGI2* (Figure 2.6H). Because CFS loci are larger than BAC probes, a CFS will have FISH signals immediately adjacent to, or split by, different breaks. 090 demonstrated this CFS FISH pattern at both *AUTS2* (28 breaks in 188 Chromosome 7 homologs) and *MAGI2* (16 breaks in 188

Chromosome 7 homologs). Consistent with predictions, HF1 only had CFS expression at *MAGI2* (21 breaks in 300 Chromosome 7 homologs) (Figure 2.6I).

Discussion

The question we initially sought to answer was why certain genomic loci are more susceptible to CNVs and CFSs under replication stress. We divided CNV regions into groups ranging from singletons (regions with only one CNV) to hotspots (regions with five or more CNVs). Although the correlation between large TUs and hotspots is the strongest, the association of transcription with CNV formation in all three groups (Fig. 2.3) indicates there is a spectrum of risk for locus instability — a risk that involves transcription.

Previous studies have also implicated transcription in genome instability (Helmrich et al. 2013; Hamperl et al. 2017; De Septenville et al. 2012; Mirkin and Mirkin 2005a; N. Kim and Jinks-Robertson 2012). But how does transcription contribute to genome instability? Does the mechanism involve a possibly antagonistic relationship between transcription and replication? Our future experiments focus on how transcription and impaired replication may interact and lead to structural rearrangements. For our CNV mechanism model, we draw upon two existing models of CNV formation — Fork Stalling and Template Switching (FoSTeS) and MMBIR (Hastings et al. 2009; Lee, Carvalho, and Lupski 2007; F. Zhang et al. 2010). We predict that *de novo* breakpoint junctions at stalled replication forks are created by invasion of nascent DNA strands into ectopic locations characterized by microhomology.

In addition, a mathematical model predicts probability of fork failure at a locus is dependent on the distance that a fork must travel (N) divided by the median distance that forks travel prior to stalling (N_s) (Newman et al. 2013). Based on this equation, inhibiting replication

with replication stress decreases N_s and leads to increased fork failures, FoSTeS/MMBIR events, and ultimately, CNV formation genome-wide.

We extend this logic and propose that increased instability risk within active TUs very likely reflects a transcription-dependent increase in fork failure due to replication-transcription collisions, R-loops that impede fork progression, or fork slowing in transcribed chromatin (Figure 2.7B). In particular, R-loops have shown to play a role in instability at CFSs, presumably by stabilizing the transcription machinery and leading to replication fork failures. Recent studies have shown that accumulation of R-loops can trigger DNA damage response and replication stress can increase the number of R-loops, suggesting that R-loops play a role in genome instability.

Replication and transcription cannot be independent factors since both processes use the same DNA template. Transcription and replication timing are linked, and our study provides an especially powerful comparison by using Bru-seq to monitor transcription of nascent mRNAs, not mature mRNAs, without relying on existing gene annotations since the data reveals the entire span of the transcription unit.

The large TUs display disproportionately increased frequencies of spontaneous and replication-stress-induced CNVs and CFSs. In our model, named Transcription-Dependent Double-Fork Failure (TrDoFF), a double-fork failure (concurrent stalling of two converging replication forks) leads to unreplicated DNA between them. The resolution of unreplicated regions in late S or even G2 phase by error-prone replication restart would lead to CNV formation via FoSTeS/MMBIR (Hastings et al. 2009; Lee, Carvalho, and Lupski 2007; F. Zhang et al. 2010; W. Gu, Zhang, and Lupski 2008).

The TrDoFF model stems from findings that large TUs correspond to late-replicating domains in which forks proceed inward from the boundaries — this will make the middle of large active TUs especially susceptible to instability, which is what we have observed in our cell lines. Although it is known that CFSs and CNV hotspots often replicate late (M. Le Beau 1998; Mi. Debatisse, Achkar, and Dutrillaux 2006), it is the act of transcribing late-replicating DNA that is especially dangerous to the stability of these loci. In other words, late replication alone does not dictate whether a region is unstable; late replication *and* large active TU dictate the instability at these sites.

An implication of the TrDoFF model is that large TUs organize the locations of fork failures in a manner consistent with CNV formation. This is predicted by the FoSTeS/MMBIR model, in which the template switching across the unreplicated DNA will lead to deletions and the switching toward already replicated DNA will lead to duplications (Verdin et al. 2013; Hastings et al. 2009) (Figure 2.7B). Thus, our observations indicate that active large TUs are not only the causative factor for CNVs in the TrDoFF model, but also the determining factor for their profile and distribution. The replication delay was the most severe in the middle of large TUs, which implies that origins are scarce in the middle and the forks will stall surrounding the middle of the gene — leading to deletions in the middle of the gene and duplications flanking the gene. The scarcity of origins in the middle of large TUs is consistent with observations in selected CFSs, where firing of late origins that typically remain dormant but fire under replication stress (Figure 2.7A).

The pre-replication complexes (pre-RCs) are only licensed in G1 and must remain bound for firing to occur in S-phase. Movement of RNA polymerase through an origin can sweep away a pre-RC, as demonstrated in yeast (Snyder, Sapolsky, and Davis 1988; Lõoke et al. 2010). We

thus suggest that dormant origins do not rescue TrDoFFs at CNV hotspots/CFSs because the transcription of the gene has persisted into S and removed those pre-RCs before they can be utilized (Fig 2.7B). Since the genes are so large, it is conceivable that they are transcribed in S, a characteristic unique to these large genes. Most genes are transcribed in G1, presumably to avoid the transcription-replication conflict.

The TrDoFF model, however, does not exclude other forms of transcription-replication conflict (e.g. R-loops). It is possible that the other forms are also contributing to the instability at the hotspots, as seen in previous studies looking at the role of R-loops in CFS expression and DNA damage response (Aguilera and García-Muse 2012; Sollier and Cimprich 2015). In Chapter 4, I tested whether manipulating R-loops changes CNV profiles at the hotspots and genome-wide.

In our study so far, two different human fibroblast cell lines with different Bru-seq profiles were used to associate large active TUs and genome instability, A potential criticism of this approach is that the two cell lines, although they are both human fibroblasts, are still different cell lines. Other differences between the two human fibroblast cell lines might be contributing to the different CNV profiles, not necessarily the differences in transcription. In Chapter 3, I addressed this by creating a mutant cell line that does not express the long isoform of a hotspot gene and comparing the CNV frequency between the mutant cell line and the isogenic, parental cell line.

Materials and Methods

Cell Lines and Culture Conditions

Culturing human 090 fibroblast cell line has been described previously (Arlt et al. 2009, 2011, 2012, and 2014). The HF1 human foreskin fibroblast cell line is the same as used in descriptions of Bru-seq (Paulsen et al. 2013). HF1 cells were cultured in MEM supplemented with 15% FBS, PenStrep, L-Glutamine, and non-essential amino acids.

Bru-seq

Bru-seq methods were previously described (Paulsen et al. 2013). Human 090 fibroblasts were processed by growing cells in the absence or presence of 0.4μM APH for 24 hours, exposing them to Bromouridine (BrU) for 30 min in the same media, immunoprecipitating labeled RNAs, constructing strand-specific RNA-seq libraries, sequencing using Illumina HiSeq, and mapping reads to the reference genome (hg19 build).

De novo CNV Creation and Analysis

Inducing CNVs in 090 cells have been described previously (Arlt et al. 2009, 2011, 2012, 2014). The CNVs were detected using Illumina HumanOmni1 and HumanOmni2.5 BeadChip SNP microarrays and NimbleGen 12 × 270k array comparative genome hybridization (aCGH) with the help from the University of Michigan Sequencing Core.

To create replication stress-induced CNVs in HF1 cells, cells were treated with 0.4 μM APH for 72 hours, followed by a 24-hour recovery period before plating at low density for single-cell clones. Cells were plated at a density of 100-200 cells per 10-cm culture dish and

individual clones isolated using cloning rings after 10-14 days and expanded. Genomic DNA was isolated using Qiagen's Blood and Tissue kit. *De novo* CNVs were detected using the HumanOmni2.5 BeadChip with the help from the University of Michigan Sequencing Core.

We designed PCR primers flanking informative SNPs that conferred a RFLP using primers and restriction enzymes listed in Table 2.3. We designed the PCR primers below to flank the informative SNPs in both 090 and HF1 fibroblasts. PCR was performed on genomic DNAs isolated using the Gentra PureGene kit from APH-treated HF1 cell clones, followed by digestion with the indicated restriction enzymes and agarose gel electrophoresis.

The number of 090 CNV alleles detected by microarray and HF1 CNV alleles detected by microarray or SNP PCR were compared using Fisher's exact test relative to the total number of alleles tested.

CFS Analysis and FISH

CFS breakage was induced by exposure of 090 or HF1 fibroblasts to 0.4μM APH for 24 to 36 hours prior to cell harvest for metaphase chromosome preparations. Cells were fixed onto slides for Giemsa banding or FISH. Chromosome breaks and gaps were analyzed in 100 metaphases from each cell line. At selected loci, the locations of fragile site breaks were refined using BAC FISH probes obtained from BlueGnome (RP5-837C9 and RP11-479M23) hybridized to metaphase spreads.

Figures

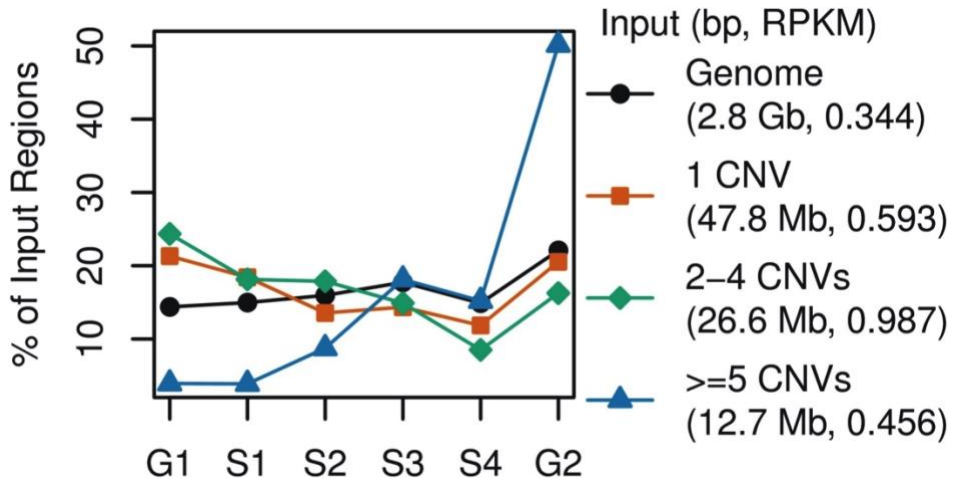


Figure 2.1. CNV hotspots replicate late compared to rest of the genome. Human fibroblast replication timing data for the entire human genome, regions with one, two to four, or more than five CNVs. The numbers on the y-axis represents the average percentage of the regions in each category that was replicating at each cell cycle stage.

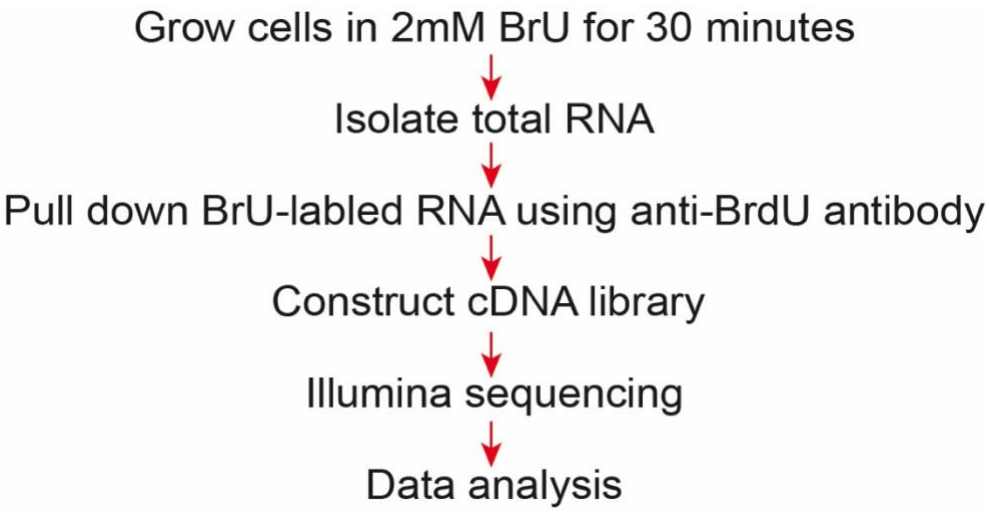


Figure 2.2. A schematic diagram summarizing Bru-seq workflow.

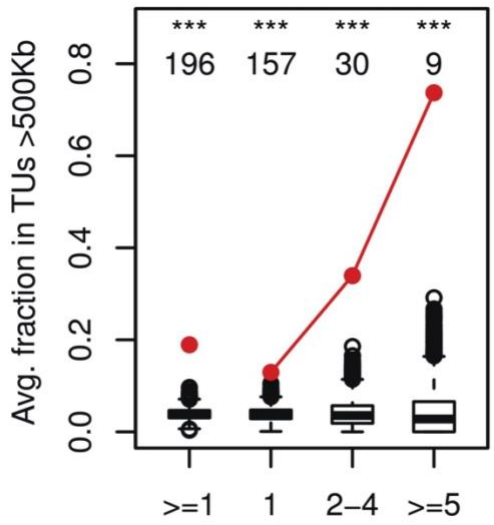


Figure 2.3. CNV hotspots correspond to active large transcription units. The 090 CNV hotspots are enriched for TUs larger than 500 kb (shown in red), shown by the significant deviation from the randomly generated data (shown in black).

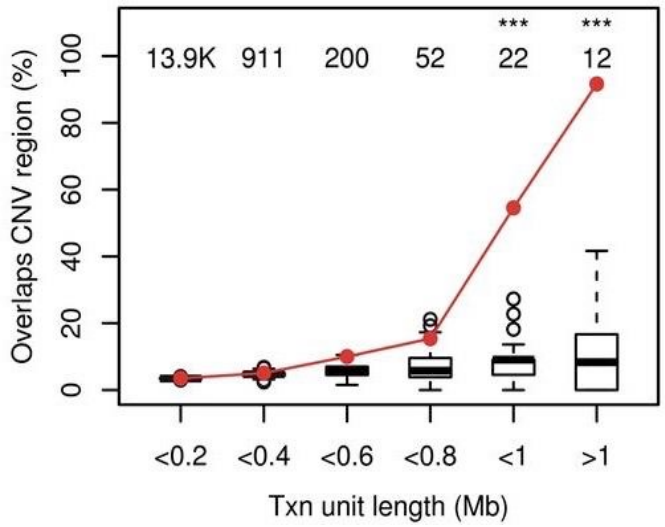


Figure 2.4. CNV hotspots correspond to large active TUs. The large TUs in 090 are enriched for CNVs (shown in red), demonstrated by the significant deviation from the randomly generated data (shown in black).

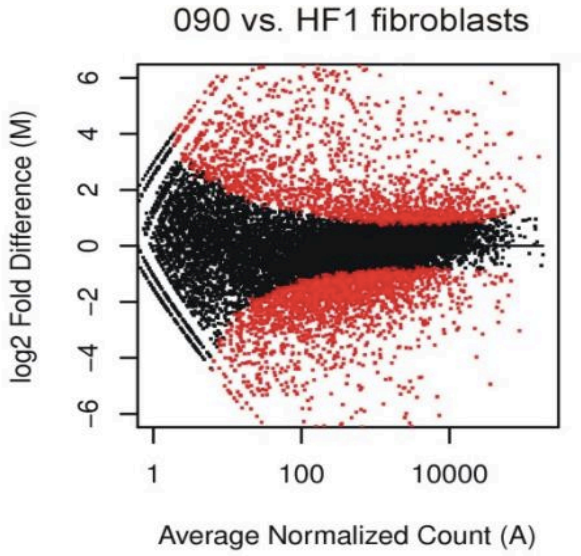


Figure 2.5. MA plots comparing human 090 to HF1 fibroblasts. Each dot represents one gene. Red dots represent genes with a significant inter-sample difference as determined by DESeq (Anders and Huber 2010) at a false discovery rate of 0.05).

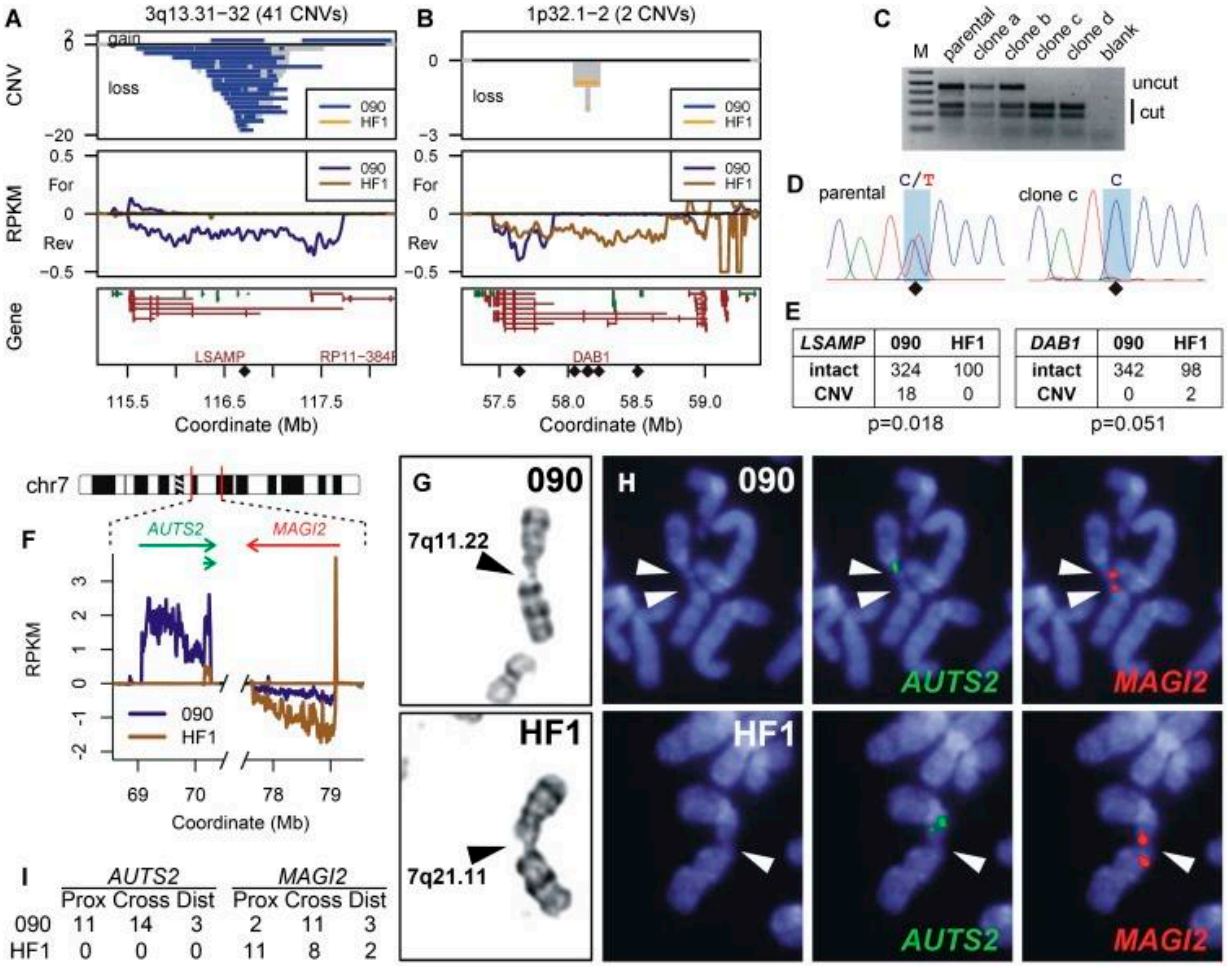


Figure 2.6. Cell-type-specific prediction of unstable loci at active large transcription units. (*A*, *B*) Chromosome region profiles, for genes *LSAMP* and *DAB1*, respectively, with CNVs colored by their detection in either 090 or HF1 fibroblasts. CNVs are drawn as horizontal bars. The number of CNVs overlapping each genome bin is plotted as a gray histogram: positive CNV counts, duplications/gains; negative counts, deletions/losses. Bru-seq transcription data are plotted as follows: positive RPKM, forward transcription; negative RPKM, reverse transcription. Genes are shown as Ensembl transcripts: green lines, forward gene orientations; red lines, reverse orientations. Diamonds mark the positions of SNP RFLPs interrogated in HF1. (*C*) BccI digestion of SNP rs79114629 PCR products for HF1 parental cells and two APH-treated clones lacking (a and b) and containing (c and d) a deletion CNV. (*D*) Sequence analysis of clone c demonstrating LOH at SNP rs79114629. (*E*) Allele counts for *LSAMP* and *DAB1*, where 090 counts only include CNVs from treated clones detectable by the HF1 RFLP analysis. (*F*) Portions of 090 and HF1 Bru-seq transcription data relevant to CFS analysis at 7q11.22–q21.11, showing differential transcription of *AUTS2*. (*G*) Examples of G-banded chromosomes demonstrating breaks at 7q11.22 in 090 (*top*) and 7q21.11 in HF1 (*bottom*). (*H*) Representative FISH on DAPI stained chromosomes using probes to *AUTS2* (green, *middle*) and *MAGI2* (red, *right*). 090 shows breaks at both loci in a single chromosome (*top*), while HF1 shows a break at *MAGI2* (*bottom*). (*I*) Summary of 090 and HF1 CFS breaks with respect to *AUTS2* and *MAGI2* FISH probes.

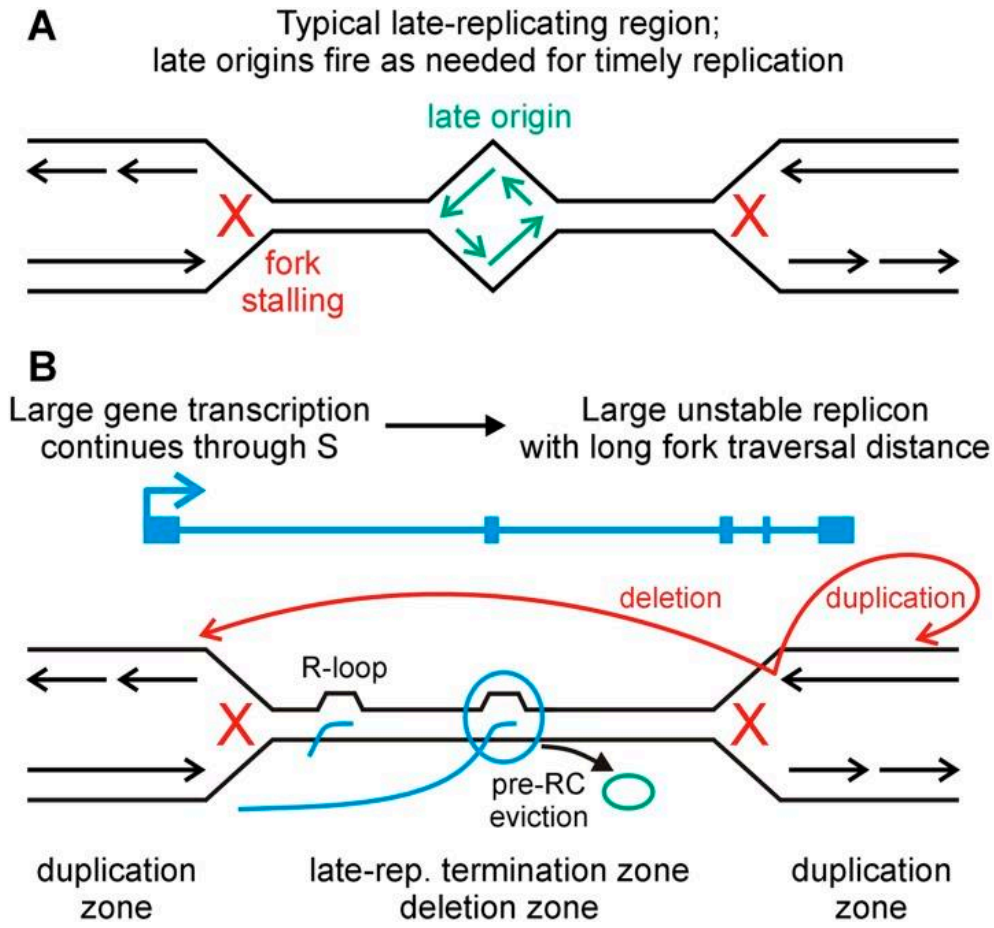


Figure 2.7. Model for CFS and CNV formation at active large transcription units. (*A*) Replication fork failures, even double-fork failures, can usually be rescued by the firing of late "dormant" origins. (*B*) The Transcription-dependent Double-Fork Failure (TrDoFF) model for extreme locus instability under replication stress proposes two mutagenic properties of active large TUs: (1) that they promote simultaneous failure of two converging forks, e.g., through the formation of R-loops; and (2) that they create large late-replicating domains where pre-RC eviction by prolonged transcription into S-phase prevents late origin firing. CFS breaks and deletion CNVs arise in the resulting unreplicated DNA, within the span of the TU, while duplications arise on the flanks, likely by template switching (red arrows).

Tables

Band	Locus	Gene	TU		CNVs	
			090	HF1	090[a]	HF1[b]
Common ≥1 Mb transcripts						
7q21.11	Chr 7: 77609500–79084500	*MAGI2*	1.5 Mb	1.5 Mb	4	2
3q26.31	Chr 3: 174152500–175527551	*NAALADL2*	1.4 Mb	1.4 Mb	4	2
10q11.23–q21.1	Chr 10: 52748500–54074055	*PRKG1*	1.0 Mb	1.3 Mb	6	8
16q23.3	Chr 16: 82659500–83831265	*CDH13*	1.2 Mb	1.2 Mb	3	2
2q33.3	Chr 2: 205409500–206518500	*PARD3B*	1.1 Mb	1.1 Mb	3	1
					20	15
≥1 Mb transcript, 090 only						
3q13.31–q13.32	Chr 3: 115516500–117717500	*LSAMP*	2.2 Mb	–	38	0
20p12.1	Chr 20: 13975500–15286500	*MACROD2*	1.3 Mb	–	1	1
7q11.22	Chr 7: 69062500–70282500	*AUTS2*	1.2 Mb	0.1 Mb	19	0
13q31.3–q32.1	Chr 13: 93879500–95083500	*GPC6*	1.2 Mb	–	1	0
2q34	Chr 2: 212240500–213403500	*ERBB4*	1.2 Mb	–	1	0
3p12.3	Chr 3: 75865500–76961500	*ROBO2*	1.1 Mb	–	5	0
2q21.2	Chr 2: 133429152–134433500	*NCKAP5*	1.0 Mb	0.6 Mb	0	0
					65	1
≥1 Mb transcript, HF1 only						
1p32.1–p32.2	Chr 1: 57454500–59013500	*DAB1*	0.4 Mb	1.6 Mb	0	0
9q33.1	Chr 9: 119162885–120177500	*ASTN2*	0.3 Mb	1.0 Mb	0	0
					0	0

[a]090 counts only include CNVs from treated cell clones, to match HF1 data.
[b]HF1 counts only include CNVs detected by genome-wide microarray.

Table 2.1. CNVs in transcription units larger than or equal to 1Mb in 090 and HF1 fibroblasts

chrom	start	end	kb	method	probes	copies	type	band
1	1527083	1733219	206	aCGH	115	3	gain	1p36.33
1	3518415	3522561	4	aCGH	12	3	gain	1p36.32
1	8961863	9556061	594	aCGH	818	3	gain	1p36.22
1	58048797	58225373	177	RFLP	4	1	loss	1p32.2
1	58138896	58148455	10	RFLP	2	1	loss	1p32.2
1	72088017	72113244	25	aCGH	17	1	loss	1p31.1
1	72239276	72352237	113	aCGH	74	1	loss	1p31.1
1	72326978	72352237	25	aCGH	16	1	loss	1p31.1
1	72401185	72637154	236	aCGH	184	1	loss	1p31.1
1	98035253	98160717	125	aCGH	116	1	loss	1p21.3
1	98117882	98149330	31	aCGH	25	1	loss	1p21.3
1	245742124	246118050	376	aCGH	564	1	loss	1q44
1	245871108	246084652	214	aCGH	320	0	loss	1q44
1	245914673	245942059	27	aCGH	49	1	loss	1q44
2	142374952	142405801	31	aCGH	50	1	loss	2q22.2
2	149055476	149161514	106	aCGH	64	3	gain	2q23.1
2	206256232	206301132	45	aCGH	38	1	loss	2q33.3
3	165940003	166065099	125	aCGH	72	3	gain	3q26.1
3	173689779	174385622	696	aCGH	579	1	loss	3q26.31
3	174813530	174831994	18	aCGH	35	1	loss	3q26.31
4	86874894	87190305	315	aCGH	200	1	loss	4q23.1
4	87119224	87460112	341	aCGH	202	1	loss	4q21.3
6	56436335	56643026	207	aCGH	152	1	loss	6p12.1
6	86295429	86523508	228	aCGH	84	1	loss	6q14.3
7	78231922	78575481	344	aCGH	329	1	loss	7q21.11
7	78547842	78569021	21	aCGH	14	1	loss	7q21.11
7	110656251	110707029	51	aCGH	33	1	loss	7q31.1
7	110665997	110684216	18	aCGH	19	1	loss	7q31.1
10	13715421	13744567	29	aCGH	66	1	loss	10p13
10	53223446	53348048	125	aCGH	147	1	loss	10q21.1
10	53268580	53504174	236	aCGH	233	1	loss	10q21.1
10	53269467	53292516	23	aCGH	26	1	loss	10q21.1
10	53336425	53435774	99	aCGH	88	1	loss	10q21.1
10	53376213	53408296	32	aCGH	33	1	loss	10q21.1
10	53388223	53407270	19	aCGH	19	0	loss	10q21.1
10	53463410	53534471	71	aCGH	70	1	loss	10q21.1
10	53875195	53914073	39	aCGH	35	3	gain	10q21.1
11	1075747	1083946	8	aCGH	13	3	gain	11p15.5
12	1067895	1236729	169	aCGH	129	1	loss	12p13.33
12	1493559	1684076	191	aCGH	294	1	loss	12p13.33
12	44573740	44743021	169	aCGH	77	1	loss	12q12
12	99765812	100058839	293	aCGH	253	1	loss	12q23.1
12	99802508	100220452	418	aCGH	353	1	loss	12q23.1
12	99934156	100262533	328	aCGH	261	1	loss	12q23.1
13	21563311	21575612	12	aCGH	9	1	loss	13q12.11
13	73373325	73390447	17	aCGH	10	1	loss	13q22.1
15	42895210	42941759	47	aCGH	18	1	loss	15q15.2
15	60972325	61088810	116	aCGH	164	1	loss	15q22.2
15	61013134	61021909	9	aCGH	16	0	loss	15q22.2

chrom	start	end	kb	method	probes	copies	type	band
15	60983429	61130639	147	aCGH	206	1	loss	15q22.2
15	71908678	72038676	130	aCGH	147	1	loss	15q23
16	2257374	2262987	6	aCGH	14	3	gain	16p13.3
16	72598884	72696237	97	aCGH	32	1	loss	16q22.2
16	78604831	78689352	85	aCGH	186	1	loss	16q23.1
16	83041537	83247237	206	aCGH	435	1	loss	16q23
16	83165910	83261615	96	aCGH	194	1	loss	16q23.3
18	7859159	7877052	18	aCGH	16	1	loss	18p11.23
18	8101814	8116625	15	aCGH	11	1	loss	18p11.23
18	34345684	34992205	647	aCGH	398	3	gain	18q12.2
20	14568758	14702440	134	aCGH	117	1	loss	20p12.1
22	33922837	33994118	71	aCGH	94	1	loss	22q12.3
X	95963620	96075055	111	aCGH	57	0	loss	Xq21.33
X	96416258	96443397	27	aCGH	27	0	loss	Xq23.33
X	96591587	96793675	202	aCGH	60	0	loss	Xq21.33

Table 2.2. List of APH-induced CNVs in HF1

gene	SNP	primers forward/reverse, 5'-3'	restriction enzyme
LSAMP	rs2090584	TGTAACGGCACAAACTTACAGT TGTCCTTGAGCAGCAGTTTC	*Mbo*II
DAB1	rs1504589	TTCCCTGCCAGACCATATTC TTCACACTTGAGCAGGATG	*Sty*I
DAB1	rs35453940	GGTGACTGGTTAAGCAGTGC ACAAGAAACCGAGGCTCAGA	*Sfc*I
DAB1	rs12566928	GCCATCTTCTTCTCGCTGTG CAGCCTGCATTCATCCATCC	*Psi*I
DAB1	rs79114629	CCCATTCTCACCACAGACCA TCTCTCTGTGCTTGGGGATC	*Bcc*I
DAB1	rs1408138	TGCCCACATCATGCAGAGTA GAATCGCTTCCTTCCGTTCC	*Xcm*I
DAB1	rs4087335	TGAGACTGCCGGCATTAAGA GAGCTGCAATTTGGACCCAT	*Rsa*I

Table 2.3. List of primers and restriction enzyme used for RFLP analysis at each informative SNP for *LSAMP* and *DAB1* genes.

Notes and Contributions

This chapter is an adaptation of a prior publication: Wilson TE, Arlt MF, Park SH, Rajendran S, Paulsen M, Ljungman M, and Glover TW. 2015. Large transcription units unify copy number variants and common fragile sites arising under replication stress. *Genome Res* 25(2): 189-200.

I contributed to the generation of Figure 2.6A-E and all the Tables. Thomas Wilson did the bioinformatics, including the Bru-seq analysis and the association studies between CNV hotspots, large TUs, and replication timing. Martin Arlt performed FISH for 090 and HF1 cells and generated Figure 2.6F-I. Michelle Paulsen was responsible for conducting the wet lab portion of Bru-seq. The model in Figure 2.7 was a collaborative effort.

Chapter 3 — Knocking Down Transcription of a Large Gene Reduces Instability at the Same Locus

Summary

Although nonrecurrent CNVs are important in normal human genetic variation and disease, the molecular mechanisms that lead to their formation is unclear. Treating cells with replication stress-inducing chemicals leads to *de novo* CNVs throughout the genome, but they are often clustered at specific genomic regions. The sites that are more prone to instability, called hotspots, were found to be late replicating and enriched in >500kb transcribed genes that are associated with various human diseases, such as cancer and neurodevelopmental disorders. Between the late replication and transcription of large underlying genes, transcription was the determining factor for CNV hotspot instability. The instability was cell-type-specific, in which the site was only unstable in cell lines that express a long isoform of the underlying gene. In this chapter, I expand on this observation and compare APH-induced CNV frequency at a hotspot gene, *FHIT*, between a wild-type cell line and two mutant cell lines with *FHIT* transcription knocked down by a promoter deletion. Both *FHIT* knockdown mutant cell lines showed reduced APH-induced deletion CNV frequency throughout *FHIT* compared to that of wild-type, although APH did induce deletion CNVs in both the wild-type and the mutant cells at *NLGN1*, another hotspot gene. This demonstrates that the transcription of the large gene is indeed the determining factor for instability at the locus and underscores the importance of monitoring transcription profiles during progression of cancer or development to predict which sites will be the most prone to genome instability.

Introduction

The study covered in Chapter 2 demonstrated a robust association between large active TUs and CFSs/CNV hotspots. It also showed that the presence of a large TU is the determining factor for whether a CFS/CNV hotspot is unstable or not, allowing us to predict the sites of instability based on nascent transcription profiles.

The caveat of the comparison made in Chapter 2 and previous CFS/CNV hotspot studies is that the analysis compared two different cell lines with many differentially expressed genes. Although both 090 and HF1 cells were both human fibroblasts, perhaps the differences in their CFS and CNV profiles were due to other differences between the cell lines, such as chromatin states, not their transcription profiles.

To address this issue, I used the CRISPR-Cas9 technology and created two mutant cell clones with a deletion spanning the promoter and the transcriptional start site of a CFS/hotspot gene. Comparing CFS and CNV hotspot instability between the promoterless mutant clones with the isogenic wild-type (WT) addresses that problem. In addition, I used two independently arisen mutant clones to account for off-target effects from the CRISPR-Cas9 modification.

I used the GM11713 mouse-human somatic cell hybrid cell line and *FHIT* gene for this experiment. GM11713 (A9+3) cell line is a mouse A9 fibroblast cell line that contains one copy of human chromosome 3 tagged with a neomycin resistance gene (Neo^R) (Ning et al. 1992). *FHIT* is a large gene (~1.5Mb long) that is located on human chromosome 3 at a well-characterized CFS called FRA3B. The long isoform of *FHIT* is expressed in GM11713 and is a known CNV hotspot in this cell line (Durkin et al. 2008). CNVs in *FHIT* have been observed in various types of human cancers (Liming Wang et al. 2017; Miyawaki et al. 2012; Karras et al. 2016; Joo et al. 2013; Saldivar, Shibata, and Huebner 2010; Saldivar et al. 2012).

We chose to work with GM11713 cells for several reasons. First, GM11713 cells were used in a previous study from our group for CNV studies. Durkin et al. (2008) characterized APH-induced deletion CNVs at *FHIT* in GM11713 cells and showed that the CNVs are clustered in the middle of the gene (Figure 3.1), resembling the deletion CNVs we observed in hotspots of human fibroblast cells and mES cells.

Second, GM11713 cells have an advantage over other cell lines with respect to genome editing. The cells have only one copy of human chromosome 3, so only one copy of the human *FHIT* promoter had to be deleted to knock down the transcription of the gene. Although the CRISPR-Cas9 technology can be used to modify both alleles, the efficiency of biallelic modification is lower compared to efficiency of a monoallelic modification (Canver et al. 2014; X. Wang et al. 2015; Y. Wu et al. 2017).

Third, since GM11713 cells only have one copy of *FHIT*, deletion CNVs in *FHIT* were easier to detect. In a normal diploid cell, a deletion CNV usually results in LOH; however, in GM11713, a deletion CNV means going from one to zero copies, rather than two to one. This allowed us to more easily measure the percentage of cells that had undergone a deletion CNV in a population using droplet digital PCR (ddPCR), a novel and sensitive technology that enables absolute quantification of nucleic acids in the sample (Hindson et al. 2011). Using ddPCR, if 20% of the GM11713 cells have a deletion CNV at *FHIT*, there is a 20% reduction in the probe signal, as opposed to a 10% reduction in signal in a biallelic cell line.

In this chapter, I describe my strategies to knock down *FHIT* transcript in GM11713 cells and to measure deletion CNVs quickly and reliably in a cell population using ddPCR. Using these two strategies, I provide additional evidence that transcription of the underlying large gene is the determining factor for CNV hotspot instability.

Results

Generating Mutant Cell Lines with No Human *FHIT* Promoter

CRISPR-Cas9 technology was used to delete the *FHIT* promoter in GM11713. Two guide RNAs (gRNAs) were designed to target two sites surrounding the gene's transcriptional start site (TSS). The target sites were chosen such that deletion of the region between the two sites would remove the RefSeq-annotated TSS, some of the CpG islands, sites with the H3K27ac mark, and transcription factor-binding sites, which are either characteristics of mammalian promoters or active transcription (Antequera 2003; Creyghton et al. 2010; Z. Wang et al. 2008) (Figure 3.2A and B). We reasoned that 'such a deletion would render the promoter silent and unable to initiate transcription even at a cryptic TSS, while leaving the deletion hotspot ~500 kb away fully intact.

Clones with successful deletion of the promoter were identified using PCR (Figure 3.2C), in which the flanking primers amplify differently sized products for the WT (no deletion; 2873bp) and promoterless mutant clones (deletion at the *FHIT* promoter; 717bp assuming the end joining occurs at the two cut sites with no resection). The two mutant clones, clone 1 (2-23) and clone 2 (2-36), were further confirmed with Sanger sequencing (Figure 3.2D and E). The two clones showed a 2-bp difference at the breakpoint junction, indicating that they arose independently. Both clones were used for subsequent analyses to control for possible CRISPR off-target effects.

Before using the two promoterless mutants for our CNV experiment, we confirmed that the two mutants had no gross rearrangements in *FHIT* via PCR assays spanning the gene, with assays targeting 7 exons and 1 intron in the gene (Figure 3.2F). The two clones were intact for all

PCR assays, ranging from the second intron to the last exon (Figure 3.2F and Table 3.1), demonstrating that no unexpected internal deletions had occurred.

Confirming Knockdown of *FHIT* Transcription

RT-qPCR and Bru-seq were performed to confirm *FHIT* is knocked down in the two promoterless mutants. RT-qPCR was performed with three different primers targeting different parts of the mature *FHIT* mRNA in case there were multiple isoforms of the transcript. For all three primers, transcription was below the level of detection in both clones 1 and 2 (Figure 3.4A). Bru-seq, which sequences nascent RNAs (Paulsen et al. 2013), was also performed on the two mutants and WT. The sequencing results were consistent with the RT-qPCR data, in which neither mutant showed transcription at *FHIT,* not even a shorter isoform (Figure 3.4B).

Because GM11713 also carries the mouse *FHIT* ortholog, we ran the RT-qPCR products on an agarose gel, sequenced the products, and aligned Bru-seq reads to hg38 and mm10 separately. We confirmed that the RT-qPCR products and the sequencing reads were from human *FHIT* (Supplemental Figures 3.1A-C), including that GM11713 was found to not express murine *Fhit*.

Design of a ddPCR assay for population CNV detection at FHIT

We used ddPCR to measure the percentage of cells with a deletion CNV at exon 5, which is near the middle of the gene and is a region where most deletion CNVs occur (Figure 3.1). We designed two ddPCR probes. To establish a control/reference probe to be used for normalization, we first performed low-coverage whole-genome paired-end sequencing of GM11713 and the two promoterless clones and used read count depth to estimate the copy number of all parts of mouse genome and human chr3. We found that the mouse chromosomes were rearranged and present at

widely varying copy numbers; however, the genome-wide chromosome copy number was highly reproducible between the parental GM11713 and two promoterless clones, suggesting a limited amount of ongoing genomic instability. Based on this information I designed a reference probe targeting a region in mouse chromosome X that has one copy based on the next-generation sequencing data on GM11713 (Supplemental Figure 3.2 and Table 3.2).

The second probe was a test probe targeting near exon 5 of *FHIT*. The test and reference probes were labeled with different fluorophores to enable multiplexing. Thus, in a GM11713 population in which no cell (or below detection limit of ddPCR) had a deletion at exon 5, the ratio of *FHIT* exon 5 to mouse chromosome X should be 1. Any ratio less than 1 suggests that some percentage of cells carry a deletion near exon 5, or possibly a gain of copies of murine chrX, but as noted above we had no reason to anticipate the latter change.

I first applied this assay to WT GM11713 cells treated with 0.6uM APH to test its validity. According to data in Durkin et al. (2008), 23.3% of APH-treated clones carried a deletion CNV crossing where the position of the *FHIT* ddPCR test probe. The ddPCR measurement was consistent with the previous dataset, showing 28.5% reduction in the ratio (Supplemental Figure 3.3) in the APH-treated population but not in untreated population.

Knockdown of Transcription Reduces Deletion CNVs

Clones 1, 2, and WT cells were treated with 0.6uM APH to induce deletion CNVs at *FHIT*, and we used ddPCR to measure the percentage of cells carrying a deletion near exon 5. The WT, again, showed a comparable reduction in test to reference ratio (19.5%) to previous data, whereas neither of the mutants showed a significant reduction (Figure 3.3A), consistent with my hypothesis. To examine whether the deletion CNVs in the promoterless mutants may

have shifted toward one end of *FHIT* or another, I designed additional test probes targeting the 5' and 3' ends of the gene. The probes were designed such that in APH-treated WT, there will be no detectable percentage of cells with deletions. The APH-treated promoterless mutants also did not show significant reduction at the ends (Figure 3.3B and C).

I next designed and performed a ddPCR assay at exon 4 of *NLGN1*, another CNV hotspot gene on human chromosome 3 in GM11713 (Arlt et al. 2009) whose transcription was unaffected the *FHIT* promoter deletion, as expected (Supplemental Figure 3.3). The *NLGN1* ddPCR assay showed that APH did induce deletion CNVs at *NLGN1* in all three cell lines at similar levels, demonstrating that the lack of APH-induced deletion CNVs in clones 1 and 2 was specific to *FHIT* (Figure 3.3D).

To examine the frequency and distribution of FHIT CNVs in more detail, I isolated clones from untreated and APH-treated clones 1, 2, and WT populations to detect deletion CNVs with PCR. Primers were designed and used throughout *FHIT* and surrounding genomic regions (Table 3.4), but we particularly targeted the two introns surrounding exon 5 to increase the likelihood of detecting a deletion CNV (Durkin et al. 2008). Since *FHIT* has large introns, I focused on the two flanking introns and exon 5, which covered approximately one-third of the annotated gene length (Figure 3.4A).

For the APH-treated group, I analyzed 18 clones for clone 1, 16 clones for clone 2, and 32 clones for GM11713. For both promoterless mutants, no deletions were detected with the PCR assays. Ten deletions were detected in APH-treated GM11713, which was significantly higher ($P \leq 0.05$) than in both mutants (Figure 3.4B). Eight of the nine deletions spanned across where the ddPCR test probe amplify, which was comparable to the proportion we detected with ddPCR (15.6% versus 23.3%). Two deletions spanned beyond *FHIT*, at least approximately

1.5Mb distal from the gene. In the untreated group, there were no detected deletions in *FHIT* for any of the tested clones, and there was no significant difference between the two mutants and the WT (Figure 3.4C).

Discussion

In this chapter, I have addressed a limitation of the experiments in Chapter 2 and tested whether manipulating transcription of a large underlying gene could alter APH-induced CNV frequency at the region. The presented data show that turning off transcription of large underlying gene reduces instability at the same locus, providing further evidence that transcription of the large gene is the determining factor for CNV hotspot instability.

The two mutant cells with *FHIT* transcription knockdown possessed a significantly lower number of APH-induced deletion CNVs at middle of *FHIT* (to an undetectable level) compared to the unedited WT. In addition, neither mutant showed a detectable amount of deletion CNVs at the 5' and 3' ends of *FHIT*, suggesting that the absence of transcription did not shift the CNVs toward one end of the gene or the other. The APH did still induce instability at another hotspot gene, *NLGN1*, in the two mutants, demonstrating that the APH still affected the cells in an expected way at another hotspot, but not at *FHIT*.

Analyzing individual APH-treated clones further illustrated that the two promoterless mutants had significantly fewer deletion CNVs compared to the WT throughout *FHIT*. Neither mutant had any deletion CNVs in the two introns surrounding exon 5, which is where at least one endpoint of all 17 deletion CNVs from Durkin et al. (2008) study were located. Together with the ddPCR data, the PCR data imply that the two mutants had no detectable CNVs across the entire gene span. These data are remarkable for the large distance of the CNV suppression effect

relative to the position of the promoter deletion, which indicates that the comparatively small 2kb deletion we introduced altered a functional property of a much larger region of human chr3. That property is almost certainly *FHIT* transcription.

Using two independently generated promoterless mutants for same experiments addressed the potential off-target effects of CRISPR-Cas9. This was particularly important for validating our findings since several recent studies have reported on the pitfalls of using CRISPR-Cas9, specifically emphasizing the introduction of numerous unanticipated off-target mutations and alterations (Anderson et al. 2018; Fu et al. 2013; Cho et al. 2014; Kosicki, Tomberg, and Bradley 2018). Nonetheless, for our study CRISPR did provide a straightforward method to delete almost a 2-kb region in the genome, effectively turning off transcription at *FHIT*. This dual-gRNA deletion method has also been successfully used in other studies using different model organisms (Bauer, Canver, and Orkin 2015; Zheng et al. 2014; Aparicio-Prat et al. 2015; X. Chen et al. 2014; Zhao et al. 2016), so it could be useful in various genetic studies investigating loss-of-function genes and genetic elements.

This study also demonstrates that ddPCR is a useful technology for CNV analysis in a population of cells. The traditional method of analyzing CNVs requires analyzing individual clones (e.g. SNP microarrays and PCR), which could be extremely time-consuming. If a specific region of interest is known, ddPCR allows in a quicker CNV detection on a population level, bypassing the cloning step. The ddPCR near exon 5 of *FHIT* and PCR analysis of individual clones show that the ddPCR results are consistent with the more traditional methods.

The two mutants generated during this study are valuable tools in further characterizing the role of transcription at CNV hotspots. Although the study further validated the importance of transcription in hotspot instability, the mechanistic role of transcription is still unknown. Based

on the TrDoFF model from Chapter 2, one possible future direction is comparing the origin activity between the two mutants and WT. In the light of transcription-replication conflicts and genome instability, it will be intriguing to investigate whether the absence of transcription changes origin activity in middle of the underlying large gene. For instance, is the replication timing at *FHIT* different in the promoterless mutants? If so, it would be interesting to determine whether transcription alters replication timing and origin activity by ejecting the licensed origins, by interfering with origin firing, or by another mechanism.

Finally, the implications of this study demonstrate the need to monitor the static and evolving transcription profiles during the progression of diseases and development to predict which sites will be the most prone to genome instability. During both cancer progression and normal development, cells go through rounds of division. If large genes are also transcribed in actively dividing cells, the corresponding sites become susceptible to genome instability since the risk of double-fork failures increase (Wilson et al. 2015). Studies have reported focal deletions at large genes in various tumors (Beroukhim et al. 2010; Zack et al. 2013; Bignell et al. 2010; Rajaram et al. 2013; Glover, Wilson, and Arlt 2017), and some of these deletions are hypothesized to be early events and are known tumor suppressor genes, thus potentially impacting the progression of the disease (Lai et al. 2010; Guler et al. 2005; Dagmar et al. 1997; Finnis et al. 2005; Glover, Wilson, and Arlt 2017).

Similarly, large genes relevant for development, particularly neuronal development, are at high risk for instability. Several studies have reported deletions and DSB clusters at large genes that are important for neuronal development (P.-C. Wei et al. 2016; Wilson et al. 2015; Glover, Wilson, and Arlt 2017; D. I. Smith et al. 2006). Wei et al. in particular have shown that 24 out of 27 recurrent DSB clusters are located in large genes linked to neuronal development

and function, and nearly all are linked to neurodevelopmental or psychiatric disorders, such as autism-spectrum disorder, schizophrenia, bipolar disorder, and intellectual disability.

Overall, the experimental evidence in this study provides a definitive demonstration of the importance of local transcription in CNV hotspot instability and can help elucidate the patterns of genome instability in actively dividing cells. However, the exact role of transcription is unclear and still needs further investigation. In Chapter 4, I explore one possible contributing factor — R-loops.

Materials and Methods

Cell Lines

GM11713 human-mouse somatic cell hybrids were cultured and maintained as described in Durkin et al (2008).

Next Generation Sequencing of GM11713

GM11713 genomic DNA was isolated using Qiagen's Gentra PureGene Cell kit. The sequencing libraries were prepared by the University of Michigan Sequencing Core staff, and the samples were run on Illumina HiSeq 4000 using the paired end 150-cycle. The reads were aligned to both hg38 and mm10 builds to accommodate for the genomic nature of GM11713. Copy numbers were determined related to a "lowest common denominator" region on chrX, which had a copy number of one.

Generating *FHIT* Promoterless Mutant Cell Lines

CRISPR-Cas9

Guide RNAs (gRNAs) surrounding the *FHIT* promoter region were designed using CRISPRdirect (Naito et al. 2015) and CHOPCHOP (Montague et al. 2014; Labun et al. 2016). The gRNA oligos were synthesized by Integrated DNA Technologies. See Table 3.3 for gRNA sequences.

We integrated the gRNA oligos into pSpCas9(BB)-2A-GFP plasmids (Addgene cat. no. 48138) using the protocol in Ran et al. 2013. Once the oligos were integrated and confirmed via Sanger Sequencing, the plasmids were transfected into GM11713 using Lipofectamine 2000 (Thermo Fisher cat. no. 11668027) using the manufacturer's instructions. See Table 3.3 for primers to confirm successful integration.

To select for cells that have successfully undergone transfection and *FHIT* promoter deletion, we followed the protocol in a previous study with the following modifications (Bauer, Canver, and Orkin 2015). 48 hours post-transfection, the top 3% GPF^{+} cells were sorted using iCyt's Synergy Cell Sorting System at the University of Michigan's Flow Core. The sorted cell population was plated in a single well in a 96-well plate for recovery.

When wells were confluent, approximately 50 cells were plated per well in a 24-well plate. These populations were screened with primers flanking the two gRNA target sites (Table 3.1) to see which populations contain the promoterless mutants. Primers were designed using Primer3Plus based on hg38 build. Genomic DNA was extracted using Qiagen's Gentra PureGene Cell kit.

The screening PCR was performed using Phusion enzyme from New England BioLabs Inc. (cat. no. M0530), using the reaction condition provided by the company using 0.2uM for

each primer and 3% DMSO (v/v). The PCR cycling condition consisted of an initial denaturation at 98°C for 30 seconds followed by 35 cycles of 98°C for 5 seconds, 60°C for 10 seconds, and 72°C for 30 seconds, final extension at 72°C for 5 minutes, and hold at 4°C.

The confirmed populations were plated at low dilution (50-100 cells per 10cm dish), and individual clones were picked approximately one week afterward for another round of PCR screening using the same flanking primers that were used to screen the populations. We confirmed the two candidate promoterless mutants (Clone 1: 2-23 and Clone 2: 2-36) with the flanking primers using Sanger sequencing at the University of Michigan sequencing core.

RT-qPCR

The total RNA was isolated using Qiagen's RNeasy mini kit (cat. no. 74104). Then the RNA was reverse transcribed to complementary DNA using the High-Capacity cDNA Reverse Transcription kit from Thermo Fisher (cat. no. 4368814), and qPCR was performed using Qiagen's QuantiTect SYBR green PCR kit (cat. no. 204143) and Applied Biosystems 7500 Real-Time PCR system. The cycling conditions were 50°C for 10 minutes, then 95°C for 15 minutes, followed by 40 cycles of 15 seconds at 94°C, 30 seconds at 60°C, then 30 seconds at 72°C. Mouse *ActB* was as a control gene to calculate ΔΔCt values. See Table 3.1 for primer information.

Bru-seq

Bru-seq was performed as described in Paulsen et al. The sequencing reads were aligned to both human build (hg38) and mouse build (mm10), separately, to ensure that the human *FHIT* reads are specifically aligning to human *FHIT*, not mouse *Fhit.*

Inducing and detecting deletion CNVs

APH treatment

GM11713 cells were treated with 0.6uM APH (Sigma-Aldrich cat. no. A0781) for 72 hours to induce the CNVs followed by a 24-hour recovery period in the absence of APH. During the 72-hour treatment period, the media was changed every 24 hours to provide fresh APH. After recovery, some of the cells were plated at low dilution to collect clones, and some were collected for DNA extraction. Genomic DNA of the population was isolated using Qiagen's Gentra PureGene kit for both ddPCR and clone analyses.

ddPCR

ddPCR was performed using the Bio-Rad QX200 system. The Taqman ddPCR assays were designed using Bio-Rad's Droplet Digital PCR assay tool online, using the FAM/HEX fluorophore combination for multiplexing. The test assays were labeled with FAM, and the reference assay was labeled with HEX. See Table 3.3 for sequence of the amplicons.

The reactions were set up using the ddPCR Supermix for Probes (No dUTP) (cat. no. 1863023) as described in manufacturer's instructions. The genomic DNA was digested with HindIII for at least two hours at 37°C prior to PCR, and 2uL of the digestion reaction directly added to each PCR reaction.

Detecting Deletion CNVs in FHIT *using PCR*

PCR primers that were used to detect deletion CNVs were designed using Primer3 based on hg38 build (Table 3.1). PCRs were carried out using native Taq polymerase from Thermo Fisher (cat. no.18038042), following the company's instructions using 0.2uM of each primer, 0.2mM dNTPs, and 1.5mM $MgCl_2$. The PCR cycling condition consisted of an initial denaturation at 94°C for 3 minutes followed by 40 cycles of 94°C for 45 seconds, 58 to 65°C for 30 seconds, and 72°C for 30 seconds to 1 minute, final extension at 72°C for 5 minutes, and hold at 4°C. Deletions were detected by running the PCR amplicons on an agarose gel, where an absence of the product was considered as a deletion.

Figures

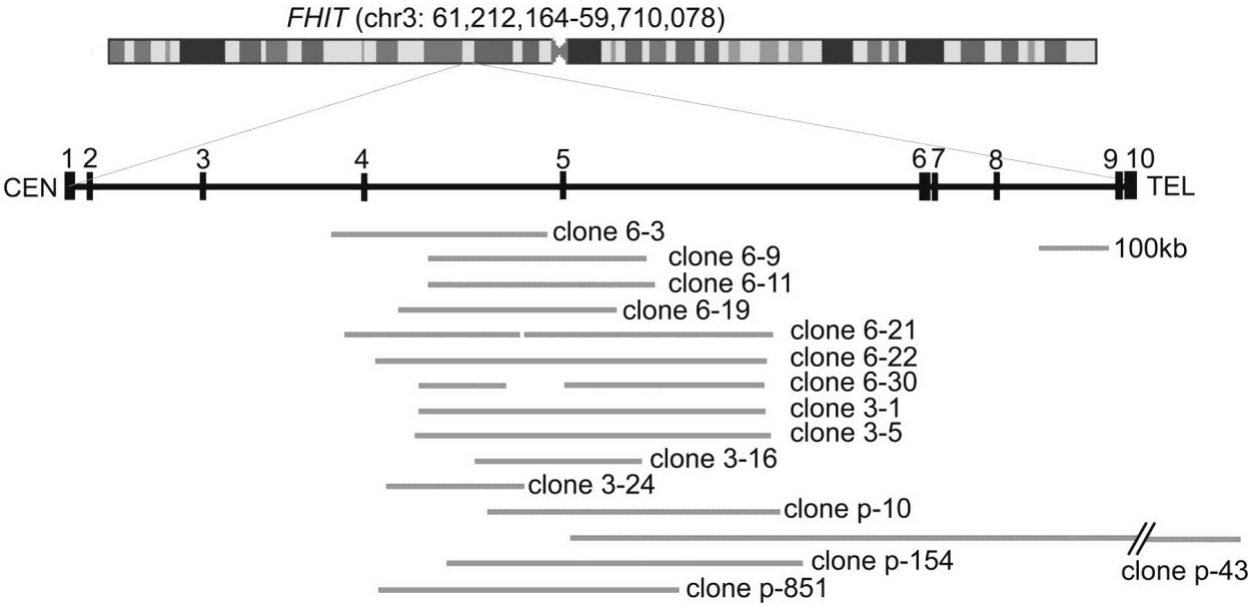
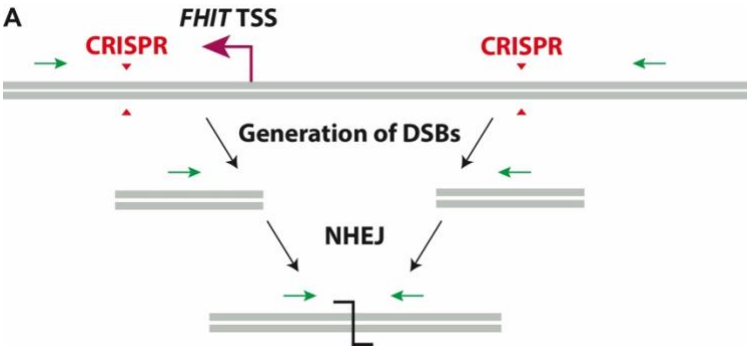
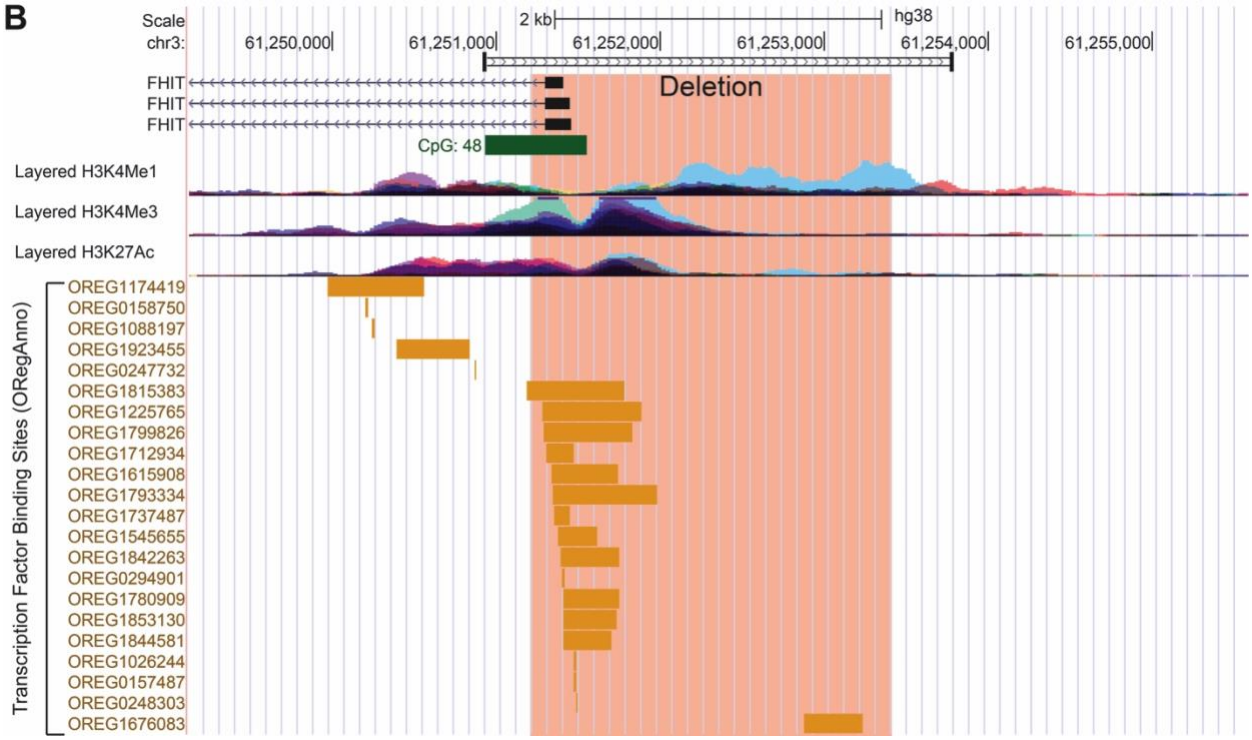
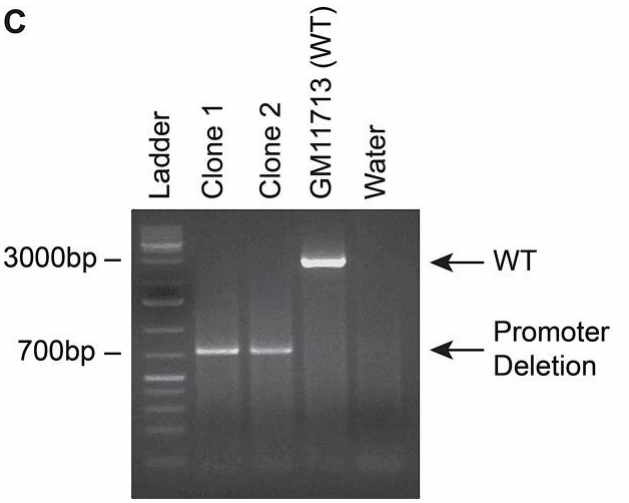


Figure 3.1. APH-induced and spontaneous deletion CNVs at *FHIT* in GM11713 cells. CEN: Centromeric and TEL: Telomeric. The numbered vertical lines represent exons. Adapted from Durkin et al (2008).


A
FHIT TSS
CRISPR
CRISPR
Generation of DSBs
NHEJ


B
Scale
chr3:
2 kb
hg38
61,250,000
61,251,000
61,252,000
61,253,000
61,254,000
61,255,000
FHIT
FHIT
FHIT
Deletion
CpG: 48
Layered H3K4Me1
Layered H3K4Me3
Layered H3K27Ac
Transcription Factor Binding Sites (ORegAnno)
OREG1174419
OREG0158750
OREG1088197
OREG1923455
OREG0247732
OREG1815383
OREG1225765
OREG1799826
OREG1712934
OREG1615908
OREG1793334
OREG1737487
OREG1545655
OREG1842263
OREG0294901
OREG1780909
OREG1853130
OREG1844581
OREG1026244
OREG0157487
OREG0248303
OREG1676083


C
Ladder
Clone 1
Clone 2
GM11713 (WT)
Water
3000bp –
700bp –
WT
Promoter
Deletion

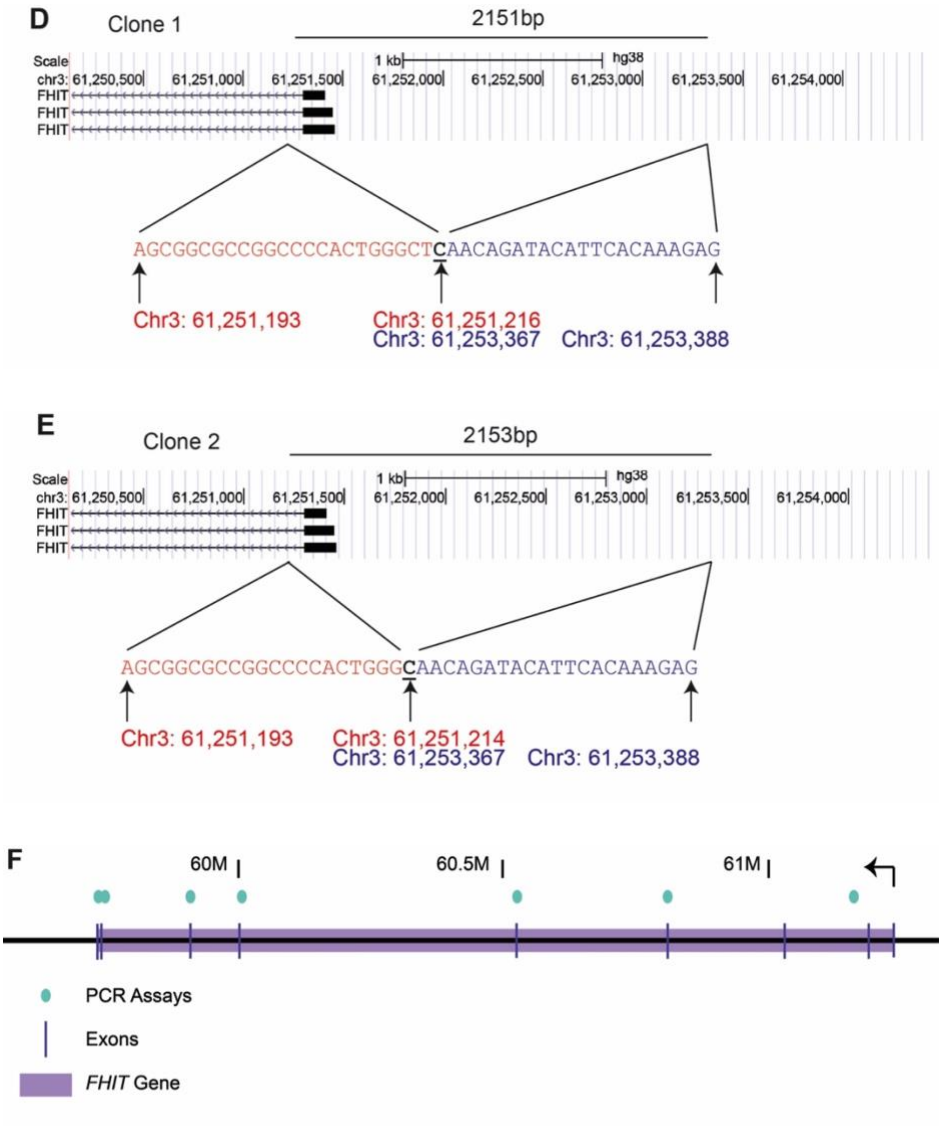


Figure 3.2. Mutants with deletion in *FHIT* promoter were generated using CRISPR-Cas9 and were confirmed by PCR and Sanger sequencing. (A) A schematic describing how the promoterless mutants were generated. Two gRNAs (gRNAs 1 and 2) surrounding the TSS (Transcriptional Start Site) and promoter were transfected into GM11713, inducing DSBs indicated by the red arrows. Mutants that deleted the region between the two break sites were detected using the green PCR primers flanking the break sites. See Table 3.1 for gRNA sequences and primer sequences. (B) A screenshot from USCS Genome Browser showing the PCR products from the green flanking primers from (A), deleted promoter region in light red box, and various features of the region, including CpG islands, histone marks, and transcription factor binding sites. Coordinates are in hg38 build. (C) A picture of an agarose gel showing PCR amplicons generated by the green flanking PCR primers from (A). PCR product sizes are 722bp (Clone 1), 720bp (Clone 2), and 2873bp (WT). (D and E) Sanger sequencing results showing the deletion breakpoints in clones 1 and 2. The red and blue segments show the two flanking segments that were joined to create the promoter deletion, and the black, underlined nucleotide represents a region of microhomology. (F) PCR assays were performed throughout the *FHIT* gene for the two mutants and WT to confirm there are no gross rearrangements or deletions. Aqua dots represent PCR assays, vertical purple lines exons, and light purple box the annotated *FHIT* gene. Name of PCR assays from left to right: Exon 10, Exon 9, Exon 8, Exon 7, Exon 6, Exon 5, Exon 4, and Intron 2. See Table 3.2 for primer information.

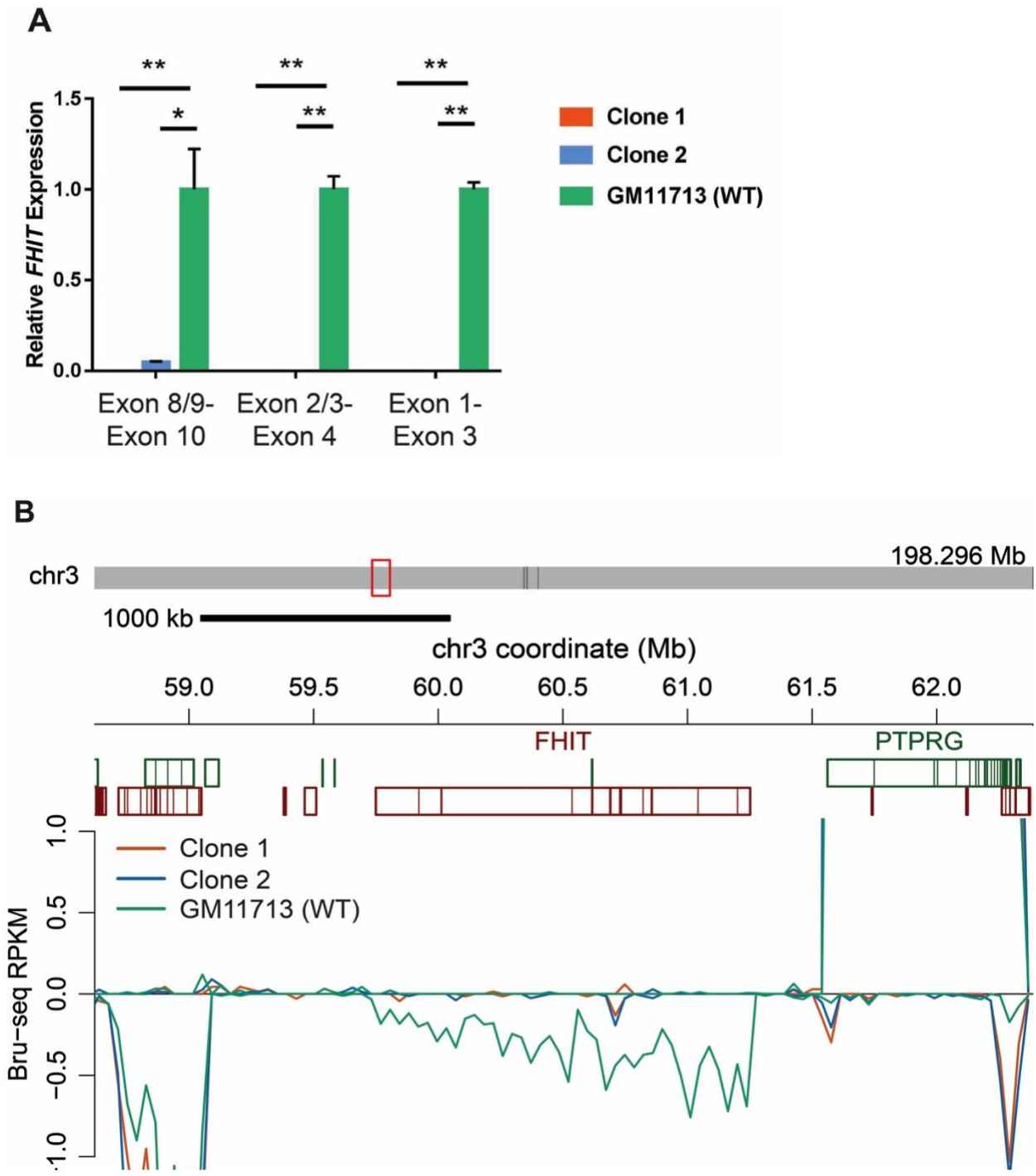


Figure 3.3. Promoterless mutants do not express *FHIT*. (A and B) Clones 1 and 2 show almost no *FHIT* transcription, shown by three RT-qPCR assay targeting different parts of *FHIT* mRNA (B) and Bru-seq (C). Statistical analyses were performed using a student's T-test. *: P≤0.05 and **: P≤0.005. See Table 3.3 for primer information.

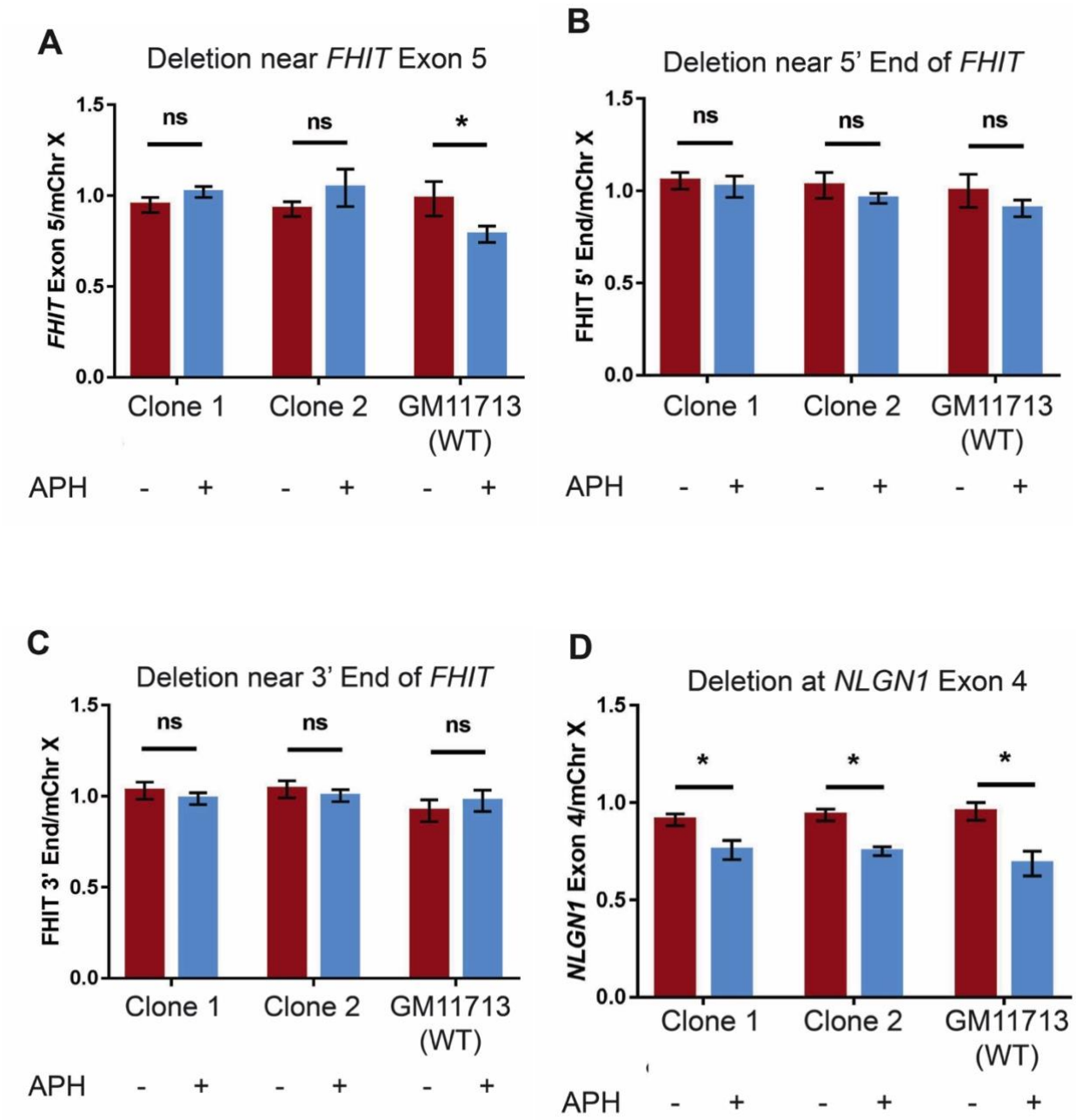
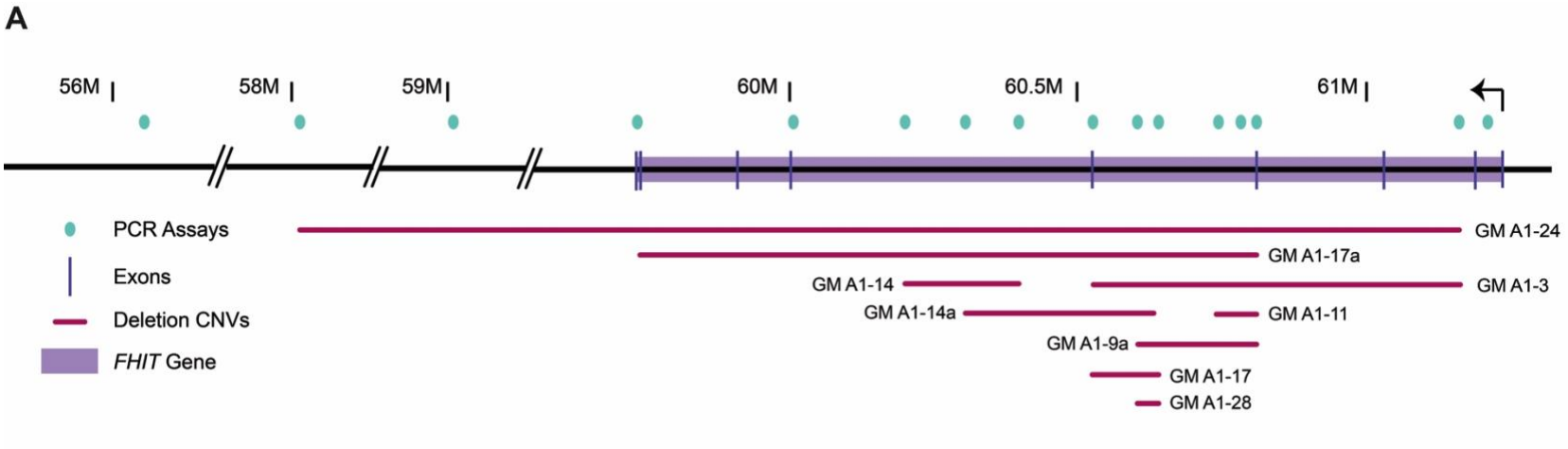


Figure 3.4. Transcription knockdown at *FHIT* reduces instability at the gene at a population level. (A) ddPCR results show that APH-treated clones 1 and 2 did not have significantly more deletion CNVs near exon 5 of *FHIT* compared to the untreated controls, whereas APH-treated WT had significantly more deletion CNVs than the untreated control. The ratio represents the average of three biological replicates, and the error bars represent the standard deviation among the replicates. Statistical analyses were performed using a student's T-test. ns: $P>0.05$ and *: $P\leq0.05$. (B and C) The two mutants and WT had no significant level of deletion CNVs near the 3' and 5' ends of *FHIT* compared to the untreated controls. Statistical analyses were performed using a student's T-test. ns: $P>0.05$. (D) Clone 1, clone 2, and WT show significant amount of APH-induced deletion CNVs at exon 4 of *NLGN1* compared to the untreated controls. Statistical analyses were performed using a student's T-test. *: $P\leq0.05$.


A
56M
58M
59M
60M
60.5M
61M
PCR Assays
Exons
Deletion CNVs
FHIT Gene
GM A1-24
GM A1-17a
GM A1-14
GM A1-3
GM A1-14a
GM A1-11
GM A1-9a
GM A1-17
GM A1-28

B

Deletion CNVs in *FHIT*

		No	Yes
+ APH	Clone 1	18	0 *
	Clone 2	16	0 *
	GM11713 (WT)	23	9

C

Deletion CNVs in *FHIT*

		No	Yes
- APH	Clone 1	18	0
	Clone 2	16	0
	GM11713 (WT)	30	0

Figure 3.5. Transcription knockdown at *FHIT* reduces instability at the gene at clonal level. (A) Diagram showing the distribution of deletion CNVs. Aqua dots represent the location of each PCR assay, vertical purple lines represent exons in *FHIT*, horizontal magenta lines represent deletion CNVs, and the light purple box shows the annotated *FHIT* gene. The span of the deletion CNVs were determined with PCR assays, in which ends of each CNV on the diagram are defined as the location of the last PCR assay that shows a deletion before the next immediate PCR assay shows an intact genomic region. The PCR assays used are as follows, from distal to proximal: Tel 3, Tel 2, Tel 1, Exon 10, Exon 6, Intron 5-5, Intron 5-4, Intron 5-3, Exon 5, Intron 4-5, Intron 4-4, Intron 4-7, Intron 4-1, Exon 4, Intron 2, Intron 1. See Table 3.1 for primer information. (B and C) Tables summarizing the number of clones that either did not have or had a deletion CNV in *FHIT*. Statistical analyses between each mutant and WT were performed with Fisher's exact test for 2x2 table. *: $P \leq 0.05$.

Tables

gRNA/Primer Name	Sequence (5'-3')
FHIT gRNA 1 Top	CACCGGGTGATGACCTGACGCGCG
FHIT gRNA 1 Bottom	AAACCGCGCGTCAGGTCATCACCC
FHIT gRNA 2 Top	CACCGTATCTGTTGCTCAAGGTCGT
FHIT gRNA 2 Bottom	AAACACGACCTTGAGCAACAGATAC
FHIT Deletion Confirmation F	CAGACCTGTTGGGACGGATT
FHIT Deletion Confirmation R	AAGCTGCCCCCAATCATTGA

Table 3.1. Sequences of gRNAs that were used to delete the *FHIT* promoter and primers that were used to screen and confirm the promoterless mutant clones. The red sequences represent the 20-nt guide sequence for each gRNA.

Primer Name	Sequence (5'-3')
FHIT Intron 2 F	CCATACTGACAGCTGCCTCT
FHIT Intron 2 R	TTTGTCGCCTCAACAGAACG
FHIT Exon 4 F	ACCTCTTCGGAGTCCTCAGT
FHIT Exon 4 R	GCCACACAGGACACTTCAGA
FHIT Exon 5 F	GGAATTTGCTTTCTTTAAGGTGGTA
FHIT Exon 5 R	CTCAGCTATGGTAGTGAAAAGGTCA
FHIT Exon 6 F	GACCTGCGTCCTGATGAAGT
FHIT Exon 6 R	TGCATGGAAAAGGTGAGAGA
FHIT Exon 7 F	CAATGTGCTGCATATGACTCG
FHIT Exon 7 R	TCTAGGATGGCCCCGAAG
FHIT Exon 8 F	CTTCCTGGGAAGAACATGGA
FHIT Exon 8 R	TCGTGGAGTAATTGGGCTTC
FHIT Exon 9 F	GAGGGTCTGGGTAATGACGA
FHIT Exon 9 R	TGTTCAAGGAGATCCCAAGG
FHIT Exon 10 F	TCTCCCCAAGAGGAACTGAA
FHIT Exon 10 R	GCTGCCGAATAAGGAGACAG

Table 3.2. Primers that were used to verify the absence of gross genome rearrangements in *FHIT* for two promoterless mutant clones.

Primer Name	Sequence (5'-3')
FHIT Assay 1 F	AGCTCCAGAAACATGACAAGG
FHIT Assay 1 R	CTGGTTGGCAATAGCTCTTTTG
FHIT Assay 2 F	GTCTGTAAAGGTCCGTAGTGC
FHIT Assay 2 R	AGACTGCTACCTCTTCGGAG
FHIT Assay 3 F	CTCCCTCTGCCTTTCATTCC
FHIT Assay 3 R	CTTTCTCTTTCTCTCCCTTCCAC
Mouse Actb Assay F	ACCTTCTACAATGAGCTGCG
Mouse Actb Assay R	CTGGATGGCTACGTACATGG

Table 3.3. qPCR primers that were used to confirm that *FHIT* is not expressed in the two promoterless clones.

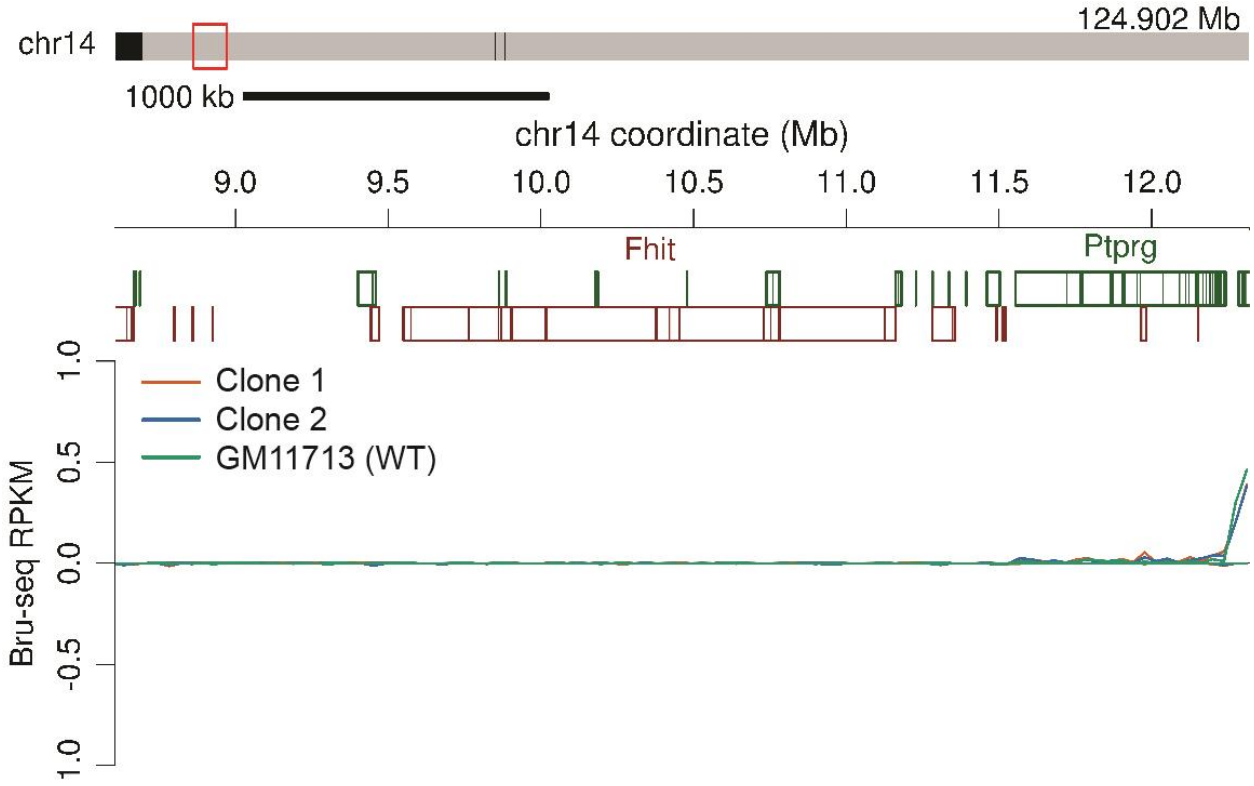
ddPCR Probe Location	Amplicon Sequence	Coordinates (hg38 or mm10)	Fluorophore
FHIT Exon 5	CATCCTAGATGATGTAGGCAAAGAAAGCCCACTG AGGTGAGCTCTATAGTTAAAGCAACTTTGTGTCCT AAGGGCATTTGCGAATTTGCATCGTGAGGGAATA AACCAAGTAGAAAGCAGCAG	chr3: 60,639,068-60,639.190	FAM
FHIT 5' End	CATCCTAGATGATGTAGGCAAAGAAAGCCCACTG AGGTGAGCTCTATAGTTAAAGCAACTTTGTGTCCT AAGGGCATTTGCGAATTTGCATCGTGAGGGAATA AACCAAGTAGAAAGCAGCAG	chr3: 60,248,902-60,249,024	FAM
FHIT 3' End	GGCCTTCTTTAACACTTTTTGTTTACAGGCCAAAC ATCACCCTCAGTCATTCCAGCAGTCTACTATGTGA ACTGAGACTAGGAGGGATAATATAGATTGCCAGAT AGATCTAGGCATGAGCTC	chr3: 59,751,402-59,515,24	FAM
NLGN1 Exon 4	GGCCTTCTTTAACACTTTTTGTTTACAGGCCAAAC ATCACCCTCAGTCATTCCAGCAGTCTACTATGTGA ACTGAGACTAGGAGGGATAATATAGATTGCCAGAT AGATCTAGGCATGAGCTC	chr3:173807701-173807823	FAM
Mouse Chromosome X (Reference)	GTTAGAATTAAAAAAGAGAAAGGCCCAAGCCATGT ATTCACATAGTTTGGGTACTGTTCTCTTCTCATTCC TGCACCAGAGGAAATAAAAGTGAAATTTGCATAGT TGATGATGTTTTATAAT	chrX: 43,500,369-43,500,491	HEX

Table 3.4. ddPCR assays that were used to measure deletion CNVs at *FHIT* and *NLGN1*.

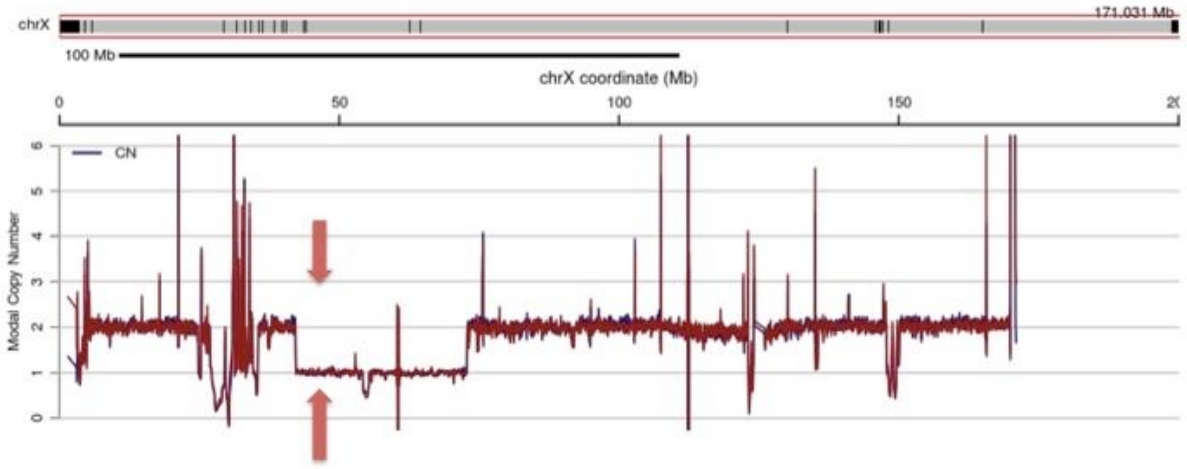
Primer Name	Sequence (5'-3')
FHIT Intron 2 F	CCATACTGACAGCTGCCTCT
FHIT Intron 2 R	TTTGTCGCCTCAACAGAACG
FHIT Exon 4 F	ACCTCTTCGGAGTCCTCAGT
FHIT Exon 4 R	GCCACACAGGACACTTCAGA
FHIT Intron 4-1 F	ACCTACCTCGGCCTCCCAAAGT
FHIT Intron 4-1 R	GCAGGAGGAGGCTCTGTTGAAACA
FHIT Intron 4-7 F	TGCCCCAGTGACTCTAGACA
FHIT Intron 4-7 R	CAGGAGAATGGCGTGAACCT
FHIT Intron 4-4 F	CGTGCCCAGCCCCAGTCTAA
FHIT Intron 4-4 R	TGGAGGCTTGCCCGGGTTTT
FHIT Intron 4-5 F	CCGGGCAAGCCTCCATGTCT
FHIT Intron 4-5 R	GGTCCCGCACTAGGCACTGT
FHIT Exon 5 F	GGAATTTGCTTTCTTTAAGGTGGTA
FHIT Exon 5 R	CTCAGCTATGGTAGTGAAAAGGTCA
FHIT Intron 5-3 F	TGGGTCAGTGCCTGCTCAGA
FHIT Intron 5-3 R	GCATGGGGGCTTGTGGTGTCA
FHIT Intron 5-4 F	TGGCCACAGCTTCCCGACTA
FHIT Intron 5-4 R	GCCCCCAGTGACTGCGGTTT
FHIT Intron 5-5 F	CGTGCCCGGCCATATCTATGCT
FHIT Intron 5-5 R	AAGGGTGCTGTTGGCTTGCT
FHIT Exon 6 F	GACCTGCGTCCTGATGAAGT
FHIT Exon 6 R	TGCATGGAAAAGGTGAGAGA
FHIT Exon 10 F	TCTCCCCAAGAGGAACTGAA
FHIT Exon 10 R	GCTGCCGAATAAGGAGACAG
FHIT Tel F	ATGGAAGCTCCTGCCTATGA
FHIT Tel R	AACTGAGGCACGCAGTTCTT
FHIT Tel2 F	TGTGGCTTGCCAGAAATTGC
FHIT Tel2 R	ACCCCCTCCACCATCTAGTC
FHIT Tel3 F	TGTGTGTGTGTGCACCTGTA
FHIT Tel3 R	TGGCGTTCTCAGCACTCTTT

Table 3.5. Primers that were used to detect deletion CNVs at *FHIT*.

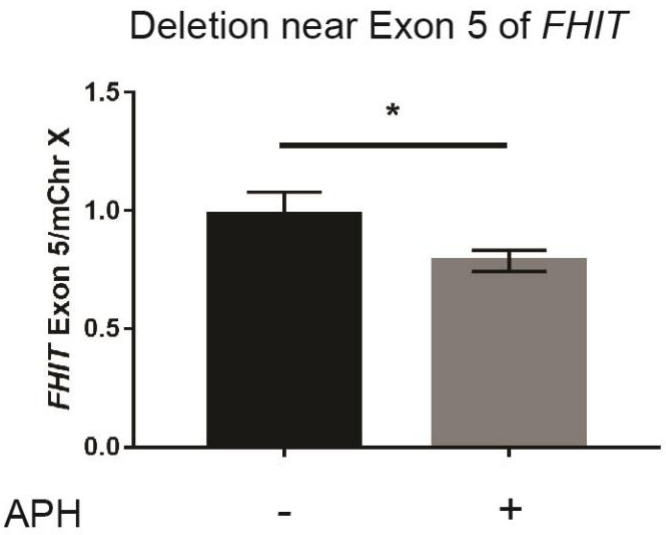
Supplemental Information



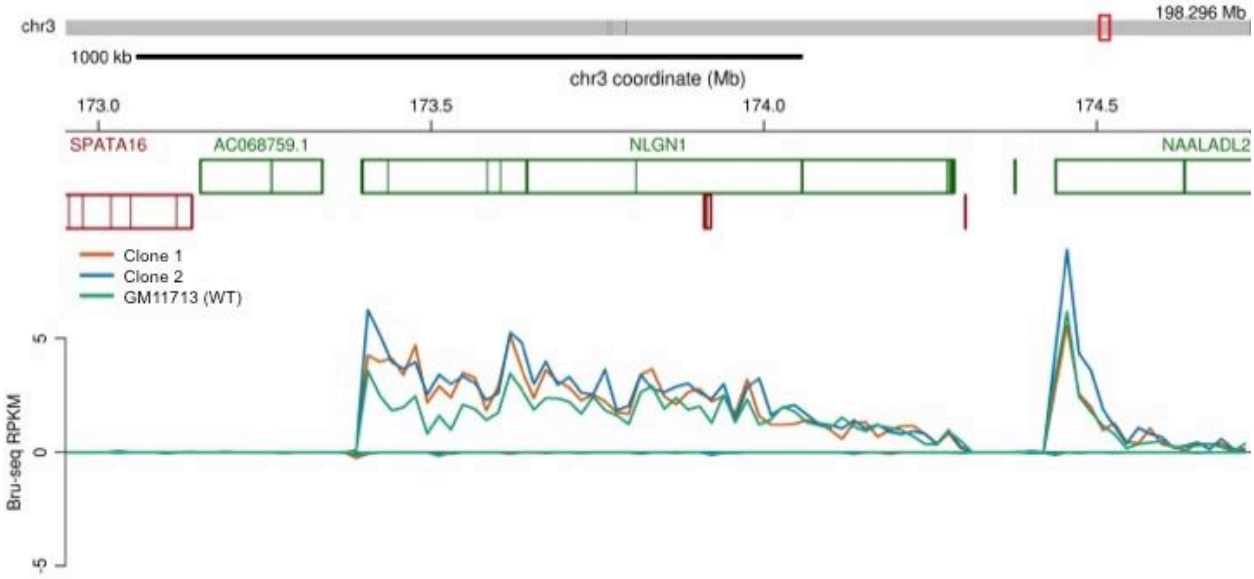
Supplemental Figure 3.1. Examination of the murine *Fhit* reveals that it is not expressed in GM11713, which additionally reveals that human *FHIT* Bru-seq reads do not align to mouse *Fhit* gene.



Supplemental Figure 3.2. Next-generation sequencing shows that certain regions on mouse chrX has copy number of one. Red arrows mark the location of the ddPCR reference probe.



Supplemental Figure 3.3. 0.6uM APH treatment induces deletion CNVs near exon 5 of *FHIT* in about 20% of GM11713 population. Statistical analysis was performed with a student's T-test. *:P≤0.05.



Supplemental Figure 3.4. *NLGN1* expression was unaffected by *FHIT* promoter knockout.

Notes and Acknowledgements

This chapter is an adaptation of a manuscript in preparation: Park SH, Ahmed S, Paulsen M, Ljungman M, Glover TW, and Wilson TE. Transcription knockdown reduces replication stress-induced CNV frequency at the same locus (tentative title).

I contributed in generating all the figures, tables, and supplemental figures in this chapter. Samreen Ahmed helped with tissue culture work and PCR assays leading to the results in Figure 3.5. Michelle Paulsen from the Ljungman lab prepared the samples for Bru-seq and Thomas Wilson aligned the Bru-seq reads to human and mouse reference genomes.

Chapter 4 — R-Loops Affect Both APH-Induced and Spontaneous CNVs

Summary

R-loops, a three-stranded nucleic acid structure composed of RNA/DNA hybrids and associated non-template ssDNA, that play a role in regulating gene expression, are also known to promote genome instability, such as DNA damage arising at sites of R-loop accumulation. Yet, how R-loops induce DNA damage at their physical location is unclear. A recent study has shown that R-loops contribute to CFS instability, in which an increased level of R-loops increased the instability and a reduced level of R-loops decreased it. In this Chapter, I investigate the potential role of R-loops in CNV formation and show that the APH-induced CNV frequency at CNV hotspots does not change with changing R-loop abundance, in contrast to published CFS results. However, changing R-loop abundance did significantly change the distribution of APH-induced CNV breakpoints. With R-loop accumulation, CNV breakpoints were more frequently located in GC-rich regions and near ends of transcripts, which are known signatures of R-loop-enriched regions, relative to wild-type cells. The opposite trend was observed with reduced level of R-loops. Taken together, these data suggest that R-loops influence the initial site of fork stalling/collapse and DSBs that lead to APH-induced CNVs. In addition, changing the R-loop abundance changed the overall frequency of spontaneous CNVs, which is consistent with previous observations that R-loops contribute to DNA damage that may lead to consequent genome rearrangements. This change in spontaneous CNV frequency provides further evidence that R-loops play an important role in genome instability.

Introduction

Chapter 2 introduced the TrDoFF model, which suggested that the transcription of the large underlying genes at CNV hotspots interferes with DNA replication to induce instability. Chapter 3 provided further evidence that transcription of large hotspot genes is crucial for CNV formation by demonstrating that knocking down transcription of a large gene reduced deletion CNVs at the region. The outstanding question still is — what is transcription doing to induce the instability?

One mechanistic possibility suggested by the literature is R-loop formation in large genes. R-loops are three-stranded nucleic acid structures with RNA/DNA hybrids and the associated ssDNA that form *in vivo* as natural intermediate structures, such as during *E. coli* plasmid replication, mitochondrial DNA replication, and immunoglobulin class switching (Aguilera and García-Muse 2012). Aside from these situations, R-loops were initially thought to be rare byproducts of transcription, but several recent studies have highlighted the role of R-loops in regulating gene expression. R-loop mapping using various methods, such as DRIP (DNA/RNA-immunoprecipitation)-seq, have shown that R-loops are preferentially located at GC-rich regions of gene promoters (i.e. CpG islands) and terminating regions (i.e. G-rich sequences near the 3'UTR) to aid transcription activation and termination (Ginno et al. 2013).

How R-loops regulate transcription is currently unknown, but recent studies have proposed that R-loops may induce epigenetic changes. A study by Skourti-Stathaki et al (2014) demonstrated that R-loops and H3K9me2 co-localize, suggesting a potential functional link between the two; however, it is not clear whether R-loops recruit the epigenetic signature or the other way around. Another study proposed that R-loop accumulation at the 5' UTR either

recruits H3K4me3 or protects the promoter from DNA methyltransferases to activate transcription (Ginno et al. 2012).

Despite their importance in regulating gene expression, R-loops, when aberrant, are detrimental. Interfering with enzymes that resolve R-loops (e.g. RNase H1, RNase H2, Senataxin, and Aquarius) can lead to R-loop accumulation and lethal effects. RNase H1 knockout causes embryonic lethality in mice due to failed mitochondrial DNA replication (Cerritelli et al. 2003), RNase H2 mutations in humans result in Aicardi-Goutieres syndrome (Crow et al. 2006), and senataxin (SETX) mutations are associated with early-onset amyotrophic lateral sclerosis (Y.-Z. Chen et al. 2004). Both RNase H2 and SETX mutations contribute to genome instability, such as DNA damage and translocations (Pizzi et al. 2015; Cohen et al. 2018; Lee-Kirsch, Wolf, and Günther 2014; Lavin, Yeo, and Becherel 2013).

Perhaps the harmful effects of R-loop accumulation are not surprising since it is a potential source of genome instability. The ssDNA associated with the RNA/DNA hybrid in the R-loop is vulnerable to DNA-modifying enzymes, including AID (Basu et al. 2011; Chiarle et al. 2011), which can lead to single-nucleotide mutations. The ssDNA is also susceptible to nucleases that cause single-stranded breaks (SSBs), such as XPG and XPF (Tian and Alt 2000). The SSBs could ultimately resolve into DSBs (Figure 4.1). The resulting DSBs are presumably processed via different DNA repair mechanisms to induce different manifestations of genome instability. The DSBs resulting from R-loop accumulation are thought to contribute to the translocation of *MYC* to the S region of Ig loci to cause sporadic Burkitt's lymphoma (Boerma et al. 2009; K. Yu et al. 2003; Duquette et al. 2005). It is also possible that the DSBs can act as a substrate for break-induced replication (BIR), a highly error-prone DNA replication that uses the 3' overhang to invade a homologous region to repair a DSB (Malkova and Ira 2013).

Furthermore, several DNA repair factors were found to be associated with R-loop accumulation, again suggesting that R-loops and genome instability are closely linked. BRCA1/2, important factors for homologous recombination, were found to increase R-loop accumulation when depleted (Bhatia et al. 2014; Hatchi et al. 2015). Similarly, FANCD2, an important factor in the Fanconi-Anemia pathway, was also found to be important in removing R-loops. FANCD2-deficient murine cells had more R-loop accumulation, γH2AX foci (phosphorylated H2AX histone protein, a biomarker for DNA double-stranded breaks), and DNA strand breaks than the control cells (García-Rubio et al. 2015). Replication stress and R-loop accumulation are also linked, as APH-treated cells had a higher global level of R-loops (Sollier et al. 2014).

R-loop accumulation also has been reported to affect CFS instability. Helmrich et al. demonstrated R-loops at CFSs affect the instability at CFSs. Upon *RNASEH1* knockdown, R-loops accumulated in the middle of three CFS genes (*FHIT*, *IMMP2L*, and *WWOX*), making the region more prone to instability in the knockdown cells compared to control cells. Upon *RNASEH1* overexpression, the authors saw the opposite. The R-loop accumulation can interfere with DNA replication because RNA/DNA hybrids are thermodynamically more stable than DNA/DNA hybrids (Roberts and Crothers 1992; Gyi et al. 1998; Chien and Davidson 1978), thus causing a roadblock for the replication fork and resulting in CFS instability. The implication of the proposed model is that large genes are more susceptible to the R-loop associated fork stalling because the RNAP has a longer mRNA tail, which is thought to produce a more stable R-loop than a RNAP at a shorter gene. Length is one of the factors that affect the stability of R-loops, in which shorter length results in lower stability (Lesnik and Freier 1995; Landgraf, Chen,

and Sigman 1995); however, whether the R-loops at the long genes are actually larger than those at shorter genes remains uninvestigated.

The relationship between R-loops and CNVs has not been studied yet. In this chapter, I present our data to support three claims. First, R-loop manipulation has no effect on the frequency of the APH-induced CNVs at the hotspots, contrary to the CFS data from Helmrich et al. Second, the APH-induced CNVs upon *RNASEH1* knockdown are preferentially located at either 5' or 3' ends of a transcript and GC-rich regions, and vice versa. Finally, R-loop manipulation alone affects the frequency of spontaneous CNVs. We present a model in which R-loop accumulation at both ends of the transcript shifts the distribution of APH-induced CNV breakpoints toward the ends, and resolving those R-loops does the opposite. In addition, the data in this Chapter provides evidence that over-resolution of R-loops reduces the number of spontaneous CNVs, and R-loop accumulation may change the structure of the spontaneous CNVs.

Results

Generating *RNASEH1* Knockdown and Overexpression Clones

For all the experiments in this Chapter, I worked with hTERT-immortalized normal human fibroblast cells line, HF1 for two main reasons. First, it is a normal human fibroblast cell line instead of a tumor cell line, which can exhibit high incidences of genome rearrangements during routine culturing and maintenance (Muff et al. 2015; Frattini et al. 2015; Orth et al. 1994; Kasai et al. 2016). Second, we have Bru-seq, CFS, and CNV data for this cell line from previous studies (Paulsen et al. 2013; Wilson et al. 2015), so we knew what HF1 cells' transcription

profile looked like, and we had an idea of where the CNV hotspots were located before we began these experiments.

RNase H1 is an enzyme that resolves R-loops by cleaving the phosphodiester bond in the RNA strand of the RNA/DNA hybrid. Helmrich et al. transiently knocked down or overexpressed the RNase H1 enzyme to study the effect of R-loops on CFS instability. We found that the HF1 cell line, however, has low transfection efficiency (<50%), implying that the majority of the cells will not knock down or overexpress the enzyme. Hence, we chose to virally transduce these cells with a Tet-On inducible construct called pTRIPZ for both the knockdown and overexpression.

The pTRIPZ construct (Supplemental Figure 4.1) is a Tet-On inducible system, in which integrants of this construct express Turbo RFP and shRNA-mir upon the addition of doxycycline. In the proprietary shRNA-mir design, the 22-nt dsRNA sequence contains 19-nt of the loop sequence from human microRNA-30 (miR-30). There are also 125-nt miR-30 flanking sequences surrounding the sequence. This design creates sites for Dicer and Drosha processing during the RNA interference process and increases knockdown efficiency and stability (Boden et al. 2004; Silva et al. 2005). The Tet-On inducible *RNASEH1* shRNA knockdown clones were generated via viral transduction as described in Materials and Methods.

The *RNASEH1* overexpression construct was generated by replacing the shRNA-mir and Turbo RFP with a *RNASEH1* cDNA sequence, using AgeI and MluI restriction enzymes (Supplemental Figure 4.1). The overexpression clones were generated also with viral transduction, as described in Materials and Methods.

Confirming RNASEH1 Knockdown and Overexpression in Transduced Clones

I first confirmed the validity of the Tet-On system by measuring RFP expression in the cells with scrambled control and knockdown shRNA constructs, since doxycycline drives shRNA expression via the linked TurboRFP gene (Supplemental Figure 4.2A and B). Fluorescence microscopy showed that adding 100ng/mL of doxycycline activates the inducible system after 48 hours (Supplemental Figure 4.2). I did not measure RFP expression in the cells with empty and overexpression constructs since in the overexpression construct, the Turbo RFP gene and shRNA-mir cassette was replaced with *RNASEH1* cDNA (Supplemental Figure 4.1 and Materials and Methods).

I then performed RT-qPCR to confirm that *RNASEH1* mRNA expression was reduced in the knockdown clones and increased in the overexpression clones. For both clones, the relative *RNASEH1* mRNA expression was measured after 48 hours of 100ng/uL doxycycline treatment. Figure 4.2A shows that the mRNA was successfully targeted, with ~75% knockdown compared to scrambled control. For the overexpression clone, the additional of doxycycline raised the mRNA level almost 75-fold compared to the empty control (Figure 4.2B). Next, I used Western Blot to confirm that the RNASEH1 protein was also predictably altered using the same experimental conditions. For both clones, the addition of doxycycline clearly increased or decreased RNASEH1 protein levels compared to the control (Figure 4.2C).

Given the lethal phenotype associated with *Rnaseh1*-knockout mice, I also conducted a cell growth assay with the *RNASEH1* knockdown mutants to see whether the knockdown slows cell growth. I did not perform the cell growth assay with the overexpression since *RNASEH1* overexpression is not associated with any cellular toxicity or diseases to date. The *RNASEH1* knockdown mutants showed no significant difference in growth compared to the scrambled

control, with and without doxycycline and APH treatment (Supplemental Figure 4.3). I also saw no apparent toxic effects of the doxycycline treatment for both scrambled and knockdown clones.

RNASEH1 Manipulation Changes Global Level of R-Loops

I performed S9.6 immunofluorescence (IF) assay to measure the effect of the *RNASEH1* knockdown and overexpression on the global level of R-loops. S9.6 antibody recognizes the RNA/DNA heteroduplex structure and is commonly used to detect R-loops (Boguslawski et al. 1986). Knockdown of *RNASEH1* is known to increase the global level of R-loops whereas *RNASEH1* overexpression reduces the level of R-loops, at least in yeast and in murine mitochondria (Wahba et al. 2011; Lima et al. 2016). The S9.6 immunofluorescence assay showed the normalized nuclear S9.6 signal was significantly increased upon *RNASEH1* knockdown (Figure 4.3A and B). The S9.6 signal was significantly reduced upon *in vitro* addition of RNase H for both clones, suggesting that the signal is specific to R-loops.

However, we saw the opposite of what we expected for the overexpression constructs. Instead of a reduced S9.6 signal, we detected a significant increase in signal (Figure 4.3C and D). Similar to the knockdown experiment, the S9.6 signal decreased with adding RNase H; however, the strong nucleolar signals persisted even with *in vitro* addition of RNase H.

RNASEH1 Manipulation Does Not Change APH-Induced CNV Structure, Overlap at Large Genes, or Size

Next, I used the *RNASEH1* knockdown and overexpression clones assess whether R-loops play a role in APH-induced CNV formation, at CNV hotspots or genome-wide. For all

clones, whether treated with APH or not, doxycycline was present throughout the entire treatment period since drug removal abolishes the *RNASEH1* mRNA knockdown or overexpression (Supplemental Figure 4.5A-D). The scrambled and empty control cells also were treated with doxycycline during the entire treatment period to control for any unanticipated effect doxycycline might have on CNVs.

The CNVs were induced, detected, and compiled as described in Materials and Methods. The final, complete lists of *de novo* CNVs from two replicates of the knockdown experiment and one replicate of the overexpression experiment are in the Appendix (Appendix 1-8). The two replicates of the knockdown experiment were set up slightly differently (Supplemental Figure 4.6A and B), but comparing the resultant CNVs from the two replicates revealed no overwhelming reason for us to believe that the two protocols yielded significantly different results (Supplemental Figures 4.7A-D). I included the separate tables for each replicate in the Appendix (Appendix 9-12). The following analyses were all performed using the combined list of CNVs from both knockdown experiments.

The first hypothesis I tested was that *RNASEH1* knockdown would induce more overall CNVs, and *RNASEH1* overexpression fewer CNVs, since in Helmrich et al. *RNASEH1* knockdown increased the number of breaks per cell and vice versa. I investigated whether there was a significant change in the number of APH-induced CNVs per clone between the control cells and knockdown or overexpression clones. There was no significant change in the number of induced CNVs per clone in either experiment (Figure 4.4A).

Because Helmrich et al. showed that *RNASEH1* manipulation changes CFS instability, I tested the hypothesis that *RNASEH1* knockdown would induce more CNVs at the CNV hotspots, with *RNASEH1* overexpression reducing CNVs. Because our CNV database on HF1 is limited, I

utilized the large (>500kb) genes as a surrogate for asking whether the CNVs were located in a hotspot or not (Wilson et al. 2015). I used the breakpoints of each CNV to determine whether a CNV was located in a transcribed large gene based on Bru-seq results from previous studies (Paulsen et al. 2013). I compared the number of CNVs that had at least one of its breakpoints in a large gene between the control and knockdown clones by constructing a 2x2 contingency table. There was no significant difference (Figure 4.4B). There was also no significant difference between control and overexpression clones (Figure 4.4C).

I then asked whether there were any qualitative differences between the APH-induced CNVs in the knockdown or overexpression clones compared to the control. Categorizing the CNVs into deletion or duplication and comparing the numbers in a 2x2 contingency table revealed that there was no significant difference between the control cells and the *RNASEH1* knockdown clones as well as control and overexpression clones (Figure 4.4D and E). Finally, I compared the sizes of the APH-induced CNVs between the control cells and the knockdown or overexpression clones. Again, there was no significant difference (Figure 4.4F). The median size of the APH-induced CNVs ranged from 86kb to 162kb, similar to the median size reported in our previous study (Arlt, Ozdemir, Birkeland, Wilson, et al. 2011).

RNASEH1 Manipulation Changes Distribution of APH-Induced CNV Breakpoints

CNV hotspots are known to be AT-rich compared to rest of the genome, whereas R-loops are known to accumulate in GC-rich regions (Ginno et al. 2013). Thus, if R-loops contribute to CNV formation, the CNV breakpoints, which are thought to be near the initial break sites, may be located in GC-rich regions. To test this hypothesis, I measured the average percentage of GC content in the 20kb window surrounding each breakpoint, 10kb on each side (Figure 4.5A) using

the UCSC genome browser's gc5Base. The 20kb window was roughly decided based on the fact that the average distance between probes in the Illumina SNP arrays (~2kb) and the largest reported resection length at a DSB site in human cells, which was about 3.5kb (Zhou et al. 2014), suggesting that the initial break site could be anywhere within a 5.5kb window on both side of the breakpoint, totaling to approximately 11kb. The GC content percentage at the breakpoints was significantly different between the control and the mutants. Consistent with the hypothesis, the breakpoints in knockdown clones were located at regions with higher percentage of GC content (Figure 4.5B). In converse, the breakpoints in the overexpression clones were located in lower percentage GC content (Figure 4.5B).

R-loops are also known to accumulate near TSSs (Transcription Start Sites) and TTSs (Transcription Terminator Sites) (Ginno et al. 2012; Ginno et al. 2013; Sanz et al. 2016; Skourti-Stathaki, Kamieniarz-Gdula, and Proudfoot 2014), so I hypothesized that the APH-induced CNV breakpoints are preferentially located near the ends of transcripts in knockdown clones. The previous studies have used gene annotations to mark TSSs and TTSs, but for my analyses I used the boundaries of transcription units (TUs) defined by Bru-seq on the HF1 cells (Wilson et al. 2015) rather than boundaries of gene annotations since previous studies also show that especially for the 3' end, the R-loop accumulation is not surrounding the end defined by gene annotation; rather, the cluster occurs past the 3' end, presumably because transcriptional activity does not cleanly stop at the end of the gene/transcript.

Significantly more APH-induced CNV breakpoints were located within the 20kb window surrounding the TU boundaries in the knockdown mutant compared to the control (Figure 4.5C). On the contrary, significantly fewer APH-induced CNV breakpoints were located near the TU boundaries in the overexpression mutants compared to in the control cells (Figure 4.5D).

RNASEH1 Manipulation Changes Spontaneous CNV Frequency and Size

I hypothesized that *RNASEH1* knockdown by itself would increase the CNV frequency, given the link between R-loop accumulation and genome instability (Aguilera and García-Muse 2012; Sollier and Cimprich 2015). However, the knockdown alone did not induce more spontaneous CNVs compared to the control (Figure 4.6A). There was also no significant difference between the control cells and the knockdown cells for most of the formerly mentioned analyses — structure, overlaps at large genes, breakpoint signature, and location at ends of transcripts (Figure 4.6B-E). However, the CNV size in the knockdown mutants was significantly smaller than that of the control cells (Figure 4.6F).

I also hypothesized that *RNASEH1* overexpression would decrease CNV frequency since previous studies have demonstrated that *RNASEH1* overexpression reverses the damaging effects from *RNASEH1* knockdown. The overexpression of *RNASEH1* significantly reduced the CNV frequency compared to the control, with only one detected CNV out of eight clones (Figure 4.7A). In other analyses, the CNV in overexpression mutant was not significantly different from the spontaneous CNV in the control cells (Figure 4.7B-E). The size comparison between the CNVs in control cells and overexpression mutant could not be performed because there was only one CNV in the overexpression mutant.

Discussion

The increasing amount of work linking R-loops with various forms of genome instability prompted us to investigate the role of R-loops in CNV formation. In particular, the study by Helmrich et al. showed that the level of R-loops directly correlated with CFS instability,

suggesting that R-loops may also impact CNV frequency at the hotspots. The Tet-On inducible *RNASEH1* knockdown and overexpression mutant cell lines were valuable tools for such question, as I was able to manipulate *RNASEH1* expression, and therefore R-loop abundance, in a controlled manner.

Although it is assumed that *RNASEH1* overexpression induces a global reduction in R-loops, the S9.6 IF assay suggested otherwise. It is possible that the *RNASEH1* overexpression showed an increase in S9.6 signal because of the antibody's cross-reactivity with other proteins. A recent study has demonstrated that S9.6 antibody recognizes a low level of various proteins even in the absence of the RNA/DNA hybrids in the sample (Cristini et al. 2018). It is not known whether the RNASEH1 protein is one of the proteins recognized by the S9.6 antibody, but the persistence of strong S9.6 signal in the nucleolus even after the RNase H treatment is consistent with the idea that the S9.6 antibody recognized the RNASEH1 protein aggregates in the nucleolus. Several studies have shown that endogenous and overexpressed RNASEH1 protein localizes in nucleoplasm, nucleolus, and mitochondria (Shen et al. 2017; Suzuki et al. 2010; H. Wu et al. 2013; Cerritelli et al. 2003). An alternative assay is necessary to confirm the reduction of R-loops in RNASEH1 overexpression mutants, such as assays using D5H6, another antibody that recognizes RNA: DNA hybrids (Shen et al. 2017; Molès et al. 2017), or gel mobility assay (K. Yu et al. 2006).

Unlike what we expected based on the published CFS results in Helmrich et al., neither form of *RNASEH1* manipulation significantly changed the CNV frequency at large, transcribed genes. There are a couple of potential explanations for this discrepancy. First, the Helmrich et al. study looked at only three CFSs (FRA3B, FRA7K, and FRA16D), so perhaps the trend they observed at these particular CFSs is not a generalized trend for all CFSs or CNV hotspots.

Second, although they share mechanistic links like late replication, CFSs and CNVs might be formed through distinct mechanisms. A recent review from our group posits that CFS gaps and breaks that represent uncondensed areas in the metaphase chromosome after MiDAS (Minocherhomji et al. 2015), while CNVs are thought to form through an incorrect fork restart or an incorrect DNA repair mechanism (Glover, Wilson, and Arlt 2017). What R-loops do to exacerbate instability at CFSs may not apply to CNV formation.

Finally, it is possible that R-loops still play the same role in CNV formation but *RNASEH1* manipulation alone was not enough to influence the CNV frequency. R-loops are resolved by various additional enzymes, such as RNase H2, Aquarius, and Senataxin, hence it is possible additional manipulation is necessary before we see an effect on CNV frequency at the hotspots, unlike at CFSs. A crucial future experiment would that would begin to address some of these conjectures would be to measure CFS instability in our mutant cells to confirm that we also observe a change in CFS instability upon *RNASEH1* knockdown and overexpression, as in Helmrich et al. In sum, it is crucial for us to measure CFS instability in the *RNASEH1* knockdown and overexpression clones.

Nonetheless, *RNASEH1* knockdown and overexpression were enough to significantly change the distribution of the APH-induced CNV breakpoints. The two tested factors, GC content percentage and the proximity to the ends of transcripts, were chosen specifically since they are associated with physical locations of R-loops (Ginno et al. 2013; Ginno et al. 2012; Sanz et al. 2016; Skourti-Stathaki, Kamieniarz-Gdula, and Proudfoot 2014). It is known that R-loops as associated with several forms of genome instability, so it is possible that the sites of R-loop accumulation become more prone to fork stalling and DSBs, which become the new sites that become extremely susceptible to template switching and misrepair. This model implies that

under replication stress, the initial events leading to the template switching or misrepair preferentially occur at the ends of the transcripts (where R-loops are located), possibly even at the large genes underlying CFSs and CNV hotspots, which goes against the model proposed in Helmrich et al.

The overexpression of *RNASEH1* significantly reduced the spontaneous CNV frequency, suggesting that R-loops also play a role in spontaneous CNV formation. The previously proposed model might help explain this result, in which R-loops induce DSBs at its sites of accumulation but the genome is protected from the breaks when they are resolved. However, the model is not consistent with the results from the knockdown of *RNASEH1*, in which the R-loop accumulation did not significantly increase the spontaneous CNV frequency. A technical explanation is that we need to analyze more CNVs to achieve significance. One possible biological explanation, as mentioned above, is that knocking down *RNASEH1* alone is not quite enough to see an effect, if any. More replicates are necessary to confirm this observation.

Materials and Methods

Cell lines

HF1 human immortalized fibroblast cells were cultured as described in Wilson et al (2015).

Generating and Confirming *RNASEH1* Knockdown and Overexpression Cell Lines

Preparing Constructs and Transduction

To generate Tet-On inducible *RNASEH1* knockdown HF1 cell line, the cells were virally transduced with one of the pTRIPZ human *RNASEH1* knockdown constructs from GE Life Sciences. We have transduced cells with one of the four *RNASEH1* knockdown constructs: V2THS_242087, V2THS_32362, V3THS_365743, and V3THS_365744. Among the four transduced populations, cells with V3THS_365744 construct were used for knockdown experiments. Cells were also transduced with a scrambled shRNA pTRIPZ plasmid for control.

To generate Tet-On inducible *RNASEH1* overexpression HF1 cell line, we first digested an empty pTRIPZ plasmid using AgeI and MluI (Supplemental Figure 4.1). The R*NASEH1* cDNA was amplified using Origene's pCMV6-AC-RNaseH1 plasmid (cat. no. SC319446) using primers with AgeI and MluI overhangs (see Table 4.1) using Phusion enzyme from New England BioLabs (cat. no. M0530). The PCR cycling condition consisted of an initial denaturation at 98°C for 30 seconds followed by 35 cycles of 98°C for 5 seconds, 58°C for 10 seconds, and 72°C for 30 seconds, final extension at 72°C for 5 minutes, and hold at 4°C.

The PCR product was completely digested with MluI-HF using overnight incubation and partially digested with AgeI-HF. The digestion reaction was cleaned up using Qiagen's PCR purification kit, and the eluted DNA brought up to 100uL total containing water and NEB's CutSmart Buffer (1x final concentration), and this tube was labeled "A." 20uL of "A" was added to tubes labeled "B," "C," and "D," and 10uL of "A" was added to tube "E." 1uL of AgeI was added to the remaining "A." Then 10uL of "A" was diluted in "B," 10uL of new "B" was diluted in "C," and so on. All five tubes were incubated at 37°C for exactly 2 minutes and were pooled into one tube. We used Qiagen's PCR purification kit to purify the digested products and ran

them out on an agarose gel to isolate the correctly sized product. Zymoclean Gel DNA Recovery Kit (cat. no. D4002) was used to extract the product from the gel.

The eluted PCR product was ligated to the AgeI and MluI-digested empty pTRIPZ plasmid using T4 DNA ligase from Thermo Fisher (cat. no. 15224017) following manufacturer's instructions. The HF1 cells were virally transduced with either an overexpression construct that was confirmed with Sanger sequencing or an empty control pTRIPZ construct. See Table 4.1 for primers that were used to confirm the integration of *RNASEH1* cDNA.

The University of Michigan Vector Core provided both empty and scrambled shRNA TRIPZ plasmid and helped with the transduction. After transduction, the integrants were selected using 0.5ug/mL Puromycin (Sigma cat no. P8833) for 10-14 days, until all non-integrants died, and plated at low dilution (200 cells per 10cm dish) to isolate individual clones.

Checking for RFP+ Cells

Both scrambled and *RNASEH1* knockdown cells were seeded in 6-well plates. Doxycycline-treated cells were treated with 100ng/uL of the drug for 48 hours, and they were supplied with fresh media and drug after 24 hours. The cells were visualized using Nikon Eclipse Ti, and images were acquired with the NIS-Elements Basic Research software.

RT-qPCR and Western Blot

The cells were treated with 100ng/mL doxycycline (Sigma Aldrich cat. no. D9891) for 48 hours before we assessed *RNASEH1* knockdown or overexpression. For RT-qPCR, the total RNA was isolated using Qiagen's RNeasy mini kit (cat. no. 74104). The total RNA was reverse transcribed to complementary DNA using the High-Capacity cDNA Reverse Transcription kit

from Thermo Fisher (cat. no. 4368814), and qPCR was performed using Qiagen's QuantiTect SYBR green PCR kit (cat. no. 204143) and Applied Biosystems 7500 Real-Time PCR system. The cycling conditions were 50°C for 10 minutes, then 95°C for 15 minutes, followed by 40 cycles of 15 seconds at 94°C, 30 seconds at 50°C, then 30 seconds at 72°C. *ACTB* was used as a control gene to calculate the ΔΔCt values. See Table 4.2 for primer sequences.

For Western Blot, proteins were detected using Abcam's anti-RNase H1 mouse monoclonal antibody (cat. no. ab56560), diluted 1:1000, or Neomarker's anti-alpha tubulin mouse monoclonal antibody (cat. no. MS-581-P), diluted 1:4000, and goat anti-mouse IgG (H+L) secondary antibody conjugated to HRP from Thermo Fisher (cat no. 62-6520), diluted 1:4000. The proteins were detected using either SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher cat. no. 34577) or Pierce ECL Western Blotting Substrate (Thermo Fisher cat. no. 32209).

Cell Count Assay

Both scrambled and *RNASEH1* knockdown cells were seeded in 6-well plates, 50,000 cells in each well. Each condition had two technical replicates. The day after seeding, I trypsinized the cells and used the Countess Automated Cell Counter and the Countess Cell Counting Chamber Slides (Thermo Fisher cat. no. C10228) to measure the number of cells for the Day 0 time point. These Day 0 trypsinized cells were not re-seeded. Except the untreated cells, the remaining cells were treated with 100ng/uL doxycycline and/or 0.4uM APH for 72 hours. The cells were supplied with fresh drug(s) every 24 hours. At the end of the 72-hour treatment period, the cells were trypsinized for the Day 3 time point. The number of cell divisions was calculated with the following formula:

$$\log 2\left(\frac{\text{Number of cells at Day 3}}{\text{Number of cells at Day 1}}\right)$$

S9.6 Immunofluorescence Assay

30,000 HF1 cells were seeded in 12-well plates on round cover slips (Fisher cat. no. 12-545-100). On the next day, they were treated with 100ng/mL doxycycline for 48 hours. Immediately after the 48-hour doxycycline treatment, the cells were fixed using 100% ice-cold methanol for 10 minutes on ice. The cells were then washed twice with PBS with 0.05% Tween-20 (v/v) (PBST) and permeabilized with ice-cold acetone for 1 minute on ice. After fixing and permeabilization, the negative control cells were treated with RNase H from New England Biolabs (cat. no. M0297) overnight at 37°C in a humid chamber to prevent drying. RNase H was diluted to 5U/mL in PBS, with RNase H buffer added to 1x final concentration.

The fixed cells were blocked with PBST supplemented with 5% normal goat serum (v/v) (Thermo Fisher cat. no. 10000C). The cells were then incubated with S9.6 antibody from Kerafast (cat no. ENH001), using 1:25 dilution, at 37°C for 1 hour in PBST with goat serum and washed three times with PBST. The cells were then incubated with the goat anti-mouse secondary antibody conjugated to Alexa Fluor 488 from Thermo Fisher (cat. no. A11001), using 1:100 dilution, at 37°C for 1 hour in PBST with goat serum. The coverslips were mounted onto the slides using SlowFade Gold Antifade Mountant with DAPI from Thermo Fisher (cat. no. S36938).

The images were obtained through Nikon's NIS-Elements Advanced Research software, using Nikon Eclipse Ti microscope. Using ImageJ (version 1.51), I used the DAPI signal to

create a mask of the nucleus. The nuclear S9.6 signal was determined by measuring the S9.6 signal in the nuclear mask. The DAPI signal in the nuclear mask also was used to normalize the nuclear S9.6 signals. The dot-plot graph was generated with GraphPad's Prism software.

Inducing CNVs

For the first replicate of the knockdown, the cells were treated with both 100ng/mL doxycycline and 0.4uM APH for 72 hours, providing fresh dose of drugs and media every 24 hours. Cells were then given 24 hours of recovery in absence of the drugs before they were plated at a low dilution (200-250 cells per 10cm dish) to isolate individual clones.

For the second repetition of knockdown, the cells were first treated with 100ng/mL doxycycline for 48 hours to induce knockdown of *RNASEH1*. Afterward, cells were treated with either 100ng/mL doxycycline alone or 100ng/mL doxycycline and 0.4uM APH for another 72 hours. Cells then recovered for 24 hours and were plated for clones.

The overexpression experiment was carried out like the second repetition of the knockdown, in which the cells were first treated with doxycycline for 48 hours and then treated with either doxycycline alone or doxycycline and 0.4uM APH for 72 hours.

Detecting *de novo* CNVs

Genomic DNA was extracted from individual clones using Qiagen's Blood and Tissue kit. *De novo* CNVs were detected using the HumanOmni2.5 BeadChip at the University of Michigan Sequencing Core. I used an in-house CNV detecting algorithm called MSVTools, which detects *de novo* CNVs by comparing the log-R ratio and B allele frequency deviations to a "reference" sample (the sample that resembles what the majority of samples looks like) and a

sample that most closely resembles what the sample in question looks like to avoid calling the same CNV in multiple samples. CNVs that were detected in multiple samples were not considered unique and were not included in the final list of *de novo* CNVs or subsequent analyses.

The final list of unique *de novo* CNVs was cross-checked with the results from Illumina's Genome Studio software's plugin called "cnvPartition 3.2.0," which detects CNVs and estimates the resulting copy number.

Analyzing the CNV Dataset

Comparing Total Number of CNVs

The p-values for the total numbers of CNVs among the different samples were calculated using Poisson statistics, using three different test methods (Krishnamoorthy and Thomson 2002; K. Gu et al. 2008). The largest p-values from the three tests were used to report statistical significance.

Comparing Deletions/Duplications Ratio

CNVs were considered deletion CNVs if the rearrangement resulted in a loss of a copy number from the reference, regardless of the actual resultant copy number. Similarly, CNVs that resulted in a gain of a copy number were considered duplication CNVs. The p-values for deletions/duplications ratios of different samples were calculated using a 2x2 contingency table with Fisher's Exact Test.

Deciding Whether Breakpoints Overlap Large Genes

Using the genomic coordinates for each breakpoint that were estimated by MSVTools algorithm, I categorized each CNV as either overlapping large (>500kb) genes or not. To be categorized as a CNV that overlaps a large gene, at least one of its breakpoints needed to lie within the annotated large gene, regardless of the relative location within the gene. Ensembl gene annotations were used to identify the start and the end of the genes. The p-values were calculated using a 2x2 contingency table with Fisher's Exact Test.

Calculating and Comparing Average Percentage of GC Content at Breakpoints

The average percentage of GC content at each breakpoint was calculated for the flanking 10kb regions surrounding the breakpoint using the UCSC Genome Browser's gc5Base, which calculates GC content percentage in each 5-base bin and calculates the mean percentage for the entire queried region. As stated in the Results section, the 20kb-window was decided based on the average distance (~2kb) between probes in the SNP array I used for this experiment and the largest reported resection length during a DSB processing, which is about 3.5kb in human cells (Zhou et al. 2014). CNVs smaller than 20kb were not used for this analysis. The dot-plot graph was generated using GraphPad Prism, and Mann-Whitney Test was used to calculate the p-values.

Comparing CNV Sizes

The CNV sizes were calculated based on the breakpoint coordinates from MSVTools. The dot-plot graph was generated using GraphPad Prism, and Mann-Whitney Test was used to calculate the p-values.

Deciding Whether Breakpoints Are Located at Ends of Transcripts

Like the former analyses, I used the breakpoint coordinates from MSVTools. The boundaries for each transcript unit were identified based on Bru-seq data for HF1 cells (Paulsen et al. 2013). Similar to calculating the GC content percentage, I queried the flanking 10kb regions surrounding each breakpoint to see whether the breakpoint is located within the 20kb-window of the transcript boundaries. Bru-seq transcription units were used to assess breakpoint locations rather than gene annotations.

Figures

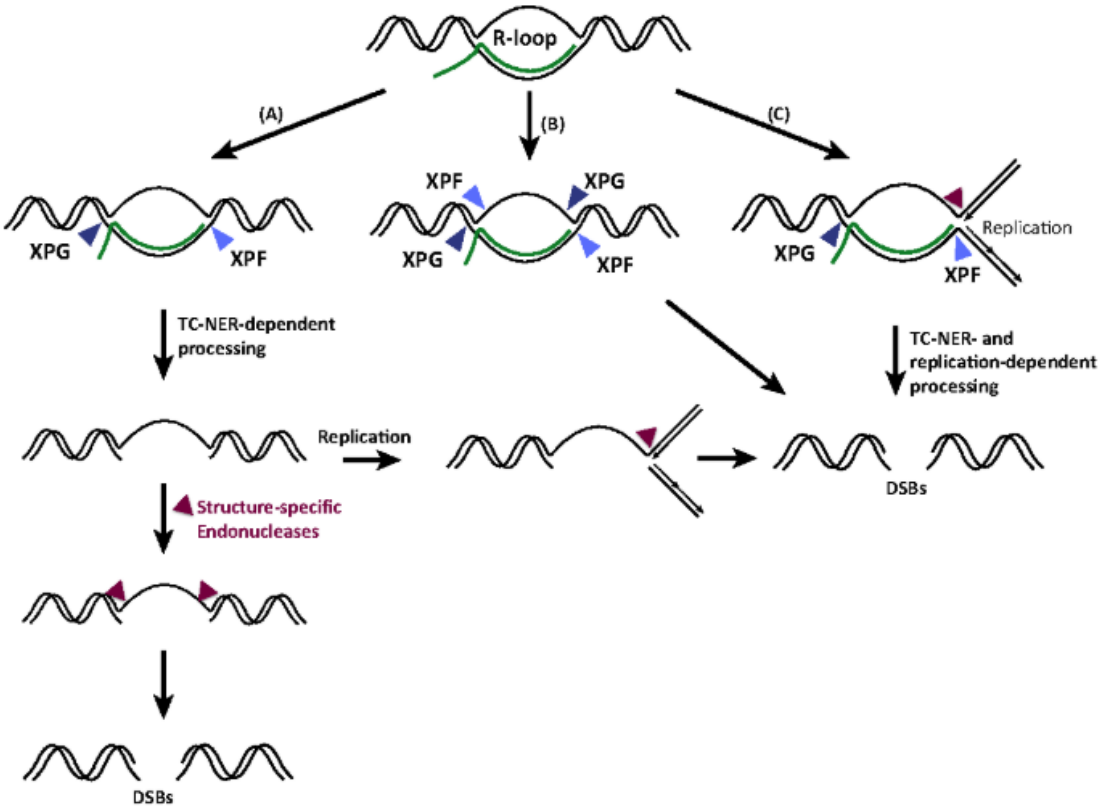


Figure 4.1. R-loop processing may lead to DSBs via various mechanisms. (A) The endonucleases XPF and XPG can excise the RNA/DNA hybrid of an R-loop. The single-stranded break (SSB) may resolve into a DSB either by structure-specific endonucleases via TC-NER (in purple) or by replication fork, which can also involve various structure-specific endonucleases. (B) XPF and XPG may also recognize both strands of an R-loop and generate a DSB. (C) Alternatively, both XPF/XPG and TC-NER endonucleases may act at the same time, where the TC-NER machinery may be recruited to R-loops during replication, and nucleases including XPF/XPG may generate a DSB. Adapted from Sollier and Cimprich (2015).

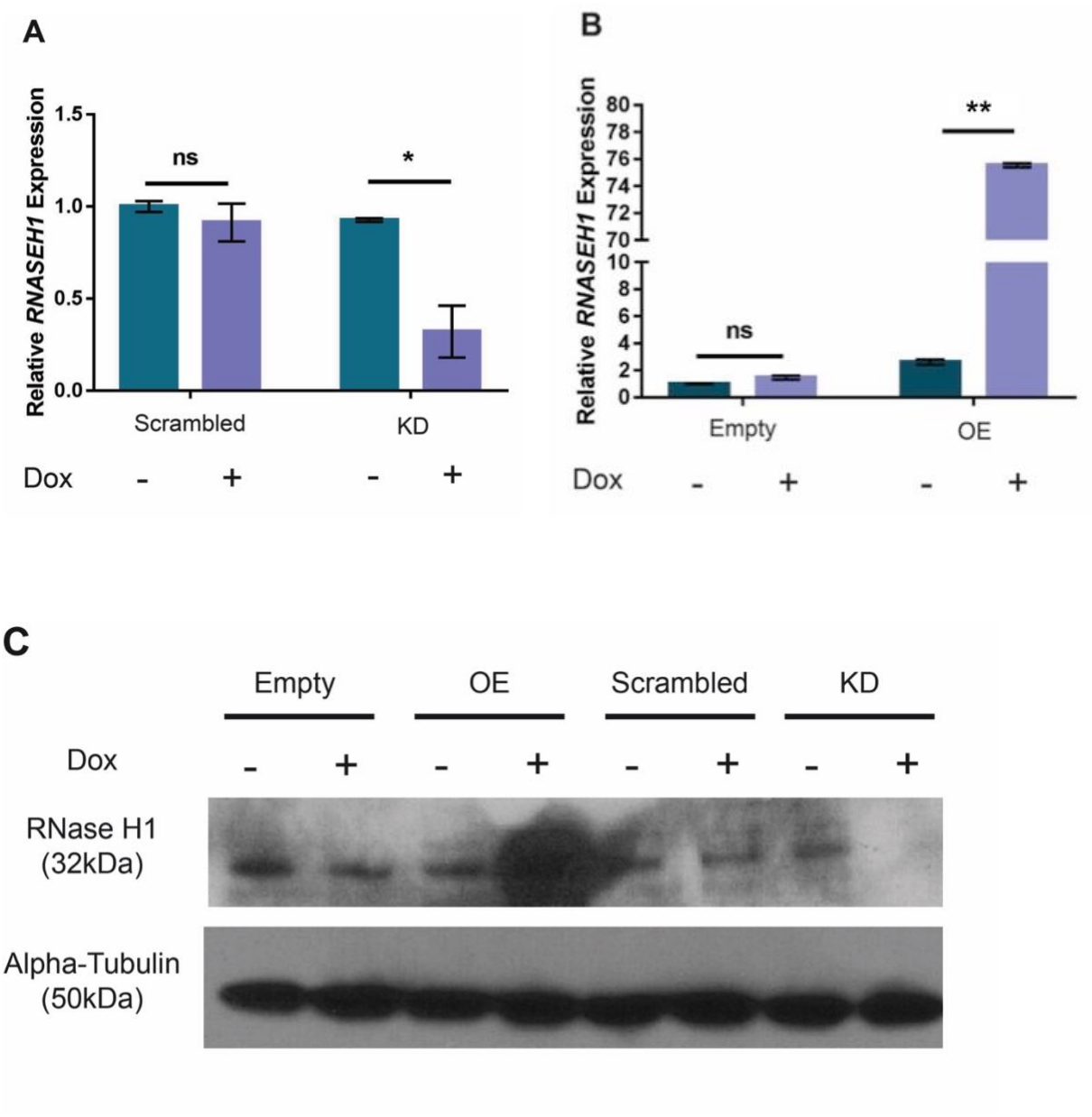
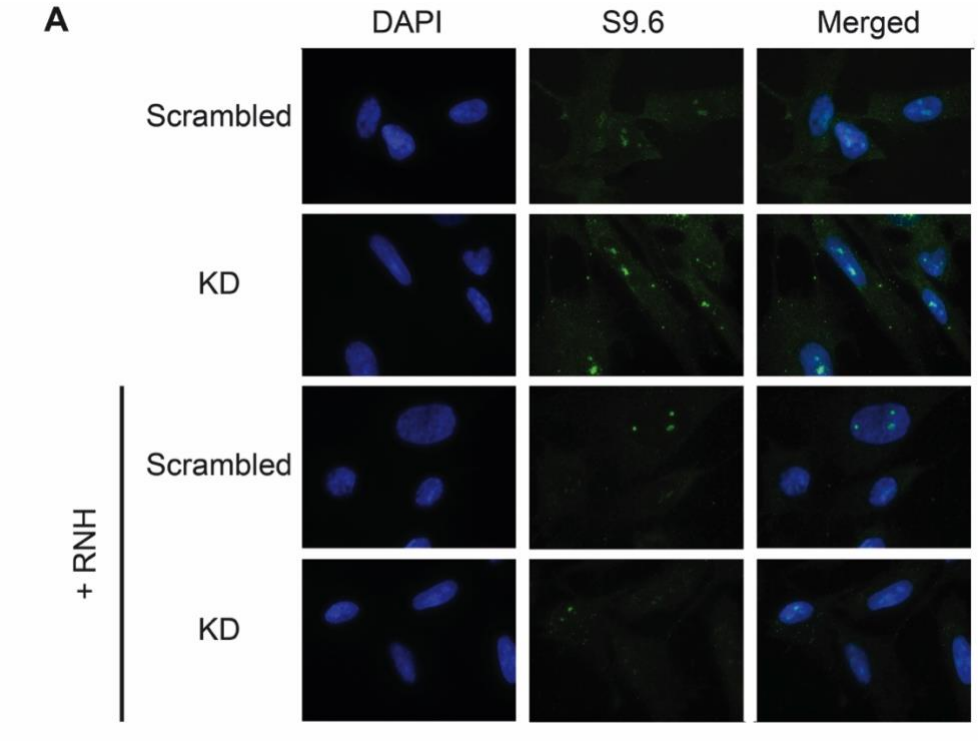
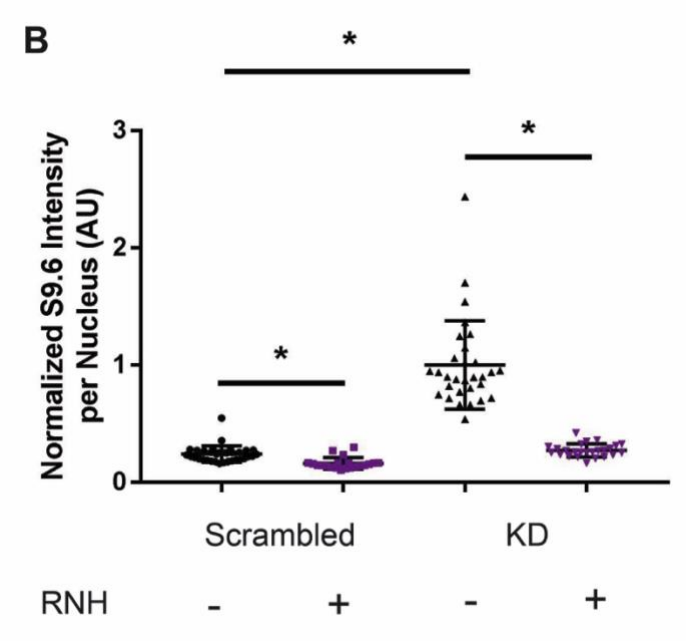


Figure 4.2. *RNASEH1* is knocked down or overexpressed in an inducible manner in cells that were transduced with the corresponding Tet-On constructs. (A and B) RT-qPCR shows that treating cells with 100ng/mL doxycycline for 48 hours reduces or increases *RNASEH1* mRNA levels in knockdown or overexpression mutants, respectively. Statistical analyses were performed using a student's T-test. ns: $P>0.05$, *: $P\leq0.05$, and **: $P\leq0.0001$. (C) Western blot shows that RNASEH1 protein levels are also reduced or increased upon doxycycline treatment in knockdown or overexpressed mutants, respectively.


A
DAPI
S9.6
Merged
Scrambled
KD
+ RNH
Scrambled
KD


B
*
*
*
Normalized S9.6 Intensity per Nucleus (AU)
3
2
1
0
Scrambled
KD
RNH
-
+
-
+

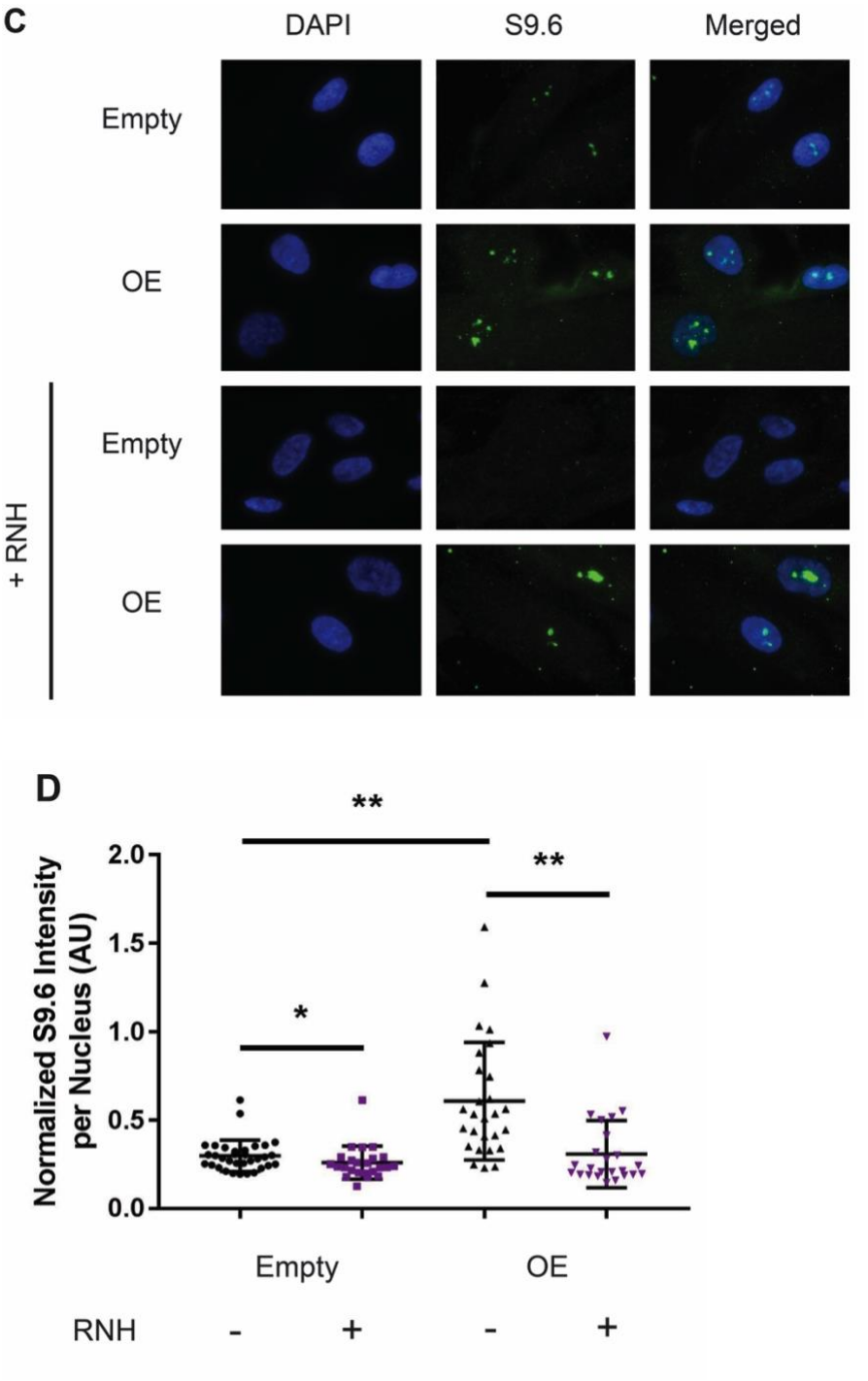


Figure 4.3. Manipulating *RNASEH1* changes global R-loop levels. (A) Representative images showing results of S9.6 immunofluorescence assay. Scr: Scrambled shRNA control and KD: *RNASEH1* knockdown. (B) Quantitation of S9.6 immunofluorescence assay. Nuclear S9.6 intensity was normalized by DAPI intensity. Statistical analyses were performed using Mann-Whitney test. *: P≤0.0001. (C) Representative images showing results of S9.6 immunofluorescence assay. Empty: Empty construct control and Over: *RNASEH1* overexpression. +RNH are control samples treated with RNase H after methanol fixation. (D) Quantitation of S9.6 immunofluorescence assay. Statistical analyses were performed using Mann-Whitney test. *: P≤0.05 and **: P≤0.0001.

A

Samples	Total Number of CNVs	Number of Clones	CNVs per Clone	P-Value
Scrambled + Dox + APH	39	19	2.1	0.36
KD + Dox + APH	32	17	1.9	
Empty + Dox + APH	26	8	3.3	0.065
OE + Dox + APH	16	8	2.0	

B

	In Large Genes	Not In Large Genes	Fraction in Large Genes	P-Value
Scrambled + Dox + APH	23	16	0.59	0.10
KD + Dox + APH	12	20	0.38	

C

	In Large Genes	Not In Large Genes	Fraction in Large Genes	P-Value
Empty + Dox + APH	11	15	0.42	0.75
OE + Dox + APH	8	8	0.5	

D

	Deletion CNVs	Duplication CNVs	Fraction of Deletions	P-Value
Scrambled + Dox + APH	24	15	0.62	0.62
KD + Dox + APH	22	10	0.69	

E

	Deletion CNVs	Duplication CNVs	Fraction of Deletions	P-Value
Empty + Dox + APH	16	10	0.62	1.0
OE + Dox + APH	10	6	0.63	

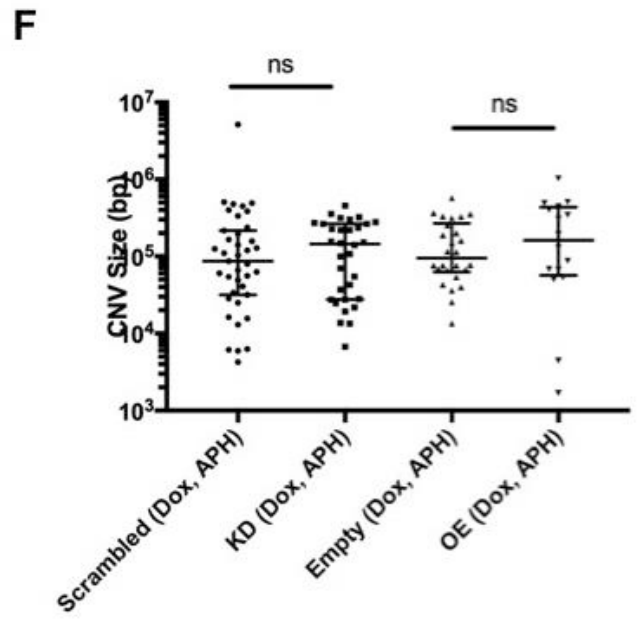
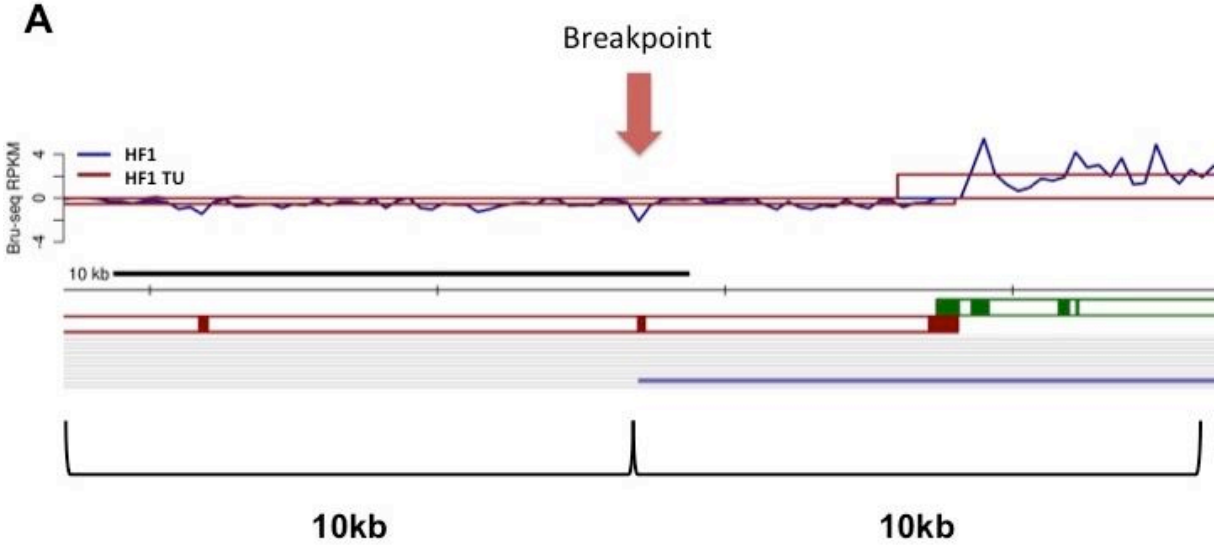
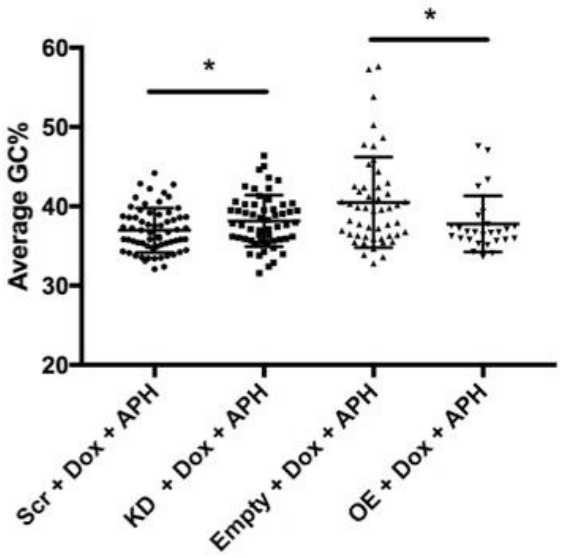


Figure 4.4. Neither *RNASEH1* knockdown nor overexpression changes overall CNV frequency (A), location in large (>500kb) genes (B and C), deletion to duplication ratio (D and E), and CNV size (F). Poisson statistics were performed for comparing the overall CNV frequency as described in Materials and Methods. Fisher's Exact Test for 2x2 contingency table was used for (B) through (E). Mann-Whitney test was performed for (F). ns: $P \geq 0.05$.

A



B



C

	Breakpoint at Ends	Breakpoint Not at Ends	Fraction at Ends	P-Value
Scrambled + Dox + APH	1	63	0.016	0.049
KD + Dox + APH	6	50	0.11	

D

	Breakpoint at Ends	Breakpoint Not at Ends	Fraction at Ends	P-Value
Empty + Dox + APH	8	42	0.16	0.045
OE + Dox + APH	0	28	0	

Figure 4.5. R-loops change the breakpoint locations in APH-induced CNVs. (A) shows an example of the breakpoint analysis. 20kb window surrounding each breakpoint (in red arrow) was analyzed for GC content and proximity to ends of TU. In the top panel, the blue trace represents raw Bru-seq data, and red box represents the predicted TU, with ends of the boxes representing TU boundaries. This particular breakpoint example would score "yes" for being at the end of the TU. (B) Average GC content % at breakpoints significantly differs between control cells and knockdown and overexpression clones. Statistical analysis was done using Mann-Whitney test. *: $P<0.05$. (C) The fraction of breakpoints within 10kb of ends of TUs for APH-induced CNVs was significantly higher in the knockdown clones. (D) The fraction of breakpoints at the ends of TU's was significantly lower in the overexpression clones.

A

Samples	Total Number of CNVs	Number of Clones	CNVs per Clone	P-Value
Scrambled + Dox	21	20	1.1	0.27
KD + Dox	30	24	1.3	
Empty + Dox	8	8	1.0	0.01
OE + Dox	1	8	0.13	

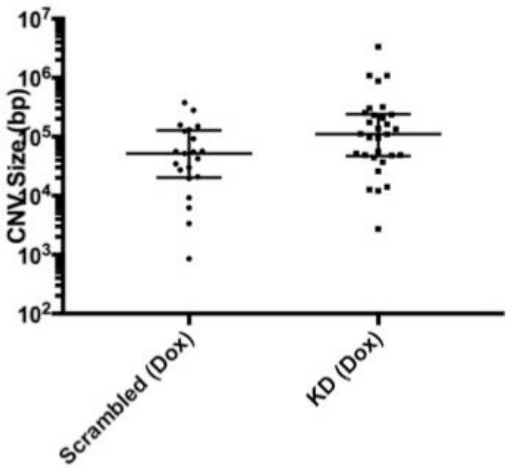
B

	In Large Genes	Not In Large Genes	Fraction in Large Genes	P-Value
Scrambled + Dox	4	17	0.19	0.43
KD + Dox	3	27	0.1	

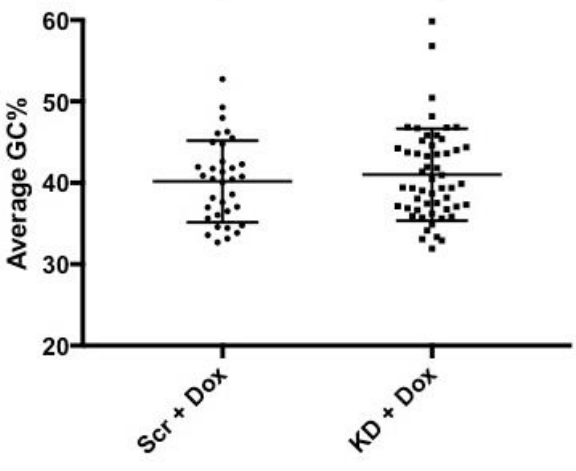
C

	Deletion CNVs	Duplication CNVs	Fraction of Deletions	P-Value
Scrambled + Dox	13	8	0.62	0.26
KD + Dox	13	17	0.43	

D



E



F

	Breakpoint at Ends	Breakpoint Not at Ends	Fraction at Ends	P-Value
Scrambled + Dox	7	27	0.21	0.37
KD + Dox	7	49	0.13	

Figure 4.6. R-loops change spontaneous CNV frequency and size. (A) The overall frequency for spontaneous CNVs is significantly lower for *RNASEH1* overexpression clones. On the other hand, knockdown clones did not have significantly more CNVs. The spontaneous CNVs in knockdown clones were not significantly deviant from the control cells in number of CNVs that overlap with large genes (B), deletion: duplication ratio (C), average GC content % (E), and number breakpoints at the ends of transcribed genes (F). The spontaneous CNVs in knockdown clones were significantly bigger than the ones in the control cells (median size: 52kb vs. 110kb). The p-values were calculated using Fisher's Exact Test for 2x2 contingency table for (A) – (C) and (F) and using Mann-Whitney test for (D) and (E). ns: $P \geq 0.05$ and *: $P < 0.05$.

Tables

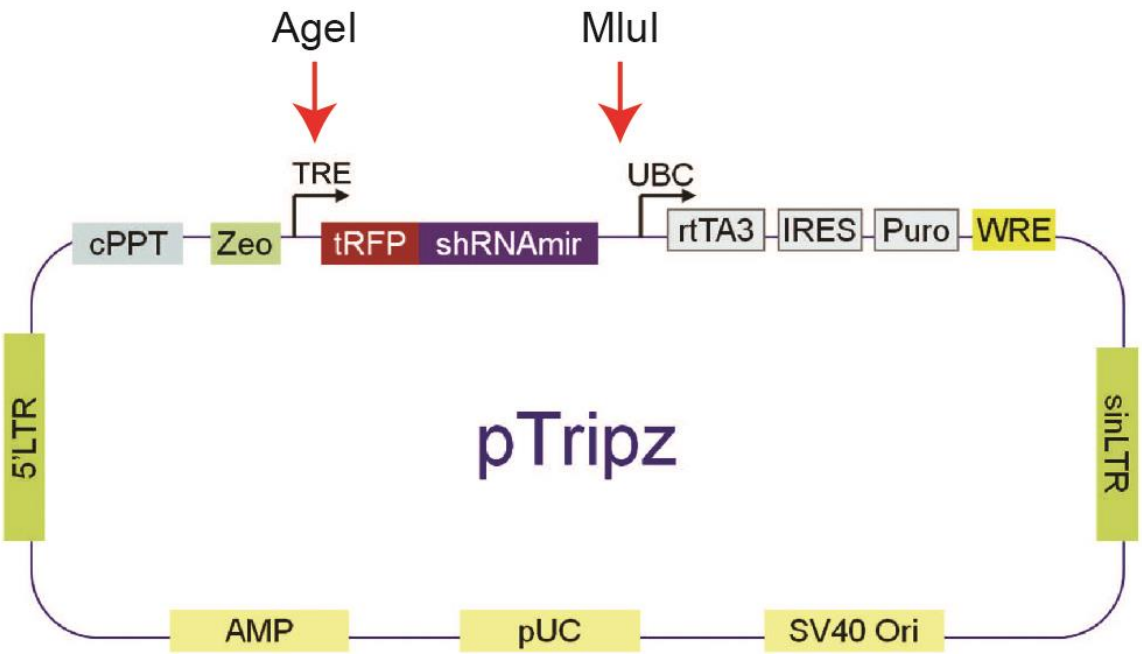
Primer Name	Sequence (5'-3')
RNase H1 cDNA F	GACTACCGGTCGTGTACGGTGGGAGGTCTA
RNase H1 cDNA R	GACTACGCGTCACTGGAGTGGCAACTTCCA
RNase H1 Confirmation 1 F	GTATGTCGAGGTAGGCGTGTA
RNase H1 Confirmation 2 F	CCAAAGAGCGGAAATTCATGCA
RNase H1 Confirmation R	TGCCCTAGAGTCACCCAAGT

Table 4.1. PCR primers that were used to amplify *RNASEH1* cDNA from the donor plasmid (pCMV6-AC-RNASEH1) and primers that were used to confirm that *RNASEH1* cDNA was successfully integrated into the Tet-On inducible expression plasmid (pTRIPZ).

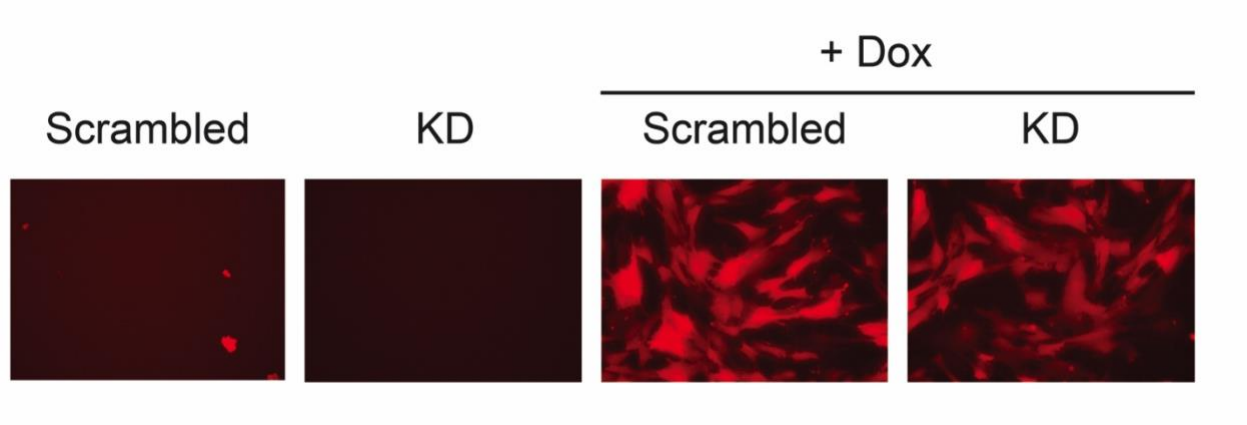
Primer Name	Sequence (5'-3')
RNase H1 Assay F	TGCAAGCCCGGAAGTTTCAG
RNase H1 Assay R	CTGCGCTTTCATGTCCATCTC
Human ACTB Assay F	ACCTTCTACAATGAGCTGCG
Human ACTB Assay R	CCTGGATAGCAACGTACATGG

Table 4.2. qPCR primers that were used to measure relative *RNASEH1* mRNA expression.

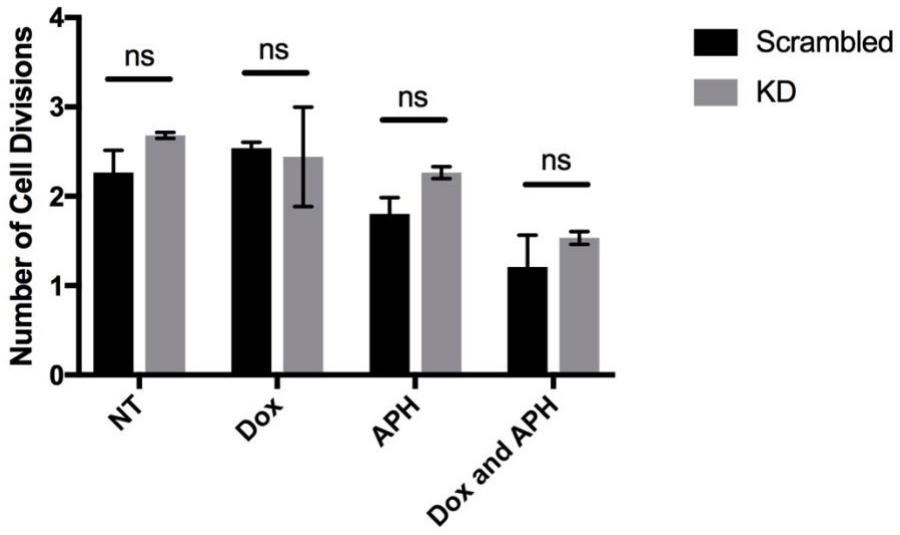
Supplemental Information



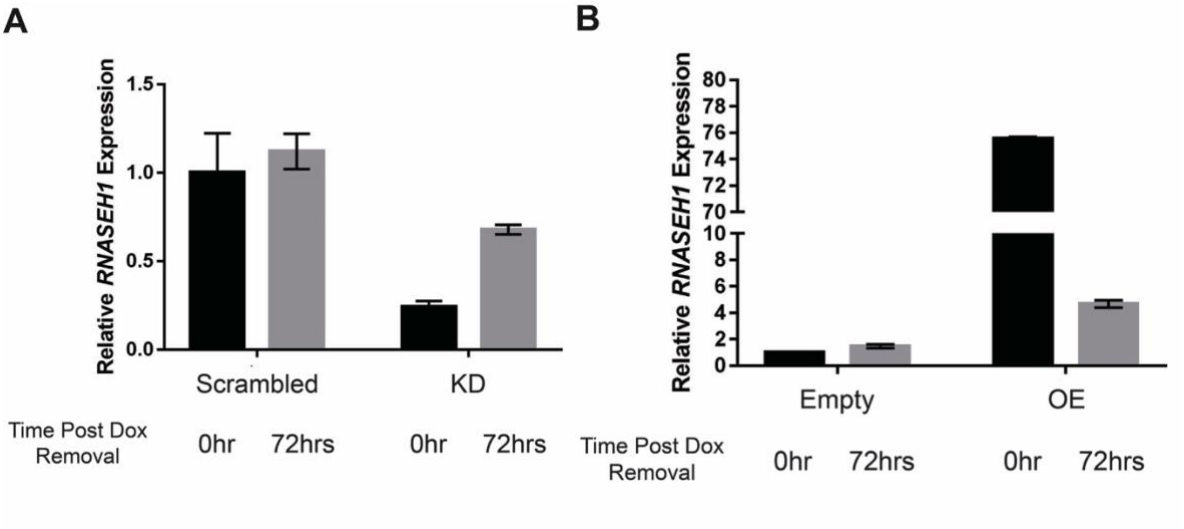
Supplemental Figure 4.1. A map of pTRIPZ. Red arrows mark AgeI and MluI recognition sites, which were used to replace the cassette with RNASEH1 cDNA in the overexpression construct. Adapted from Yu et al (2016).



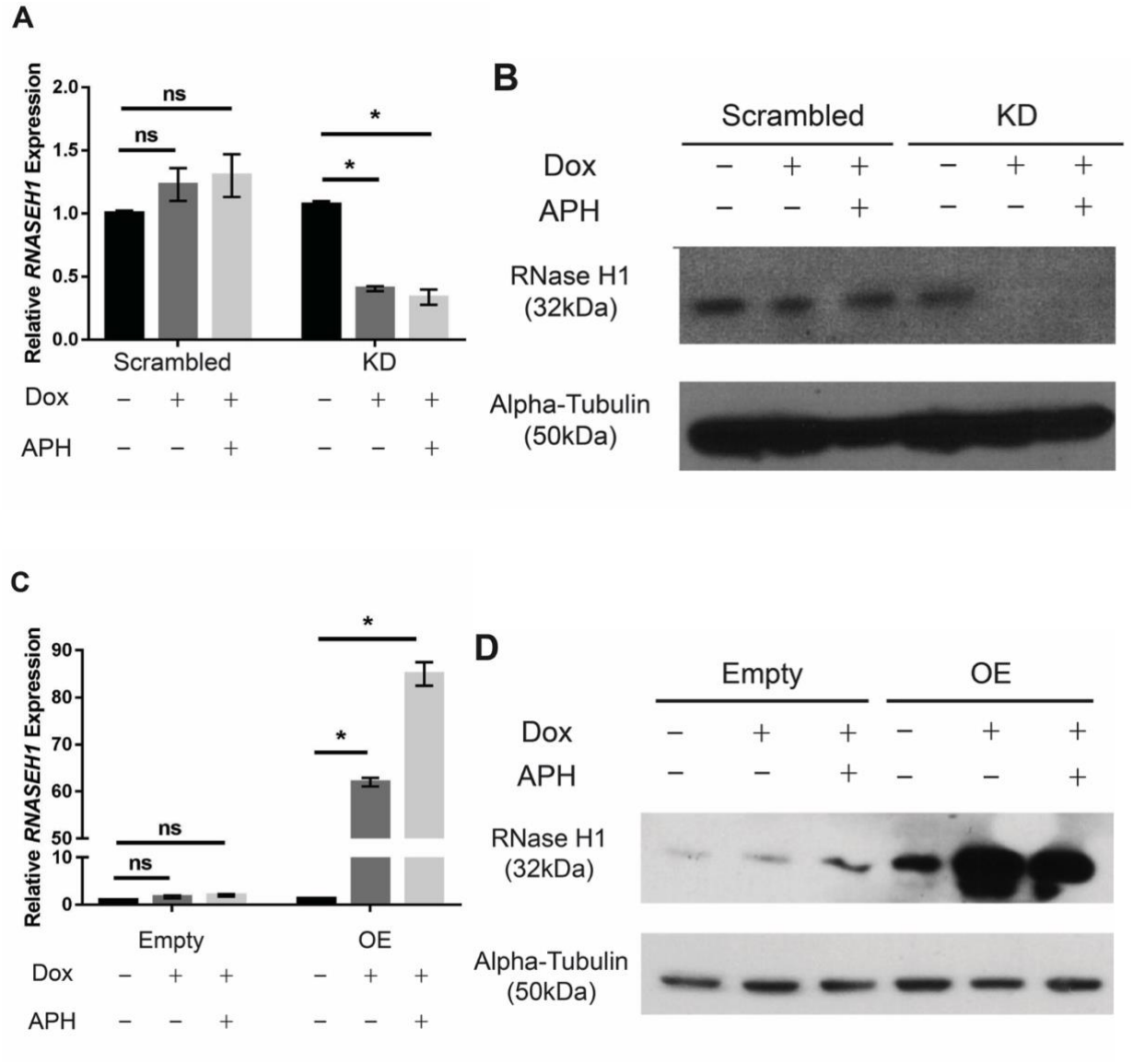
Supplemental Figure 4.2. A 48-hour treatment with 100ng/mL doxycycline turns on the Tet-On system in both scrambled control and *RNASEH1* knockdown cells as assessed by detection of Turbo-RFP by fluorescence microscopy.



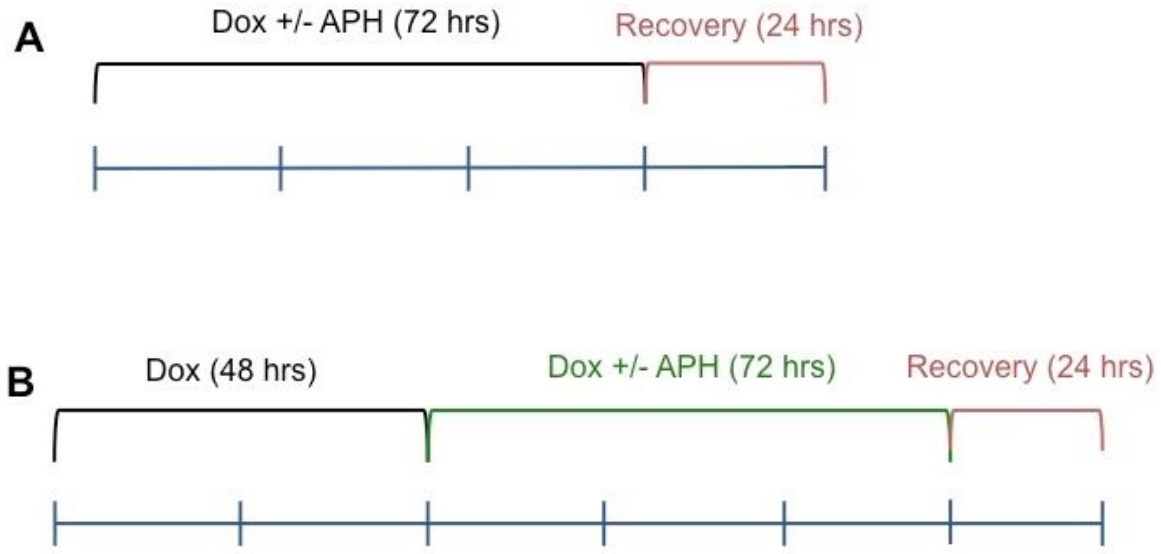
Supplemental Figure 4.3. There is no significant difference in cell growth between scrambled control and *RNASEH1* knockdown clones (KD) for all four treatment conditions after 72 hours. NT: untreated, Dox: doxycycline only, APH: APH only, and Dox and APH: Both doxycycline and APH.



Supplemental Figure 4.4. RT-qPCR shows that knockdown (KD) clones (A) and overexpression (OE) clones (B) lose the doxycycline-induced effect after 72 hours post drug removal.



Supplemental Figure 4.5. *RNASEH1* knockdown (A and B) and overexpression (C and D) persist through APH treatment. Statistical analyses were performed using a student's T-test. ns: P>0.05, *: P≤0.001, and **: P≤0.0001. Scrambled: Scrambled control, KD: Knockdown clone, Empty: Empty control, OE: Overexpression clone.



Supplemental Figure 4.6. The two replicates for *RNASEH1* knockdown CNV experiments were set up differently. In the first replicate (A), either doxycycline or doxycycline and APH were added to the cells for 72 hours, followed by 24 hours of recovery before the cells were plated for clone isolation, In the second replicate (B), doxycycline was added first for 48 hours, then doxycycline +/- APH were added for an additional 72 hours, followed by 24 hours of recovery. The overexpression experiment was set up following the timeline in (B).

A

Samples (Rep 1)	Deletion	Duplication	Total	Number of Clones	CNVs per Clone	Fraction of Deletions
Scrambled + Dox	7	7	14	12	1.2	0.50
Scrambled + Dox + APH	16	6	22	11	2.0	0.73
KD + Dox	9	14	23	16	1.4	0.39
KD + Dox + APH	8	6	14	9	1.6	0.57

B

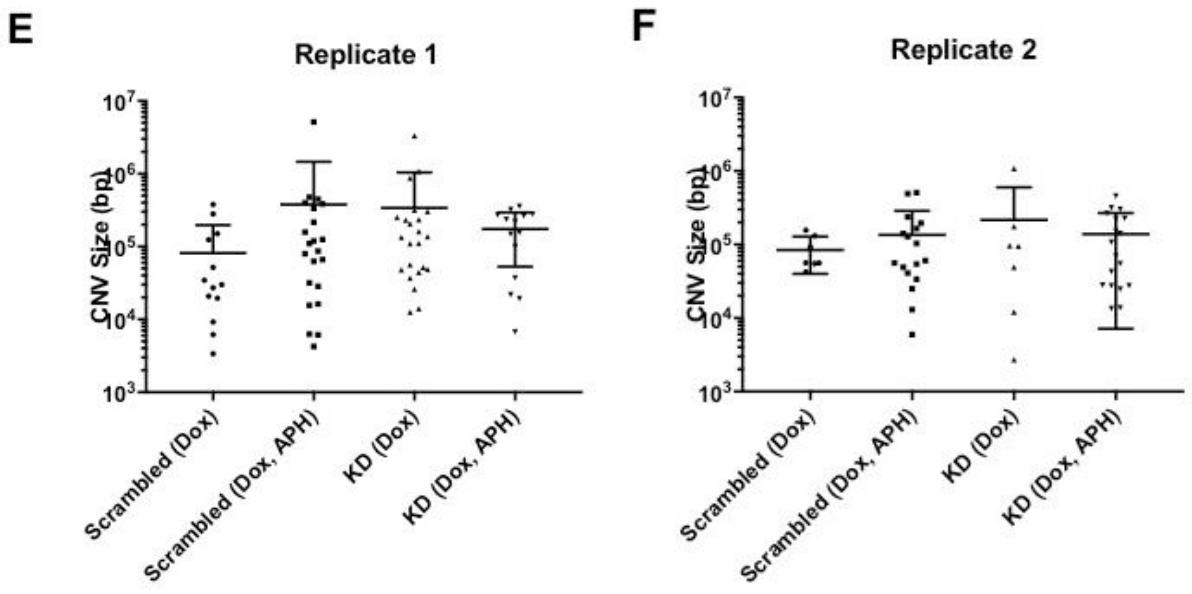
Samples (Rep 2)	Deletion	Duplication	Total	Number of Clones	CNVs per Clone	Fraction of Deletions
Scrambled + Dox	6	1	7	8	0.88	0.86
Scrambled + Dox + APH	8	9	17	8	2.1	0.47
KD + Dox	4	3	7	8	0.88	0.57
KD + Dox + APH	15	4	19	8	2.4	0.79

C

Samples (Rep 1)	In Large Genes	Not in Large Genes	Fraction in Large Genes
Scrambled + Dox	4	10	0.29
Scrambled + Dox + APH	14	8	0.64
KD + Dox	4	19	0.17
KD + Dox + APH	7	9	0.44

D

Samples (Rep 2)	In Large Genes	Not in Large Genes	Fraction in Large Genes
Scrambled + Dox	2	4	0.33
Scrambled + Dox + APH	10	7	0.59
KD + Dox	2	5	0.29
KD + Dox + APH	9	11	0.45



Supplemental Figure 4.7. There are no apparent differences between the two replicates of *RNASEH1* knockdown experiment in overall CNV frequency and deletion to duplication CNV ratio (A and B), the number of CNVs overlapping with large genes (C and D), and size (E and F).

Notes and Acknowledgements

I contributed to generating all the figures, tables, and supplemental figures in this chapter. Thomas Wilson built the bioinformatics pipeline for CNV detection, including writing the code for our in-house CNV detecting algorithm, MSVTools.

I would like to thank the University of Michigan vector core for their help with viral transduction and supplying the control pTRIPZ plasmids, the Markovitz lab and the Kalantry lab for their generosity in letting me use their microscope for the checking the RFP expression and the S9.6 immunofluorescence assay, respectively, and the University of Michigan sequencing core for their help with the SNP arrays.

Chapter 5 — Conclusions and Future Directions

Overview

Genome instability is a threat to an organism's genome integrity, as it is associated with a myriad of human diseases and conditions including cancer, neurodevelopmental disorders, and aging. Identifying the causes and mechanisms of genome instability is essential for understanding the aforementioned diseases and conditions, as well as understanding the genomic dynamics in a normal cell. Among numerous genome rearrangements resulting from genome instability, this dissertation focused on instability at CFSs and CNV hotspots. Recent reports, including ones from our group, have increasingly demonstrated the importance of replication-transcription conflicts in the instability at these sites, in which transcription interferes with replication and may cause fork stalling/collapse and DSBs (Wilson et al. 2015; Helmrich, Ballarino, and Tora 2011; Pomerantz and O'Donnell 2010b; M. Debatisse et al. 2012; Aguilera and García-Muse 2012; De Septenville et al. 2012).

Data from Chapter 2 demonstrate the predictability of the instability at the sites based on the transcription profile for the particular cell line used in the study. Based on the cell line's nascent transcription alone, I was able to predict where deletion CNVs would form in the human fibroblast cell line, HF1, which has no prior CNV data. I detected two deletion CNVs in *DAB1*, whose 1.5Mb isoform is expressed in the cell line. In addition, I demonstrated that transcription of the large genes is the determining factor for instability. We chose two human fibroblast cells (HF1 and 090) that have different expression patterns at two large genes to test whether that state

of transcription is correlated to CNV frequency at the site. I found that the presence of transcription determined whether the large genes were unstable or not. Combining the data from previous studies and the data in the Chapter, we propose the TrDoFF model, in which the late-replicating nature of CFSs and CNV hotspots combined with the transcription of the large underlying genes render the regions especially susceptible to double-fork failures, leading to incorrect fork restarts and misrepair, leading to CFS gaps/breaks and CNVs.

One caveat of the experiments and data presented in Chapter 2 is that I used two different cell lines to reach the conclusion about transcription's determining role in CNV formation. Data from Chapter 3 address this caveat. I deleted the promoter of *FHIT*, a hotspot gene, using the CRISPR-Cas9 technology to knockdown the transcriptional activity at the gene and showed that the transcription knockdown reduced CNV frequency at *FHIT* compared to the unedited, parental cell line. This data expands on the model presented in Chapter 2 and underscore the pivotal role of transcription in CNV formation. In addition, the transcription knockout mutants are valuable tools in exploring other replication-transcription conflict mechanisms at CFSs and CNV hotspots.

However, questions about the exact role of transcription still remain. I explored a potential mechanism of the replication-transcription conflict in Chapter 4, in which R-loops might create a barrier for DNA replication forks, thereby causing fork stalling and collapse. According to a previous study, R-loop accumulation led to higher CFS instability, especially rendering the middle of the site susceptible to transcription-replication collisions (Helmrich, Ballarino, and Tora 2011). I manipulated R-loop abundance by either knocking down or overexpressing *RNASEH1* and analyzed the effect on APH-induced CNVs at the hotspots. However, the data revealed a change in breakpoint signatures in APH-induced CNVs when R-

loop abundance was changed, suggesting that R-loops contribute to fork stalling/collapse in presence of replication stress, and this changes where the breaks occur.

In addition, reducing the level of R-loops resulted in a reduction in the spontaneous CNV frequency, which is consistent with previous reports linking R-loops with genome instability (Sollier and Cimprich 2015; Aguilera and García-Muse 2012). Previous studies have shown that R-loop accumulation increases genome instability by causing DNA damage and inducing DNA damage response/gross chromosomal rearrangements, all of which are reversed when the level of R-loops is reduced (Amon and Koshland 2016; García-Pichardo et al. 2017; Herrera-Moyano et al. 2014; Gómez-González et al. 2011; Bhatia et al. 2014; García-Rubio et al. 2015; Herrero, Marı, and Moreno 2013). Furthermore, the total number of APH-induced gaps and breaks on metaphase chromosomes was reduced with reduced R-loop abundance (Helmrich, Ballarino, and Tora 2011).

Interestingly, R-loop accumulation significantly reduced the size of the spontaneous CNVs, and showed more duplications than deletions (at a non-significant level). It is possible that the R-loop accumulation does not induce more spontaneous CNVs, but changes the mechanism of how the CNVs form. Future experiments should focus on further characterizing both APH-induced CNVs and spontaneous CNVs in the *RNASEH1* knockdown and overexpression mutants to improve on the interpretations and hypotheses proposed later in this Chapter.

Exploring Different Transcription-Associated Models for CNV Formation

As mentioned earlier, an increasing number of studies have focused on transcription as a potential source of genome instability, especially in the context of replication-transcription

conflict. Replication and transcription cannot be independent factors since both processes use the same DNA template, thus the cell must have ways to minimize the conflict between the two processes.

Cells have evolved to minimize the conflict in various ways, such as organizing the genome so the transcriptional direction is aligned with replication. Yet studies have shown that conflicts still occur and cause damage to the genome. Our data from Chapter 2 and 3 provide evidence that transcription of large genes is the determining factor for instability at CFSs and CNV hotspots. Based on previously proposed models (e.g. dormant origin and collision models) and properties of CFSs and CNV hotspots, we propose with our TrDoFF model that at CFSs and CNV hotspots, the middle of the regions are especially susceptible to instability because the transcription's interference with DNA replication origin activity is the most severe in the center. TrDoFF model specifically posits that transcription actually strips licensed origins off the DNA template (Snyder, Sapolsky, and Davis 1988; Lõoke et al. 2010) and reduces the number of available origins, thus reducing origin activity at the CFSs and CNV hotspots. The reduced origin activity is predicted to increase the risk of fork stalling/collapse, thereby leading to CNV formation through erroneous mechanisms like FoSTeS, MMBIR, and MMEJ.

The R-loop model for CFS instability gives a different role to transcription. The R-loop model draws upon the implications of replication-transcription collision model, in which the collision between replication and transcription machineries put the complexes at a stalemate. The collision cannot be resolved unless one of the machineries is ejected from the DNA strand and makes way for the remaining machinery to proceed. At least in *E. coli*, a study has shown that replication machineries can eject the transcription machineries during a head-on collision when the RNAP-DNA complex is not stable (Pomerantz and O'Donnell 2010a).

Conversely, if the RNAP-DNA complex was stable, the replication machinery might not be able to push through the collision and complete replication. RNA/DNA hybrids are more thermodynamically stable than dsDNA (Roberts and Crothers 1992; Chien and Davidson 1978; Gyi et al. 1998). If the R-loops at the large genes are longer stretches of the stable RNA/DNA hybrids, the replication machinery may not be able to remove the RNAP, and the replication fork may stall/collapse and lead to CNVs via mechanisms mentioned earlier.

Are the two models, TrDoFF and R-loops model, mutually exclusive? The TrDoFF model predicts that that the likelihood of double-fork failure depends on two factors, which are the distance forks need to travel (N) and the distance the fork has traveled before stalling (N_s). Since the likelihood is a function of N/N_s, factors that reduce the distance the fork has traveled increases the likelihood, such as aphidicolin and possibly R-loops; hence, R-loops are compatible with the assumptions of the TrDoFF model. However, what contributes to the increase in N or reduction in N_s remains unclear.

The two mechanisms that might increase the likelihood, origin scarcity and R-loops, mainly differ on the required timings of replication and transcription. The detail about when transcription ejects licensed origins in the TrDoFF model implies that transcription proceeds through the middle of the gene before replication does. On the other hand, the R-loop model suggests that transcription and replication are occurring at the same time since the two machineries have to collide. Assessing the relative timing of replication and transcription in high resolution in the same cell line will help elucidate which model is more accurate. The transcription timing assay in Helmrich et al. has limitations since the authors only analyzed two fractions of S phase cells (early and late S), which may not be enough resolution to deconstruct and analyze the relative timing of transcription and replication. In addition, using RT-qPCR to

measure transcription activity has caveats since even if the primers are designed to probe pre-mRNAs, the assay is still not directly measuring transcriptional activity.

However, the two models are not mutually exclusive; the main limitation for the proposed approaches are that they are population-based, and therefore we are making assumptions about events in individual cells based on the results. Do the replication timing data imply that *every* cell is replicating the middle of CNV hotspot late in the cell cycle? And in each cell, when does transcription occur and how does that compare to the same cell's replication timing? Unless we utilize single-cell technology, we will not be able to answer these questions accurately. In addition, Chapter 4 has demonstrated that R-loops possibly contribute to fork stalling and collapsing to induce CNVs, so perhaps what the future experiments also should focus on is refining the model so it is more consistent with the existing dataset.

The promoterless mutants generated in Chapter 3 could be valuable tools in testing the two models. According to TrDoFF, the absence of transcription will maintain the dormant origins in the middle of the hotspot gene, which will fire when replication stress-inducing drugs are added to the cells. The R-loop model, on the other hand, does not predict that the replication timing will change with replication stress in absence of transcription.

R-Loops and Genome Instability

Nonetheless, R-loops still may contribute to CNV formation even if the replication timing experiment reveals that dormant origins are fired in the promoterless clones. The data from Chapter 4 of this dissertation support the idea that R-loops contribute to CNVs possibly in other ways besides collision.

The exposed ssDNA in the R-loop structure is known to be susceptible to various enzymes, such as AID, XPF, and XPG, which induces SSBs that can ultimately become DSBs via various mechanisms and lead to the consequences of genome instability (Sollier and Cimprich 2015), including CFS gaps/breaks and CNVs. In Chapter 4, overexpressing *RNASEH1* (thus over-resolving R-loops throughout the genome) significantly reduced the number of spontaneous CNVs in the cells. This is consistent with prior studies that demonstrated that the reduction of R-loop abundance is associated with less DNA Damage Response and different genome rearrangements (Cerritelli et al. 2003; Gan et al. 2011; Huertas and Aguilera 2003; Hraiky, Raymond, and Drolet 2000; García-Rubio et al. 2015).

Curiously, the *RNASEH1* knockdown did not induce significantly more CNVs, but the CNVs were smaller in size and seem to have lower deletion to duplication ratio, possibly suggesting that although the knockdown may not induce more CNVs, but changes the profile in size and structure. The lower deletion to duplication ratio (albeit not reaching significance) was particularly striking since APH-induced CNVs tend to be biased toward deletion CNVs, and the bias even exists for spontaneous CNVs. Sequencing across the breakpoint junctions of these CNVs may help in identifying which other mechanism (if different) led to their formation. One possible mechanism that fits the characteristics of the CNVs is BIR, which is known to generate duplications in human cells at sites of DNA damage (Costantino et al. 2014).

Chapter 4 also revealed that the breakpoints of the APH-induced CNVs shift when *RNASEH1* is knocked down or overexpressed. In the knockdown cells, the CNV breakpoints tended to be in regions that share the same characteristics as regions enriched with R-loops, and the opposite was observed for the overexpression cells. These results imply that the initial fork stall and collapse and/or DSBs occur at regions that are enriched with R-loops, and this location

bias increases with more R-loops. The breakpoints of the spontaneous CNVs did not show this bias, at least for the knockdown clones (since no comparisons could be made for the overexpression clones due to low number of spontaneous clones), which suggests that this bias is specific for APH-induced CNVs.

What happened between the R-loop accumulation and the template switching or DSB generation is unclear. Are the DSBs result of breaks in the ssDNA in the R-loops, independent of replication forks? Or did the fork stall and collapse near the ends of the transcripts because of the stable R-loops, similar to the ideas proposed in the R-loops model? Similar to the spontaneous CNVs, breakpoint junction sequencing may help identify the mechanism leading up to the DSBs or template switching; for example, is there microhomology at the breakpoint junctions of these CNVs that suggests mechanisms like FoSTeS, MMBIR, or MMEJ?

Reflecting on Experimental Procedures in This Dissertation

In Chapter 3, I utilized the CRISPR-Cas9 technology to generate two promoterless mutants. I used two independently risen promoterless mutants for same experiments to address the potential off-target effects of CRISPR-Cas9, which have been a focus of several recent papers (Anderson et al. 2018; Fu et al. 2013; Cho et al. 2014; Kosicki, Tomberg, and Bradley 2018). Unlike typical uses of CRISPR (e.g. introducing small indels with NHEJ or replacing a sequence with HR), which utilizes one gRNA, I used two gRNAs to delete the promoter region. Although I needed to add extra steps to enrich the population with the deletion mutants, the CRISPR technology did provide a straightforward method to delete almost a 2-kb region in the genome, efficiently turning off transcription at *FHIT*. Similarly, this method could be useful in studying the functions of other genetic elements by deleting them and studying the phenotype.

I also used ddPCR in Chapter 3 for CNV analysis in a population of cells. The traditional method of analyzing CNVs requires analyzing individual clones using methods like SNP microarrays and PCR, which could be extremely time-consuming. Since we can now guess where CNV hotspots might be and know that the middle of the hotspots is the most unstable, we can use ddPCR to query the middle of the gene in a treated population without going through the time-consuming cloning, expanding, and harvesting steps. ddPCR could be extremely useful when wanting to quickly test what effect different conditions and manipulations have on CNV frequency at a hotspot gene.

The CNV analyses in Chapter 4 required me to work with several layers of uncertainty. Since the CNVs were detected using SNP arrays, the exact breakpoint junction could not be located. The Illumina SNP array has about 1.7 million probes for the human genome, which means that on average, each probe is about 2kb apart. Thus, the estimated breakpoints from MSVTools may be off by approximately 2kb even if the initial calling was accurate. In addition, DSBs are sometimes resected before they are repaired, and this resection in human cells could be as large as 3.5kb (Zhou et al. 2014).

Future Directions

Immediate future experiments should focus on filling the gaps in the presented data to make them more robust evidence to support our models. First, we can investigate the effect of transcription knockdown in CFS instability and replication timing using the promoterless mutants generated in Chapter 3. Second, we should look at CFS instability in the *RNASEH1* knockdown and overexpression clones to confirm that we are getting data that are consistent with the results presented in Helmrich et al. and that the difference in the data are due to biological

reasons, not technical reasons. Third, we must repeat some parts of the CNV experiments in Chapter 4. In particular, the overexpression experiment must be repeated since my conclusions in this dissertation were based on one replicate. I will also be interested to see the knockdown experiment repeated for the doxycycline-only treatment to test whether the *RNASEH1* knockdown alone actually induces more CNVs and/or change the CNV profile and size since the small number of CNVs in this discussion might have not been enough to make the true effect more obvious. Last, the CNV breakpoint junctions for *RNASEH1* knockdown and overexpression clones should be characterized further to see if the CNVs are similar to APH-induced CNVs in the control cells, and if not, start identifying the mechanisms that led to the formation of those CNVs.

The potential long-term future experiments are further characterizing the relationship between APH-induced CNVs in the *RNASEH1* knockdown clones to R-loops. In this dissertation, I analyzed the CNVs using two characteristics of R-loop-enriched regions, high GC-content and proximity to ends of transcripts. The obvious next step will be to actually measure the association between the CNVs and R-loops themselves, using techniques like DRIP-seq, and couple that dataset with other datasets that measure genome dynamics, such as Repli-seq (Hansen et al. 2010), to get more accurate picture of how R-loops are detrimental to genome integrity with respect to its relationship with DNA replication and DNA repair mechanisms. Furthermore, it will be relevant to observe what effects *RNASEH1* had on nascent transcription profile. The analyses performed in Chapter 4 assumed that neither *RNASEH1* knockdown nor overexpression changed the Bru-seq profile. A study by Lima et al. has demonstrated that *Rnaseh1* knockout in mice induced a change in the transcriptome in murine hepatocytes, particularly in genes relevant to mitochondrial function (Lima et al. 2016).

Finally, since the R-loop model and the TrDoFF model are not mutually exclusive, more studies should be done to explore the TrDoFF model further. One possible experiment is mapping licensed origins in different stages of the cell cycle and/or characterizing the relative replication and transcription timing.

Larger Implications of This Dissertation

Many of the hotspot genes are clinically relevant, as they are associated with numerous types of human cancer and neurodevelopmental disorders (Wilson et al. 2015). In many instances, the genomic rearrangements of the gene disrupt the gene's function, serving as the etiology for the onset and the progression of the disease. Other times, the instability at these genes renders the sites susceptible to viral integration. The distribution of CNVs observed in tumors and autism patients has a striking resemblance to the distribution of induced CNVs in our cell culture system (Glover, Wilson, and Arlt 2017; Beunders et al. 2013), suggesting that the findings using our cell culture model can be used to study what causes the instability in people. Additional investigations of the mechanism, such as uncovering the exact role of transcription at large gene in CFS/CNV hotspot instability, can help researchers predict the sites of extreme instability and learn to possibly prevent it in the context of diseases.

Although about 40% of all induced CNVs are located in the hotspots (Wilson et al. 2015), the majority of CNVs are not in the hotspots. Most of our efforts so far have been focused on studying the mechanisms that lead to CNVs at the hotspots because they are the sites of extreme instability. Although the same mechanism (e.g. TrDoFF model) may apply to non-hotspot CNVs formation, it is possible that these CNVs are formed with a different mechanism altogether. Our data from Chapter 4 suggest R-loops may also play a role in CNV formation, and

along with a number of other studies that have reported the relationship between R-loops and various aspects of genome instability (Aguilera and García-Muse 2012; Sollier and Cimprich 2015), they can perhaps help to explain how some of these non-hotspot CNVs occur. Furthermore, further investigations will also help identifying casual factors and potential preventive measures for R-loop-associated human diseases.

Appendix

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
Scr1	chr1	72,823,712	72,954,883	131171	G
Scr10	chr1	72,842,545	72,899,190	56645	L
Scr1-5	chr10	103,733,135	103,767,614	34479	L
Scr1-9	chr10	69,750,596	69,780,717	30121	G
Scr2-17	chr10	131,575,757	131,627,834	52077	L
Scr1-2	chr13	110,896,738	110,900,111	3373	L
Scr2-7	chr13	58,581,310	58,600,904	19594	G
Scr10-1	chr15	71,631,172	71,673,459	42287	L
Scr1	chr16	17,263,999	17,356,258	92259	L
Scr1-9	chr16	23,563,500	23,714,311	150811	G
Scr2-1	chr3	50,008,117	50,064,090	55973	L
Scr1-1	chr3	4,160,142	4,538,738	378596	G
Scr1-10	chr3	7,380,247	7,386,444	6197	G
Scr1-2	chr4	372,995	497,138	124143	L
Scr6	chr5	60,507,328	60,664,356	157028	L
Scr1-10	chr5	64,467,769	64,749,379	281610	L
Scr2-13	chr6	26,405,263	26,414,488	9225	G
Scr2-13	chr6	26,465,383	26,466,234	851	G
Scr2-7	chr7	110,459,283	110,486,408	27125	L
Scr6	chr8	141,004,454	141,059,415	54961	L
Scr1-10	chr9	9,652,477	9,673,320	20843	L

Table A.1. List of unique, *de novo* CNVs in Scrambled control + doxycycline clones. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication).

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
Scr14A	chr1	72,221,729	72,362,338	140609	L
Scr17A	chr1	72,367,960	72,424,117	56157	L
Scr18A	chr1	72,508,295	72,703,860	195565	G
Scr1-21A	chr1	72,108,797	72,234,070	125273	L
Scr1-21A	chr1	72,296,122	72,512,864	216742	L
scr1-11A	chr1	51,866,768	51,953,306	86538	L
scr1-11A	chr1	98,203,819	98,539,032	335213	L
Scr2-18A	chr1	72,336,912	77,493,510	5156598	L
Scr10A	chr10	53,871,454	54,377,691	506237	G
Scr17A	chr10	53,326,699	53,455,408	128709	L
Scr10A	chr11	114,449,607	114,474,610	25003	G
Scr2-18A	chr12	99,855,731	100,255,744	400013	L
Scr2-10A	chr14	56,078,738	56,110,539	31801	L
Scr10-1	chr15	71,631,172	71,673,459	42287	L
Scr1-16A	chr15	61,074,035	61,080,335	6300	L
Scr10A	chr16	72,327,024	72,381,460	54436	G
Scr1-21A	chr16	72,423,462	72,439,749	16287	L
Scr2-18A	chr16	78,617,930	78,680,728	62798	L
Scr2-18A	chr16	83,088,211	83,198,489	110278	G
Scr2-21A	chr16	78,634,831	78,792,498	157667	L
Scr1-16A	chr18	68,719,952	68,726,105	6153	G
Scr7-1A	chr19	1,912,948	1,926,027	13079	G
Scr14A	chr2	36,738,629	36,842,336	103707	G
Scr18A	chr2	214,702,401	214,762,709	60308	G
Scr1-16A	chr2	136,157,318	136,236,787	79469	L
Scr1-2A	chr2	206,472,559	206,858,172	385613	G
Scr1-16A	chr3	60,604,488	60,620,114	15626	L
Scr2-18A	chr3	67,529,226	67,648,908	119682	G
Scr2-20A	chr3	60,243,577	60,693,253	449676	L
Scr2-21A	chr3	59,999,964	60,004,228	4264	L
Scr10A	chr4	32,447,037	32,452,973	5936	G
Scr12A	chr5	145,503,563	145,669,064	165501	L
Scr10A	chr7	77,797,289	78,289,970	492681	L
Scr10A	chr7	78,315,216	78,348,879	33663	L
Scr18A	chr7	110,941,859	110,982,750	40891	L
Scr1-16A	chr7	77,866,710	78,345,207	478497	L
Scr17A	chr8	50,374,788	50,424,019	49231	G
Scr18A	chr8	4,170,496	4,407,443	236947	L
Scr1-16A	chrX	36,308,773	36,374,854	66081	G
Scr2-10A	chrX	36,509,237	36,537,763	28526	G

Table A.2. List of unique, *de novo* CNVs in Scrambled control + doxycycline and APH clones. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication).

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
sh2a_dox	chr1	236,702,209	236,714,291	12082	G
sh1-4	chr1	172,252,304	172,487,743	235439	G
sh2-6_2	chr1	2,870,849	2,980,578	109729	G
sh2-6_2	chr10	126,642,073	126,656,168	14095	L
sh2a_dox	chr11	35,215,725	35,218,445	2720	L
sh1-15	chr11	121,916,950	121,973,583	56633	G
sh1-4	chr12	88,837,978	88,949,124	111146	G
sh1-21	chr13	110,877,045	111,014,716	137671	L
sh1-4	chr13	110,811,450	111,043,309	231859	L
sh1-4	chr13	111,044,407	111,081,408	37001	L
sh1-4	chr13	111,082,057	111,094,728	12671	L
sh2-1	chr13	28,186,344	28,235,239	48895	L
sh2-6_2	chr13	73,876,054	73,920,464	44410	G
sh1-1	chr15	101,155,565	101,458,429	302864	L
sh2-9_2	chr15	85,207,361	85,341,859	134498	L
sh1-19	chr16	12,040,510	12,092,575	52065	G
sh1-21	chr16	7,760,959	11,085,603	3324644	G
sh2-8_2	chr2	234,889,976	235,096,984	207008	G
sh12a	chr20	9,153,061	9,202,117	49056	G
sh7a_dox	chr21	40,075,872	40,251,856	175984	G
sh1-21	chr3	87,665,971	88,752,258	1086287	G
sh1-4	chr3	105,142,722	105,190,870	48148	G
sh1-9	chr4	2,791,990	3,110,509	318519	L
sh2-14	chr4	32,524,423	32,550,474	26051	G
sh11a	chr5	103,536,192	104,616,894	1080702	L
sh11a	chr7	78,170,842	78,267,691	96849	L
sh11a	chr7	134,493,343	134,588,145	94802	L
sh1-4	chr7	137,667,969	137,921,040	253071	G
sh1-6	chr7	82,100,279	82,262,193	161914	G
sh1-4	chrX	88,852,133	89,729,690	877557	G

Table A.3. List of unique, *de novo* CNVs in *RNASEH1* knockdown + doxycycline clones. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication).

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
sh2A_APH	chr1	245,448,885	245,904,391	455506	L
sh6A	chr1	85,980,226	86,004,787	24561	L
sh6A	chr1	85,868,701	86,021,169	152468	L
sh1-6A	chr1	71,633,718	71,868,535	234817	G
sh1-6A	chr1	72,315,819	72,637,154	321335	L
sh2A_APH	chr10	129,051,516	129,366,742	315226	G
sh2a_dox	chr11	35,215,725	35,218,445	2720	L
sh2-5A	chr11	129794949	129950592	155643	G
sh2A_APH	chr12	99,915,661	100,056,529	140868	L
sh1-13A	chr13	71,905,357	72,125,641	220284	G
sh2A_APH	chr15	60,681,114	60,694,820	13706	L
sh2-6A	chr15	46696827	46969077	272250	G
sh5A_APH	chr16	17,204,226	17,312,099	107873	L
sh6A	chr16	83,735,129	83,805,328	70199	G
sh1-6A	chr17	26,410,260	26,510,217	99957	L
sh3A	chr2	191,745,388	191,772,740	27352	L
sh1-12A	chr2	36,660,276	36,808,732	148456	G
sh1-12A	chr2	204,216,558	204,238,402	21844	L
sh1-6A	chr20	15,077,029	15,356,533	279504	L
sh5A_APH	chr3	37,402,748	37,430,761	28013	L
sh7A_APH	chr3	60,735,899	60,749,321	13422	L
sh7A_APH	chr3	118,175,121	118,439,274	264153	L
sh1-12A	chr3	60,664,488	60,683,737	19249	L
sh1-5A	chr3	33,880,056	33,917,150	37094	L
sh1-5A	chr3	187,924,901	187,931,652	6751	L
sh1-7A	chr3	41,358,860	41,622,410	263550	L
sh4A	chr5	64,492,944	64,719,601	226657	L
sh10A	chr6	56,948,181	57,249,995	301814	G
sh6A	chr7	77,874,687	77,917,440	42753	L
sh9A_APH	chr8	112,626,847	112,681,758	54911	G
sh2-6A	chr8	109735350	110092044	356694	G

Table A.4. List of unique, *de novo* CNVs in *RNASEH1* knockdown + doxycycline and APH clones. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication).

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
E16	chr1	85,725,117	85,981,757	256640	L
E13	chr10	34876391	34959157	82766	L
E2	chr10	52,719,600	53,081,138	361538	G
E3	chr11	95,809,116	96,052,081	242965	L
E3	chr13	110,889,715	111,084,656	194941	L
E3	chr5	118,456,696	118,516,913	60217	L
E3	chr7	111,092,477	111,196,093	103616	L
E6	chr7	110,976,750	111,118,947	142197	G

Table A.5. List of unique, *de novo* CNVs in Empty control + doxycycline clones. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication).

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
E13A	chr1	72,377,151	72,565,293	188142	L
E13A	chr1	244,210,091	244,562,829	352738	G
E7A	chr1	32,673,183	32,786,845	113662	G
E14A	chr10	53,762,920	53,922,819	159899	L
E2	chr10	52,719,600	53,081,138	361538	G
E6A	chr10	53,444,657	53,510,461	65804	G
E6A	chr10	79,037,372	79,290,584	253212	L
E8A	chr10	53,304,988	53,373,421	68433	L
E8A	chr11	78,742,527	78,785,279	42752	G
E10A	chr12	1,760,477	2,091,257	330780	G
E10A	chr12	124,821,402	124,897,565	76163	L
E10A	chr14	56,725,209	56,801,445	76236	L
E14A	chr14	35,988,106	36,013,716	25610	L
E8A	chr16	83,020,062	83,220,123	200061	L
E10A	chr17	35,469,924	35,833,923	363999	L
E10A	chr18	7,404,948	7,556,208	151260	G
E13A	chr18	37,050,275	37,117,792	67517	L
E13A	chr19	39,728,625	40,056,462	327837	G
E8A	chr2	179,315,141	179,886,265	571124	L
E7A	chr3	3,947,755	4,001,655	53900	G
E9A	chr3	130,653,300	130,768,830	115530	L
E10A	chr4	101,964,133	102,041,236	77103	L
E6A	chr4	12,997,150	13,010,694	13544	G
E8A	chr4	143,181,688	143,221,777	40089	L
E13A	chr6	45,468,073	45,781,941	313868	G
E6A	chr7	78,407,125	78,483,068	75943	L
E10A	chr8	195,606	231,871	36265	L

Table A.6. List of unique, *de novo* CNVs in Empty control + doxycycline and APH clones. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication).

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
O6	chr2	36,622,940	36,648,398	25458	G

Table A.7. List of unique, *de novo* CNVs in *RNASEH1* overexpression + doxycycline clones. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication).

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
O13A	chr1	174,908,226	175,416,603	508377	G
O3A	chr1	201,661,704	201,711,959	50255	L
O6A	chr1	72,184,477	72,629,795	445318	L
O13A	chr14	65,208,572	65,210,267	1695	G
O8A	chr16	53,749,127	54,056,879	307752	L
O8A	chr16	78,540,591	78,607,832	67241	L
O6A	chr19	1,492,620	1,497,060	4440	L
O12A	chr2	18,852,966	19,346,130	493164	L
O3A	chr4	142,982,420	143,328,376	345956	G
O8A	chr5	160,753,462	161,155,692	402230	G
O8A	chr6	78,735,367	78,803,881	68514	L
O3A	chr7	78,130,026	78,183,552	53526	L
O3A	chr7	124,433,600	125,461,881	1028281	G
O6A	chr7	78,118,215	78,205,557	87342	L
O8A	chr7	110,655,892	110,850,063	194171	L
O5A	chrX	96,229,117	96,358,980	129863	G

Table A.8. List of unique, *de novo* CNVs in *RNASEH1* overexpression + doxycycline and APH clones. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication).

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
Scr1-21A	chr1	72,108,797	72,234,070	125273	L
Scr1-21A	chr1	72,296,122	72,512,864	216742	L
scr1-11A	chr1	51,866,768	51,953,306	86538	L
scr1-11A	chr1	98,203,819	98,539,032	335213	L
Scr2-18A	chr1	72,336,912	77,493,510	5156598	L
Scr1-5	chr10	103,733,135	103,767,614	34479	L
Scr1-9	chr10	69,750,596	69,780,717	30121	G
Scr2-17	chr10	131575757	131627834	52077	L
Scr2-18A	chr12	99,855,731	100,255,744	400013	L
Scr1-2	chr13	110,896,738	110,900,111	3373	L
Scr2-7	chr13	58581310	58600904	19594	G
Scr2-10A	chr14	56,078,738	56,110,539	31801	L
Scr1-16A	chr15	61,074,035	61,080,335	6300	L
Scr1-21A	chr16	72,423,462	72,439,749	16287	L
Scr1-9	chr16	23,563,500	23,714,311	150811	G
Scr2-18A	chr16	78,617,930	78,680,728	62798	L
Scr2-18A	chr16	83,088,211	83,198,489	110278	G
Scr2-21A	chr16	78,634,831	78,792,498	157667	L
Scr1-16A	chr18	68,719,952	68,726,105	6153	G
Scr1-16A	chr2	136,157,318	136,236,787	79469	L
Scr1-2A	chr2	206,472,559	206,858,172	385613	G
Scr1-16A	chr3	60,604,488	60,620,114	15626	L
Scr1-1	chr3	4,160,142	4,538,738	378596	G
Scr1-10	chr3	7,380,247	7,386,444	6197	G
Scr2-18A	chr3	67,529,226	67,648,908	119682	G
Scr2-20A	chr3	60,243,577	60,693,253	449676	L
Scr2-21A	chr3	59,999,964	60,004,228	4264	L
Scr1-2	chr4	372,995	497,138	124143	L
Scr1-10	chr5	64,467,769	64,749,379	281610	L
Scr2-13	chr6	26405263	26414488	9225	G
Scr2-13	chr6	26465383	26466234	851	G
Scr1-16A	chr7	77,866,710	78,345,207	478497	L
Scr2-7	chr7	110459283	110486408	27125	L
Scr1-10	chr9	9,652,477	9,673,320	20843	L
Scr1-16A	chrX	36,308,773	36,374,854	66081	G
Scr2-10A	chrX	36,509,237	36,537,763	28526	G

Table A.9. List of unique, *de novo* CNVs in Scrambled control clones in the first replicate of the knockdown experiment. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication). The APH-treated clones are in red.

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
Scr14A	chr1	72,221,729	72,362,338	140609	L
Scr17A	chr1	72,367,960	72,424,117	56157	L
Scr18A	chr1	72,508,295	72,703,860	195565	G
Scr1	chr1	72,823,712	72,954,883	131171	G
Scr10	chr1	72,842,545	72,899,190	56645	L
Scr10A	chr10	53,871,454	54,377,691	506237	G
Scr17A	chr10	53,326,699	53,455,408	128709	L
Scr10A	chr11	114,449,607	114,474,610	25003	G
Scr10-1	chr15	71,631,172	71,673,459	42287	L
Scr10-1	chr15	71,631,172	71,673,459	42287	L
Scr10A	chr16	72,327,024	72,381,460	54436	G
Scr1	chr16	17,263,999	17,356,258	92259	L
Scr7-1A	chr19	1,912,948	1,926,027	13079	G
Scr14A	chr2	36,738,629	36,842,336	103707	G
Scr18A	chr2	214,702,401	214,762,709	60308	G
Scr2-1	chr3	50,008,117	50,064,090	55973	L
Scr10A	chr4	32,447,037	32,452,973	5936	G
Scr12A	chr5	145,503,563	145,669,064	165501	L
Scr6	chr5	60,507,328	60,664,356	157028	L
Scr10A	chr7	77,797,289	78,289,970	492681	L
Scr10A	chr7	78,315,216	78,348,879	33663	L
Scr18A	chr7	110,941,859	110,982,750	40891	L
Scr17A	chr8	50,374,788	50,424,019	49231	G
Scr18A	chr8	4,170,496	4,407,443	236947	L
Scr6	chr8	141,004,454	141,059,415	54961	L

Table A.10. List of unique, *de novo* CNVs in Scrambled control clones in the second replicate of the knockdown experiment. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication). The APH-treated clones are in red.

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
sh1-4	chr1	172,252,304	172,487,743	235439	G
sh1-6A	chr1	71,633,718	71,868,535	234817	G
sh1-6A	chr1	72,315,819	72,637,154	321335	L
sh2-6_2	chr1	2870849	2980578	109729	G
sh2-6_2	chr10	126642073	126656168	14095	L
sh1-15	chr11	121,916,950	121,973,583	56633	G
sh2-5A	chr11	129,794,949	129,950,592	155643	G
sh1-4	chr12	88,837,978	88,949,124	111146	G
sh1-21	chr13	110,877,045	111,014,716	137671	L
sh1-4	chr13	110,811,450	111,043,309	231859	L
sh1-4	chr13	111,044,407	111,081,408	37001	L
sh1-4	chr13	111,082,057	111,094,728	12671	L
sh1-13A	chr13	71,905,357	72,125,641	220284	G
sh2-1	chr13	28186344	28235239	48895	L
sh2-6_2	chr13	73876054	73920464	44410	G
sh1-1	chr15	101,155,565	101,458,429	302864	L
sh2-6A	chr15	46,696,827	46,969,077	272250	G
sh2-9_2	chr15	85207361	85341859	134498	L
sh1-19	chr16	12,040,510	12,092,575	52065	G
sh1-21	chr16	7,760,959	11,085,603	3324644	G
sh1-6A	chr17	26,410,260	26,510,217	99957	L
sh1-12A	chr2	36,660,276	36,808,732	148456	G
sh1-12A	chr2	204,216,558	204,238,402	21844	L
sh2-8_2	chr2	234889976	235096984	207008	G
sh1-6A	chr20	15,077,029	15,356,533	279504	L
sh1-21	chr3	87,665,971	88,752,258	1086287	G
sh1-4	chr3	105,142,722	105,190,870	48148	G
sh1-12A	chr3	60,664,488	60,683,737	19249	L
sh1-5A	chr3	33,880,056	33,917,150	37094	L
sh1-5A	chr3	187,924,901	187,931,652	6751	L
sh1-7A	chr3	41,358,860	41,622,410	263550	L
sh1-9	chr4	2,791,990	3,110,509	318519	L
sh2-14	chr4	32524423	32550474	26051	G
sh1-4	chr7	137,667,969	137,921,040	253071	G
sh1-6	chr7	82,100,279	82,262,193	161914	G
sh2-6A	chr8	109,735,350	110,092,044	356694	G
sh1-4	chrX	88,852,133	89,729,690	877557	G

Table A.11. List of unique, *de novo* CNVs in *RNASEH1* knockdown clones in the first replicate of the knockdown experiment. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication). The APH-treated clones are in red.

SAMPLE	CHROM	START	END	SPAN	LOSS/GAIN
sh2A_APH	chr1	245,448,885	245,904,391	455506	L
sh6A	chr1	85,980,226	86,004,787	24561	L
sh6A	chr1	85,868,701	86,021,169	152468	L
sh2a_dox	chr1	236,702,209	236,714,291	12082	G
sh2A_APH	chr10	129,051,516	129,366,742	315226	G
sh2a_dox	chr11	35,215,725	35,218,445	2720	L
sh2a_dox	chr11	35,215,725	35,218,445	2720	L
sh2A_APH	chr12	99,915,661	100,056,529	140868	L
sh2A_APH	chr15	60,681,114	60,694,820	13706	L
sh5A_APH	chr16	17,204,226	17,312,099	107873	L
sh6A	chr16	83,735,129	83,805,328	70199	G
sh3A	chr2	191,745,388	191,772,740	27352	L
sh12a	chr20	9,153,061	9,202,117	49056	G
sh7a_dox	chr21	40,075,872	40,251,856	175984	G
sh5A_APH	chr3	37,402,748	37,430,761	28013	L
sh7A_APH	chr3	60,735,899	60,749,321	13422	L
sh7A_APH	chr3	118,175,121	118,439,274	264153	L
sh4A	chr5	64,492,944	64,719,601	226657	L
sh11a	chr5	103,536,192	104,616,894	1080702	L
sh10A	chr6	56,948,181	57,249,995	301814	G
sh6A	chr7	77,874,687	77,917,440	42753	L
sh11a	chr7	78,170,842	78,267,691	96849	L
sh11a	chr7	134,493,343	134,588,145	94802	L
sh9A_APH	chr8	112,626,847	112,681,758	54911	G

Table A.12. List of unique, *de novo* CNVs in *RNASEH1* knockdown clones in the second replicate of the knockdown experiment. SAMPLE: Sample name, CHROM: Chromosome, START & END: Breakpoint coordinates, SPAN: Size of the CNV, LOSS/GAIN: Denotes whether the CNV is a loss (deletion) or a gain (duplication). The APH-treated clones are in red.

Bibliography

Abbas, Tarek, Mignon A Keaton, and Anindya Dutta. 2013. "Genomic Instability in Cancer." *Cold Spring Harbor Perspectives in Biology* 5 (3). Cold Spring Harbor Laboratory Press: a012914. doi:10.1101/cshperspect.a012914.

Adams, R.L.P., S. Berryman, and A. Thomson. 1971. "Deoxyribonucleoside Triphosphate Pools in Synchronized and Drug-Inhibited L929 Cells." *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 240 (4). Elsevier: 455–62. doi:10.1016/0005-2787(71)90702-7.

Aguilera, Andrés, and Tatiana García-Muse. 2012. "R Loops: From Transcription Byproducts to Threats to Genome Stability." *Molecular Cell* 46 (2): 115–24. doi:10.1016/j.molcel.2012.04.009.

Alver, Robert C., Gaganmeet Singh Chadha, and J. Julian Blow. 2014. "The Contribution of Dormant Origins to Genome Stability: From Cell Biology to Human Genetics." *DNA Repair* 19 (July). Elsevier: 182–89. doi:10.1016/J.DNAREP.2014.03.012.

Amon, Jeremy D, and Douglas Koshland. 2016. "RNase H Enables Efficient Repair of R-Loop Induced DNA Damage." *ELife* 5. eLife Sciences Publications, Ltd. doi:10.7554/eLife.20533.

Anand, Ranjith P, Susan T Lovett, and James E Haber. 2013. "Break-Induced DNA Replication." *Cold Spring Harbor Perspectives in Biology* 5 (12). Cold Spring Harbor Laboratory Press: a010397. doi:10.1101/cshperspect.a010397.

Anderson, Keith R., Maximilian Haeussler, Colin Watanabe, Vasantharajan Janakiraman, Jessica Lund, Zora Modrusan, Jeremy Stinson, et al. 2018. "CRISPR Off-Target Analysis in Genetically Engineered Rats and Mice." *Nature Methods* 15 (7). Nature Publishing Group: 512–14. doi:10.1038/s41592-018-0011-5.

Antequera, F. 2003. "Structure, Function and Evolution of CpG Island Promoters." *Cellular and Molecular Life Sciences (CMLS)* 60 (8). Birkhäuser-Verlag: 1647–58. doi:10.1007/s00018-003-3088-6.

Aparicio-Prat, Estel, Carme Arnan, Ilaria Sala, Núria Bosch, Roderic Guigó, and Rory Johnson. 2015. "DECKO: Single-Oligo, Dual-CRISPR Deletion of Genomic Elements Including Long Non-Coding RNAs." *BMC Genomics* 16 (1). doi:10.1186/s12864-015-2086-z.

Arlt, Martin F, Jennifer G Mulle, Valerie M Schaibley, Ryan L Ragland, Sandra G Durkin, Stephen T Warren, and Thomas W Glover. 2009. "Replication Stress Induces Genome-Wide Copy Number Changes in Human Cells That Resemble Polymorphic and Pathogenic Variants." *American Journal of Human Genetics* 84 (3): 339–50. doi:10.1016/j.ajhg.2009.01.024.

Arlt, Martin F, Alev Cagla Ozdemir, Shanda R Birkeland, Robert H Lyons, Thomas W Glover, Thomas E. Wilson, and Thomas E. Wilson. 2011. "Comparison of Constitutional and Replication Stress-Induced Genome Structural Variation by SNP Array and Mate-Pair Sequencing." *Genetics* 187 (3). Genetics Society of America: 675–83.

doi:10.1534/genetics.110.124776.
Arlt, Martin F, Alev Cagla Ozdemir, Shanda R Birkeland, Thomas E Wilson, and Thomas W Glover. 2011. "Hydroxyurea Induces de Novo Copy Number Variants in Human Cells." *Proceedings of the National Academy of Sciences of the United States of America* 108 (42): 17360–65. doi:10.1073/pnas.1109272108.
Arlt, Martin F, Sountharia Rajendran, Shanda R Birkeland, Thomas E Wilson, and Thomas W Glover. 2012. "De Novo CNV Formation in Mouse Embryonic Stem Cells Occurs in the Absence of Xrcc4-Dependent Nonhomologous End Joining." *PLoS Genetics* 8 (9). Public Library of Science: e1002981. doi:10.1371/journal.pgen.1002981.
———. 2014. "Copy Number Variants Are Produced in Response to Low-Dose Ionizing Radiation in Cultured Cells." *Environmental and Molecular Mutagenesis* 55 (2): 103–13. doi:10.1002/em.21840.
Arlt, Martin F, Thomas E Wilson, and Thomas W Glover. 2012. "Replication Stress and Mechanisms of CNV Formation." *Current Opinion in Genetics & Development* 22 (3): 204–10. doi:10.1016/j.gde.2012.01.009.
Barlow, Jacqueline H, Robert B Faryabi, Elsa Callén, Nancy Wong, Amy Malhowski, Hua Tang Chen, Gustavo Gutierrez-Cruz, et al. 2013. "Identification of Early Replicating Fragile Sites That Contribute to Genome Instability." *Cell* 152 (3): 620–32. doi:10.1016/j.cell.2013.01.006.
Basu, Uttiya, Fei-Long Meng, Celia Keim, Veronika Grinstein, Evangelos Pefanis, Jennifer Eccleston, Tingting Zhang, et al. 2011. "The RNA Exosome Targets the AID Cytidine Deaminase to Both Strands of Transcribed Duplex DNA Substrates." *Cell* 144 (3): 353–63. doi:10.1016/j.cell.2011.01.001.
Bauer, Daniel E, Matthew C Canver, and Stuart H Orkin. 2015. "Generation of Genomic Deletions in Mammalian Cell Lines via CRISPR/Cas9." *Journal of Visualized Experiments : JoVE*, no. 95 (January). MyJoVE Corporation: e52118. doi:10.3791/52118.
Baumann, Christoph, Roman Körner, Kay Hofmann, and Erich A. Nigg. 2007. "PICH, a Centromere-Associated SNF2 Family ATPase, Is Regulated by Plk1 and Required for the Spindle Checkpoint." *Cell* 128 (1): 101–14. doi:10.1016/j.cell.2006.11.041.
Beck, H., V. Nahse-Kumpf, M. S. Y. Larsen, K. A. O'Hanlon, S. Patzke, C. Holmberg, J. Mejlvang, et al. 2012. "Cyclin-Dependent Kinase Suppression by WEE1 Kinase Protects the Genome through Control of Replication Initiation and Nucleotide Consumption." *Molecular and Cellular Biology* 32 (20): 4226–36. doi:10.1128/MCB.00412-12.
BedingerS, Patricia, Maureen Munng, and Bruce M Alberts. 1989. "THE JOURNAL OF BIOLOGICAL CHEMISTRY Sequence-Specific Pausing during in Vitro DNA Replication on Double-Stranded DNA Templates*." Vol. 264. http://www.jbc.org/content/264/28/16880.full.pdf.
Bermejo, R., Y. Doksani, T. Capra, Y.-M. Katou, H. Tanaka, K. Shirahige, and M. Foiani. 2007. "Top1- and Top2-Mediated Topological Transitions at Replication Forks Ensure Fork Progression and Stability and Prevent DNA Damage Checkpoint Activation." *Genes & Development* 21 (15): 1921–36. doi:10.1101/gad.432107.
Bermejo, Rodrigo, Thelma Capra, Victor Gonzalez-Huici, Daniele Fachinetti, Andrea Cocito, Gioacchino Natoli, Yuki Katou, et al. 2009. "Genome-Organizing Factors Top2 and Hmo1 Prevent Chromosome Fragility at Sites of S Phase Transcription." *Cell* 138 (5): 870–84. doi:10.1016/j.cell.2009.06.022.
Beroukhim, Rameen, Craig H. Mermel, Dale Porter, Guo Wei, Soumya Raychaudhuri, Jerry

Donovan, Jordi Barretina, et al. 2010. “The Landscape of Somatic Copy-Number Alteration across Human Cancers.” *Nature* 463 (7283). Macmillan Publishers Limited. All rights reserved: 899–905. doi:10.1038/nature08822.

Bester, Assaf C., Maayan Roniger, Yifat S. Oren, Michael M. Im, Dan Sarni, Malka Chaoat, Aaron Bensimon, Gideon Zamir, Donna S. Shewach, and Batsheva Kerem. 2011. “Nucleotide Deficiency Promotes Genomic Instability in Early Stages of Cancer Development.” *Cell* 145 (3): 435–46. doi:10.1016/j.cell.2011.03.044.

Beunders, Gea, Els Voorhoeve, Christelle Golzio, Luba M Pardo, Jill a Rosenfeld, Michael E Talkowski, Ingrid Simonic, et al. 2013. “Exonic Deletions in AUTS2 Cause a Syndromic Form of Intellectual Disability and Suggest a Critical Role for the C Terminus.” *American Journal of Human Genetics* 92 (2): 210–20. doi:10.1016/j.ajhg.2012.12.011.

Bhatia, Vaibhav, Sonia I. Barroso, María L. García-Rubio, Emanuela Tumini, Emilia Herrera-Moyano, and Andrés Aguilera. 2014. “BRCA2 Prevents R-Loop Accumulation and Associates with TREX-2 MRNA Export Factor PCID2.” *Nature* 511 (7509): 362–65. doi:10.1038/nature13374.

Bhowmick, Rahul, Sheroy Minocherhomji, and Ian D. Hickson. 2016. “RAD52 Facilitates Mitotic DNA Synthesis Following Replication Stress.” *Molecular Cell* 64 (6): 1117–26. doi:10.1016/j.molcel.2016.10.037.

Bianco, P R, R B Tracy, and S C Kowalczykowski. 1998. “DNA Strand Exchange Proteins: A Biochemical and Physical Comparison.” *Frontiers in Bioscience : A Journal and Virtual Library* 3 (June): D570-603. http://www.ncbi.nlm.nih.gov/pubmed/9632377.

Bicknell, Louise S, Ernie M H F Bongers, Andrea Leitch, Stephen Brown, Jeroen Schoots, Margaret E Harley, Salim Aftimos, et al. 2011. “Mutations in the Pre-Replication Complex Cause Meier-Gorlin Syndrome.” *Nature Genetics* 43 (4). Europe PMC Funders: 356–59. doi:10.1038/ng.775.

Biebricher, Andreas, Seiki Hirano, Jacqueline H. Enzlin, Nicola Wiechens, Werner W. Streicher, Diana Huttner, Lily H.-C. Wang, et al. 2013. “PICH: A DNA Translocase Specially Adapted for Processing Anaphase Bridge DNA.” *Molecular Cell* 51 (5): 691–701. doi:10.1016/j.molcel.2013.07.016.

Bignell, Graham R., Chris D. Greenman, Helen Davies, Adam P. Butler, Sarah Edkins, Jenny M. Andrews, Gemma Buck, et al. 2010. “Signatures of Mutation and Selection in the Cancer Genome.” *Nature* 463 (7283). Nature Publishing Group: 893–98. doi:10.1038/nature08768.

Blow, J. Julian, Xin Quan Ge, and Dean A. Jackson. 2011. “How Dormant Origins Promote Complete Genome Replication.” *Trends in Biochemical Sciences* 36 (8): 405–14. http://www.sciencedirect.com/science/article/pii/S0968000411000703.

Blow, J Julian, and Xin Quan Ge. 2009. “A Model for DNA Replication Showing How Dormant Origins Safeguard against Replication Fork Failure.” *EMBO Reports* 10 (4): 406–12. doi:10.1038/embor.2009.5.

Blumenfeld, Britny, Micha Ben-Zimra, and Itamar Simon. 2017. “Perturbations in the Replication Program Contribute to Genomic Instability in Cancer.” Edited by Vassilis G Gorgoulis. *International Journal of Molecular Sciences* 18 (6). MDPI: 1138. doi:10.3390/ijms18061138.

Boden, D., Oliver Pusch, Rebecca Silbermann, Fred Lee, Lynne Tucker, and Bharat Ramratnam. 2004. “Enhanced Gene Silencing of HIV-1 Specific SiRNA Using MicroRNA Designed Hairpins.” *Nucleic Acids Research* 32 (3): 1154–58. doi:10.1093/nar/gkh278.

Boerma, E. G., R. Siebert, P. M. Kluin, and M. Baudis. 2009. “Translocations Involving 8q24 in

Burkitt Lymphoma and Other Malignant Lymphomas: A Historical Review of Cytogenetics in the Light of Todays Knowledge." *Leukemia*. doi:10.1038/leu.2008.281.

Boguslawski, Sophie J., Dennis E. Smith, Mary A. Michalak, Kenneth E. Mickelson, Clifford O. Yehle, William L. Patterson, and Robert J. Carrico. 1986. "Characterization of Monoclonal Antibody to DNA · RNA and Its Application to Immunodetection of Hybrids." *Journal of Immunological Methods* 89 (1). Elsevier: 123–30. doi:10.1016/0022-1759(86)90040-2.

Boye, E, A Løbner-Olesen, and K Skarstad. 2000. "Limiting DNA Replication to Once and Only Once." *EMBO Reports* 1 (6). European Molecular Biology Organization: 479–83. doi:10.1093/embo-reports/kvd116.

Breier, Adam M, Heinz-Ulrich G Weier, and Nicholas R Cozzarelli. 2005. "Independence of Replisomes in Escherichia Coli Chromosomal Replication." *Proceedings of the National Academy of Sciences of the United States of America* 102 (11): 3942–47. doi:10.1073/pnas.0500812102.

Brewer, B J. 1988. "When Polymerases Collide: Replication and the Transcriptional Organization of the E. Coli Chromosome." *Cell* 53 (5): 679–86. http://www.ncbi.nlm.nih.gov/pubmed/3286014.

Brewer, B J, and W L Fangman. 1988. "A Replication Fork Barrier at the 3' End of Yeast Ribosomal RNA Genes." *Cell* 55 (4): 637–43. http://www.ncbi.nlm.nih.gov/pubmed/3052854.

Burhans, William C, and Martin Weinberger. 2007. "DNA Replication Stress, Genome Instability and Aging." *Nucleic Acids Research* 35 (22). Oxford University Press: 7545–56. doi:10.1093/nar/gkm1059.

Burrow, Allison A., Allison Marullo, Lindsay R. Holder, and Yuh-Hwa Wang. 2010. "Secondary Structure Formation and DNA Instability at Fragile Site FRA16B." *Nucleic Acids Research* 38 (9). Oxford University Press: 2865–77. doi:10.1093/nar/gkp1245.

Byrnes, John J., Kathleen M. Downey, Vicky L. Black, and Antero G. So. 1976. "A New Mammalian DNA Polymerase with 3' to 5' Exonuclease Activity: DNA Polymerase δ." *Biochemistry* 15 (13). American Chemical Society: 2817–23. doi:10.1021/bi00658a018.

Cadet, Jean, and J Richard Wagner. 2013. "DNA Base Damage by Reactive Oxygen Species, Oxidizing Agents, and UV Radiation." *Cold Spring Harbor Perspectives in Biology* 5 (2). Cold Spring Harbor Laboratory Press. doi:10.1101/cshperspect.a012559.

Campbell, Peter J, Philip J Stephens, Erin D Pleasance, Sarah O'Meara, Heng Li, Thomas Santarius, Lucy A Stebbings, et al. 2008. "Identification of Somatically Acquired Rearrangements in Cancer Using Genome-Wide Massively Parallel Paired-End Sequencing." *Nature Genetics* 40 (6): 722–29. doi:10.1038/ng.128.

Campuzano, V, L Montermini, M D Moltò, L Pianese, M Cossée, F Cavalcanti, E Monros, et al. 1996. "Friedreich's Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion." *Science (New York, N.Y.)* 271 (5254): 1423–27. http://www.ncbi.nlm.nih.gov/pubmed/8596916.

Canver, Matthew C, Daniel E Bauer, Abhishek Dass, Yvette Y Yien, Jacky Chung, Takeshi Masuda, Takahiro Maeda, Barry H Paw, and Stuart H Orkin. 2014. "Efficient CRISPR-Mediated Deletion Characterization of Genomic Deletion Efficiency Mediated by CRISPR/Cas9 in Mammalian Cells*." doi:10.1074/jbc.M114.564625.

Carvalho, Claudia M B, and James R Lupski. 2016. "Mechanisms Underlying Structural Variant Formation in Genomic Disorders." *Nature Reviews. Genetics* 17 (4). NIH Public Access: 224–38. doi:10.1038/nrg.2015.25.

Cerritelli, Susana M, Ella G Frolova, Chiguang Feng, Alexander Grinberg, Paul E Love, and Robert J Crouch. 2003. "Failure to Produce Mitochondrial DNA Results in Embryonic Lethality in Rnaseh1 Null Mice." *Molecular Cell* 11 (3): 807–15. http://www.ncbi.nlm.nih.gov/pubmed/12667461.

Chan, Kok-Lung, Phillip S North, and Ian D Hickson. 2007. "BLM Is Required for Faithful Chromosome Segregation and Its Localization Defines a Class of Ultrafine Anaphase Bridges." *The EMBO Journal* 26 (14): 3397 LP-3409. http://emboj.embopress.org/content/26/14/3397.abstract.

Chan, Kok Lung, and Ian D. Hickson. 2011. "New Insights into the Formation and Resolution of Ultra-Fine Anaphase Bridges." *Seminars in Cell & Developmental Biology* 22 (8): 906–12. doi:10.1016/j.semcdb.2011.07.001.

Chaudhuri, Jayanta, Ming Tian, Chan Khuong, Katrin Chua, Eric Pinaud, and Frederick W. Alt. 2003. "Transcription-Targeted DNA Deamination by the AID Antibody Diversification Enzyme." *Nature* 422 (6933). Nature Publishing Group: 726–30. doi:10.1038/nature01574.

Chen, Xiangyang, Fei Xu, Chengming Zhu, Jiaojiao Ji, Xufei Zhou, Xuezhu Feng, and Shouhong Guang. 2014. "Dual SgRNA-Directed Gene Knockout Using CRISPR/Cas9 Technology in Caenorhabditis Elegans." *Scientific Reports* 4. doi:10.1038/srep07581.

Chen, Ying-Zhang, Craig L. Bennett, Huy M. Huynh, Ian P. Blair, Imke Puls, Joy Irobi, Ines Dierick, et al. 2004. "DNA/RNA Helicase Gene Mutations in a Form of Juvenile Amyotrophic Lateral Sclerosis (ALS4)." *The American Journal of Human Genetics* 74 (6): 1128–35. doi:10.1086/421054.

Chiarle, Roberto, Yu Zhang, Richard L. Frock, Susanna M. Lewis, Benoit Molinie, Yu-Jui Ho, Darienne R. Myers, et al. 2011. "Genome-Wide Translocation Sequencing Reveals Mechanisms of Chromosome Breaks and Rearrangements in B Cells." *Cell* 147 (1): 107–19. doi:10.1016/j.cell.2011.07.049.

Chien, Y H, and N Davidson. 1978. "RNA:DNA Hybrids Are More Stable than DNA:DNA Duplexes in Concentrated Perchlorate and Trichloroacetate Solutions." *Nucleic Acids Research* 5 (5). Oxford University Press: 1627–37. http://www.ncbi.nlm.nih.gov/pubmed/208059.

Cho, Seung Woo, Sojung Kim, Yongsub Kim, Jiyeon Kweon, Heon Seok Kim, Sangsu Bae, and Jin Soo Kim. 2014. "Analysis of Off-Target Effects of CRISPR/Cas-Derived RNA-Guided Endonucleases and Nickases." *Genome Research* 24 (1): 132–41. doi:10.1101/gr.162339.113.

Cohen, Sarah, Nadine Puget, Yea-Lih Lin, Thomas Clouaire, Marion Aguirrebengoa, Vincent Rocher, Philippe Pasero, Yvan Canitrot, and Gaëlle Legube. 2018. "Senataxin Resolves RNA:DNA Hybrids Forming at DNA Double-Strand Breaks to Prevent Translocations." *Nature Communications* 9 (1). Nature Publishing Group: 533. doi:10.1038/s41467-018-02894-w.

Conrad, Donald F., Dalila Pinto, Richard Redon, Lars Feuk, Omer Gokcumen, Yujun Zhang, Jan Aerts, et al. 2010. "Origins and Functional Impact of Copy Number Variation in the Human Genome." *Nature* 464 (7289): 704–12. doi:10.1038/nature08516.

Conrad, Donald F, Christine Bird, Ben Blackburne, Sarah Lindsay, Lira Mamanova, Charles Lee, Daniel J Turner, and Matthew E Hurles. 2010. "Mutation Spectrum Revealed by Breakpoint Sequencing of Human Germline CNVs." *Nature Genetics* 42 (5). Nature Publishing Group: 385–91. doi:10.1038/ng.564.

Core, Leighton J, Joshua J Waterfall, and John T Lis. 2008. "Nascent RNA Sequencing Reveals

Widespread Pausing and Divergent Initiation at Human Promoters." *Science (New York, N.Y.)* 322 (5909). American Association for the Advancement of Science: 1845–48. doi:10.1126/science.1162228.

Cortez, David. 2015. "Preventing Replication Fork Collapse to Maintain Genome Integrity." *DNA Repair* 32 (August). Elsevier: 149–57. doi:10.1016/J.DNAREP.2015.04.026.

Costantino, Lorenzo, Sotirios K. Sotiriou, Juha K. Rantala, Simon Magin, Emil Mladenov, Thomas Helleday, James E. Haber, George Iliakis, Olli P. Kallioniemi, and Thanos D. Halazonetis. 2014. "Break-Induced Replication Repair of Damaged Forks Induces Genomic Duplications in Human Cells." *Science* 343 (6166): 88–91. doi:10.1126/science.1243211.

Creyghton, Menno P, Albert W Cheng, G Grant Welstead, Tristan Kooistra, Bryce W Carey, Eveline J Steine, Jacob Hanna, et al. 2010. "Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State." *Proceedings of the National Academy of Sciences of the United States of America* 107 (50). National Academy of Sciences: 21931–36. doi:10.1073/pnas.1016071107.

Cristini, Agnese, Matthias Groh, Maiken S. Kristiansen, and Natalia Gromak. 2018. "RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage." *Cell Reports* 23 (6): 1891–1905. doi:10.1016/j.celrep.2018.04.025.

Crow, Yanick J, Andrea Leitch, Bruce E Hayward, Anna Garner, Rekha Parmar, Elen Griffith, Manir Ali, et al. 2006. "Mutations in Genes Encoding Ribonuclease H2 Subunits Cause Aicardi-Goutières Syndrome and Mimic Congenital Viral Brain Infection." *Nature Genetics* 38 (8). Nature Publishing Group: 910–16. doi:10.1038/ng1842.

Dagmar, Michael, David G Beer, Charles W Wilke, Diane E Miller, and Thomas W Glover. 1997. "Frequent Deletions of FHIT and FRA3B in Barrett's Metaplasia and Esophageal Adenocarcinomas." *Oncogene* 15 (14): 1653–59. doi:10.1038/sj.onc.1201330.

Darzacq, Xavier, Yaron Shav-Tal, Valeria de Turris, Yehuda Brody, Shailesh M Shenoy, Robert D Phair, and Robert H Singer. 2007. "In Vivo Dynamics of RNA Polymerase II Transcription." *Nature Structural & Molecular Biology* 14 (9): 796–806. doi:10.1038/nsmb1280.

Davis, Allison P, and Lorraine S Symington. 2004. "RAD51-Dependent Break-Induced Replication in Yeast." *Molecular and Cellular Biology* 24 (6). American Society for Microbiology Journals: 2344–51. doi:10.1128/MCB.24.6.2344-2351.2004.

de Munnik, Sonja A, Louise S Bicknell, Salim Aftimos, Jumana Y Al-Aama, Yolande van Bever, Michael B Bober, Jill Clayton-Smith, et al. 2012. "Meier–Gorlin Syndrome Genotype–Phenotype Studies: 35 Individuals with Pre-Replication Complex Gene Mutations and 10 without Molecular Diagnosis." *European Journal of Human Genetics* 20 (6): 598–606. doi:10.1038/ejhg.2011.269.

De Septenville, Anne L, Stéphane Duigou, Hasna Boubakri, and Bénédicte Michel. 2012. "Replication Fork Reversal after Replication-Transcription Collision." *PLoS Genetics* 8 (4): e1002622. doi:10.1371/journal.pgen.1002622.

Debacker, Kim, and R.F. Frank Kooy. 2007. "Fragile Sites and Human Disease." *Human Molecular Genetics* 16 Spec No (R2). Oxford University Press: R150-8. doi:10.1093/hmg/ddm136.

Debatisse, MIchelle, Eliane El Achkar, and Bernard Dutrillaux. 2006. "Common Fragile Sites Nested at the Interfaces of Early and Late-Replicating Chromosome Bands : Cis Acting Components of the G2/M Checkpoint?" *Cell Cycle* 5 (6): 578–81. doi:10.4161/cc.5.6.2574.

Debatisse, Michelle, Benoît Le Tallec, Anne Letessier, Bernard Dutrillaux, and Olivier Brison. 2012. "Common Fragile Sites: Mechanisms of Instability Revisited." *Trends in Genetics : TIG* 28 (1): 22–32. doi:10.1016/j.tig.2011.10.003.

Deshpande, A M, and C S Newlon. 1996. "DNA Replication Fork Pause Sites Dependent on Transcription." *Science (New York, N.Y.)* 272 (5264): 1030–33. http://www.ncbi.nlm.nih.gov/pubmed/8638128.

Di Micco, Raffaella, Marzia Fumagalli, Angelo Cicalese, Sara Piccinin, Patrizia Gasparini, Chiara Luise, Catherine Schurra, et al. 2006. "Oncogene-Induced Senescence Is a DNA Damage Response Triggered by DNA Hyper-Replication." *Nature* 444 (7119): 638–42. doi:10.1038/nature05327.

Dickerson, Sarah K, Eleonora Market, Eva Besmer, and F Nina Papavasiliou. 2003. "AID Mediates Hypermutation by Deaminating Single Stranded DNA." *The Journal of Experimental Medicine* 197 (10). The Rockefeller University Press: 1291–96. doi:10.1084/jem.20030481.

Dillon, Laura W, Levi C T Pierce, Maggie C Y Ng, and Yuh-Hwa Wang. 2013. "Role of DNA Secondary Structures in Fragile Site Breakage along Human Chromosome 10." *Human Molecular Genetics* 22 (7). Oxford University Press: 1443–56. doi:10.1093/hmg/dds561.

Dittwald, Piotr, Tomasz Gambin, Claudia Gonzaga-Jauregui, Claudia M.B. Carvalho, James R. Lupski, Paweł Stankiewicz, and Anna Gambin. 2013. "Inverted Low-Copy Repeats and Genome Instability-A Genome-Wide Analysis." *Human Mutation* 34 (1): 210–20. doi:10.1002/humu.22217.

Donley, Nathan, and Mathew J Thayer. 2013. "DNA Replication Timing, Genome Stability and Cancer: Late and/or Delayed DNA Replication Timing Is Associated with Increased Genomic Instability." *Seminars in Cancer Biology* 23 (2). NIH Public Access: 80–89. doi:10.1016/j.semcancer.2013.01.001.

Donovan, S., J. Harwood, L. S. Drury, and J. F. X. Diffley. 1997. "Cdc6p-Dependent Loading of Mcm Proteins onto Pre-Replicative Chromatin in Budding Yeast." *Proceedings of the National Academy of Sciences* 94 (11): 5611–16. doi:10.1073/pnas.94.11.5611.

Duquette, Michelle L., Phuong Pham, Myron F. Goodman, and Nancy Maizels. 2005. "AID Binds to Transcription-Induced Structures in c-MYC That Map to Regions Associated with Translocation and Hypermutation." *Oncogene* 24 (38): 5791–98. doi:10.1038/sj.onc.1208746.

Durkin, Sandra G, Ryan L Ragland, Martin F Arlt, Jennifer G Mulle, Stephen T Warren, and Thomas W Glover. 2008. "Replication Stress Induces Tumor-like Microdeletions in FHIT/FRA3B." *Proceedings of the National Academy of Sciences of the United States of America* 105 (1): 246–51. doi:10.1073/pnas.0708097105.

Edwards, Melissa C., Antonin V Tutter, Christin Cvetic, Catherine H Gilbert, Tatyana A Prokhorova, and Johannes C Walter. 2002. "MCM2-7 Complexes Bind Chromatin in a Distributed Pattern Surrounding the Origin Recognition Complex in Xenopus Egg Extracts." *The Journal of Biological Chemistry* 277 (36): 33049–57. doi:10.1074/jbc.M204438200.

Ekholm-Reed, Susanna, Juan Méndez, Donato Tedesco, Anders Zetterberg, Bruce Stillman, and Steven I. Reed. 2004. "Deregulation of Cyclin E in Human Cells Interferes with Prereplication Complex Assembly." *The Journal of Cell Biology* 165 (6): 789–800. doi:10.1083/jcb.200404092.

Elia, J., X. Gai, H. M. Xie, J. C. Perin, E. Geiger, J. T. Glessner, M. D'Arcy, et al. 2010. "Rare

Structural Variants Found in Attention-Deficit Hyperactivity Disorder Are Preferentially Associated with Neurodevelopmental Genes." *Molecular Psychiatry* 15 (6): 637–46. doi:10.1038/mp.2009.57.

Finnis, Merran, Sonia Dayan, Lynne Hobson, Georgia Chenevix-Trench, Kathryn Friend, Karin Ried, Deon Venter, Erica Woollatt, Elizabeth Baker, and Robert I. Richards. 2005. "Common Chromosomal Fragile Site FRA16D Mutation in Cancer Cells." *Human Molecular Genetics* 14 (10): 1341–49. doi:10.1093/hmg/ddi144.

Frattini, Annalisa, Marco Fabbri, Roberto Valli, Elena De Paoli, Giuseppe Montalbano, Laura Gribaldo, Francesco Pasquali, and Emanuela Maserati. 2015. "High Variability of Genomic Instability and Gene Expression Profiling in Different HeLa Clones." *Scientific Reports* 5 (1). Nature Publishing Group: 15377. doi:10.1038/srep15377.

Friedman, J. I., T. Vrijenhoek, S. Markx, I. M. Janssen, W. A. Van Der Vliet, B. H.W. Faas, N. V. Knoers, et al. 2008. "CNTNAP2 Gene Dosage Variation Is Associated with Schizophrenia and Epilepsy." *Molecular Psychiatry* 13 (3): 261–66. doi:10.1038/sj.mp.4002049.

Fu, Yanfang, Jennifer A. Foden, Cyd Khayter, Morgan L. Maeder, Deepak Reyon, J. Keith Joung, and Jeffry D. Sander. 2013. "High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells." *Nature Biotechnology* 31 (9): 822–26. doi:10.1038/nbt.2623.

Fungtammasan, Arkarachai, Erin Walsh, Francesca Chiaromonte, Kristin a Eckert, and Kateryna D Makova. 2012. "A Genome-Wide Analysis of Common Fragile Sites: What Features Determine Chromosomal Instability in the Human Genome?" *Genome Research* 22 (6). Cold Spring Harbor Laboratory Press: 993–1005. doi:10.1101/gr.134395.111.

Gaillard, Hélène, Tatiana García-Muse, and Andrés Aguilera. 2015. "Replication Stress and Cancer." *Nature Publishing Group* 15. doi:10.1038/nrc3916.

Gan, Wenjian, Zhishuang Guan, Jie Liu, Ting Gui, Keng Shen, James L Manley, and Xialu Li. 2011. "R-Loop-Mediated Genomic Instability Is Caused by Impairment of Replication Fork Progression." *Genes & Development* 25 (19): 2041–56. doi:10.1101/gad.17010011.

García-Pichardo, Desiré, Juan C Cañas, María L García-Rubio, Belén Gómez-González, Ana G Rondón, and Andrés Aguilera. 2017. "Histone Mutants Separate R Loop Formation from Genome Instability Induction." doi:10.1016/j.molcel.2017.05.014.

García-Rubio, María L., Carmen Pérez-Calero, Sonia I. Barroso, Emanuela Tumini, Emilia Herrera-Moyano, Iván V. Rosado, and Andrés Aguilera. 2015. "The Fanconi Anemia Pathway Protects Genome Integrity from R-Loops." Edited by Jeff Sekelsky. *PLOS Genetics* 11 (11). Public Library of Science: e1005674. doi:10.1371/journal.pgen.1005674.

Ge, Xin Quan, Dean a Jackson, and J Julian Blow. 2007. "Dormant Origins Licensed by Excess Mcm2-7 Are Required for Human Cells to Survive Replicative Stress." *Genes & Development* 21 (24). Cold Spring Harbor Laboratory Press: 3331–41. doi:10.1101/gad.457807.

Ginno, Paul A, Yoong Wearn Lim, Paul L Lott, Ian Korf, and Frédéric Chédin. 2013. "GC Skew at the 5' and 3' Ends of Human Genes Links R-Loop Formation to Epigenetic Regulation and Transcription Termination." *Genome Research* 23 (10). Cold Spring Harbor Laboratory Press: 1590–1600. doi:10.1101/gr.158436.113.

Ginno, Paul A, Paul L Lott, Holly C Christensen, Ian Korf, and Frédéric Chédin. 2012. "R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters." *Molecular Cell* 45 (6). NIH Public Access: 814–25. doi:10.1016/j.molcel.2012.01.017.

Girirajan, Santhosh, Catarina D. Campbell, and Evan E. Eichler. 2011. "Human Copy Number Variation and Complex Genetic Disease." *Annual Review of Genetics* 45 (1). Annual Reviews : 203–26. doi:10.1146/annurev-genet-102209-163544.

Glover, Thomas W., Carol Berger, Jane Coyle, and Barbara Echo. 1984. "DNA Polymerase ? Inhibition by Aphidicolin Induces Gaps and Breaks at Common Fragile Sites in Human Chromosomes." *Human Genetics* 67 (2): 136–42. doi:10.1007/BF00272988.

Glover, Thomas W., Thomas E. Wilson, and Martin F. Arlt. 2017. "Fragile Sites in Cancer: More than Meets the Eye." *Nature Reviews Cancer* 17 (8). Nature Publishing Group: 489–501. doi:10.1038/nrc.2017.52.

Glover, Thomas W, Martin F Arlt, Anne M Casper, and Sandra G Durkin. 2005. "Mechanisms of Common Fragile Site Instability." *Human Molecular Genetics* 14 Spec No (2): R197-205. doi:10.1093/hmg/ddi265.

Gómez-González, Belén, María García-Rubio, Rodrigo Bermejo, Hélène Gaillard, Katsuhiko Shirahige, Antonio Marín, Marco Foiani, and Andrés Aguilera. 2011. "Genome-Wide Function of THO/TREX in Active Genes Prevents R-Loop-Dependent Replication Obstacles." *The EMBO Journal* 30 (15). European Molecular Biology Organization: 3106–19. doi:10.1038/emboj.2011.206.

Goranov, Alexi I, Adam M Breier, Houra Merrikh, and Alan D Grossman. 2009. "YabA of Bacillus Subtilis Controls DnaA-Mediated Replication Initiation but Not the Transcriptional Response to Replication Stress." *Molecular Microbiology* 74 (2): 454–66. doi:10.1111/j.1365-2958.2009.06876.x.

Gregor, Anne, Beate Albrecht, Ingrid Bader, Emilia K. Bijlsma, Arif B. Ekici, Hartmut Engels, Karl Hackmann, et al. 2011. "Expanding the Clinical Spectrum Associated with Defects in CNTNAP2 and NRXN1." *BMC Medical Genetics* 12. doi:10.1186/1471-2350-12-106.

Groh, Matthias, Michele M. P. Lufino, Richard Wade-Martins, and Natalia Gromak. 2014. "R-Loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome." Edited by Andrés Aguilera. *PLoS Genetics* 10 (5): e1004318. doi:10.1371/journal.pgen.1004318.

Gu, Kangxia, Hon Keung Tony Ng, Man Lai Tang, and William R. Schucany. 2008. "Testing the Ratio of Two Poisson Rates." *Biometrical Journal* 50 (2). Wiley-Blackwell: 283–98. doi:10.1002/bimj.200710403.

Gu, Wenli, Feng Zhang, and James R Lupski. 2008. "Mechanisms for Human Genomic Rearrangements." *PathoGenetics* 1 (1). BioMed Central: 4. doi:10.1186/1755-8417-1-4.

Guler, Gulnur, Aysegul Uner, Nilufer Guler, Shuang-Yin Han, Dimitrios Iliopoulos, Peter McCue, and Kay Huebner. 2005. "Concordant Loss of Fragile Gene Expression Early in Breast Cancer Development." *Pathology International* 55 (8): 471–78. doi:10.1111/j.1440-1827.2005.01855.x.

Gyi, J I, A N Lane, G L Conn, and T Brown. 1998. "The Orientation and Dynamics of the C2'-OH and Hydration of RNA and DNA.RNA Hybrids." *Nucleic Acids Research* 26 (13). Oxford University Press: 3104–10. http://www.ncbi.nlm.nih.gov/pubmed/9628906.

Hamperl, Stephan, Michael J. Bocek, Joshua C. Saldivar, Tomek Swigut, and Karlene A. Cimprich. 2017. "Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses." *Cell* 170 (4): 774–786.e19. doi:10.1016/j.cell.2017.07.043.

Hamperl, Stephan, and Karlene A Cimprich. 2016. "Conflict Resolution in the Genome: How Transcription and Replication Make It Work." *Cell* 167 (6). NIH Public Access: 1455–67.

doi:10.1016/j.cell.2016.09.053.
Hansen, R Scott, Sean Thomas, Richard Sandstrom, Theresa K Canfield, Robert E Thurman, Molly Weaver, Michael O Dorschner, Stanley M Gartler, and John a Stamatoyannopoulos. 2010. "Sequencing Newly Replicated DNA Reveals Widespread Plasticity in Human Replication Timing." *Proceedings of the National Academy of Sciences of the United States of America* 107 (1): 139–44. doi:10.1073/pnas.0912402107.
Hastings, P J, James R Lupski, Susan M Rosenberg, and Grzegorz Ira. 2009. "Mechanisms of Change in Gene Copy Number." *Nature Reviews. Genetics* 10 (8): 551–64. doi:10.1038/nrg2593.
Hatchi, Elodie, Konstantina Skourti-Stathaki, Steffen Ventz, Luca Pinello, Angela Yen, Kinga Kamieniarz-Gdula, Stoil Dimitrov, et al. 2015. "BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair." *Molecular Cell* 57 (4): 636–47. doi:10.1016/j.molcel.2015.01.011.
Hellman, A., A. Rahat, S. W. Scherer, A. Darvasi, L.-C. Tsui, and B. Kerem. 2000. "Replication Delay along FRA7H, a Common Fragile Site on Human Chromosome 7, Leads to Chromosomal Instability." *Molecular and Cellular Biology* 20 (12): 4420–27. doi:10.1128/MCB.20.12.4420-4427.2000.
Helmrich, Anne, Monica Ballarino, Evgeny Nudler, and Laszlo Tora. 2013. "Transcription-Replication Encounters, Consequences and Genomic Instability." *Nature Structural & Molecular Biology* 20 (4): 412–18. doi:10.1038/nsmb.2543.
Helmrich, Anne, Monica Ballarino, and Laszlo Tora. 2011. "Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes." *Molecular Cell* 44 (6): 966–77. http://www.sciencedirect.com/science/article/pii/S1097276511008896.
Herrera-Moyano, Emilia, Xénia Mergui, María L García-Rubio, Sonia Barroso, and Andrés Aguilera. 2014. "The Yeast and Human FACT Chromatin-Reorganizing Complexes Solve R-Loop-Mediated Transcription-Replication Conflicts." *Genes & Development* 28 (7): 735–48. doi:10.1101/gad.234070.113.
Herrero, Ana B, Antonio Marı, and Sergio Moreno. 2013. "The Npl3 HnRNP Prevents R-Loop-Mediated Transcription – Replication Conflicts and Genome Instability," 2445–58. doi:10.1101/gad.229880.113.associated.
Hewett, D R, O Handt, L Hobson, M Mangelsdorf, H J Eyre, E Baker, G R Sutherland, S Schuffenhauer, J I Mao, and R I Richards. 1998. "FRA10B Structure Reveals Common Elements in Repeat Expansion and Chromosomal Fragile Site Genesis." *Molecular Cell* 1 (6): 773–81. http://www.ncbi.nlm.nih.gov/pubmed/9660961.
Hills, Stephanie A. a, and John F.X. F X Diffley. 2014. "DNA Replication and Oncogene-Induced Replicative Stress." *Current Biology : CB* 24 (10). Elsevier: R435-44. doi:10.1016/j.cub.2014.04.012.
Hindson, Benjamin J., Kevin D. Ness, Donald A. Masquelier, Phillip Belgrader, Nicholas J. Heredia, Anthony J. Makarewicz, Isaac J. Bright, et al. 2011. "High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number." *Analytical Chemistry* 83 (22). American Chemical Society: 8604–10. doi:10.1021/ac202028g.
Hiratani, Ichiro, Tyrone Ryba, Mari Itoh, Tomoki Yokochi, Michaela Schwaiger, Chia-Wei Chang, Yung Lyou, Tim M Townes, Dirk Schübeler, and David M Gilbert. 2008. "Global Reorganization of Replication Domains during Embryonic Stem Cell Differentiation." *PLoS Biology* 6 (10): e245. doi:10.1371/journal.pbio.0060245.

Houtsmuller, A B, S Rademakers, A L Nigg, D Hoogstraten, J H Hoeijmakers, and W Vermeulen. 1999. “Action of DNA Repair Endonuclease ERCC1/XPF in Living Cells.” *Science (New York, N.Y.)* 284 (5416): 958–61. http://www.ncbi.nlm.nih.gov/pubmed/10320375.

Hraiky, Chadi, Marc André Raymond, and Marc Drolet. 2000. “RNase H Overproduction Corrects a Defect at the Level of Transcription Elongation during RRNA Synthesis in the Absence of DNA Topoisomerase I in Escherichia Coli.” *Journal of Biological Chemistry* 275 (15): 11257–63. doi:10.1074/jbc.275.15.11257.

Hu, Zheng, Da Zhu, Wei Wang, Weiyang Li, Wenlong Jia, Xi Zeng, Wencheng Ding, et al. 2015. “Genome-Wide Profiling of HPV Integration in Cervical Cancer Identifies Clustered Genomic Hot Spots and a Potential Microhomology-Mediated Integration Mechanism.” *Nature Genetics* 47 (2): 158–63. doi:10.1038/ng.3178.

Huertas, Pablo, and Andrés Aguilera. 2003. “Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination.” *Molecular Cell* 12 (3): 711–21. doi:10.1016/j.molcel.2003.08.010.

Hwang, D S, and A Kornberg. 1992. “Opening of the Replication Origin of Escherichia Coli by DnaA Protein with Protein HU or IHF.” *The Journal of Biological Chemistry* 267 (32): 23083–86. http://www.ncbi.nlm.nih.gov/pubmed/1429655.

Hyrien, Olivier. 2016. “How MCM Loading and Spreading Specify Eukaryotic DNA Replication Initiation Sites.” *F1000Research* 5. Faculty of 1000 Ltd. doi:10.12688/f1000research.9008.1.

Ibarra, Arkaitz, Etienne Schwob, and Juan Méndez. 2008. “Excess MCM Proteins Protect Human Cells from Replicative Stress by Licensing Backup Origins of Replication.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (26). National Academy of Sciences: 8956–61. doi:10.1073/pnas.0803978105.

IKEGAMI, SUSUMU, TAKAHIKO TAGUCHI, MOCHIHIKO OHASHI, MIEKO OGURO, HIROSHI NAGANO, and YOSHITAKE MANO. 1978. “Aphidicolin Prevents Mitotic Cell Division by Interfering with the Activity of DNA Polymerase-α.” *Nature* 275 (5679). Nature Publishing Group: 458–60. doi:10.1038/275458a0.

Iliopoulos, Dimitrios, Gulnur Guler, Shuang-Yin Han, Teresa Druck, Michelle Ottey, Kelly A. McCorkell, and Kay Huebner. 2006. “Roles of FHIT and WWOX Fragile Genes in Cancer.” *Cancer Letters* 232 (1): 27–36. doi:10.1016/j.canlet.2005.06.048.

Inoue, Ken, Hitoshi Osaka, Virginia C Thurston, Joe T R Clarke, Akira Yoneyama, Lisa Rosenbarker, Thomas D Bird, M E Hodes, Lisa G Shaffer, and James R Lupski. 2002. “Genomic Rearrangements Resulting in PLP1 Deletion Occur by Nonhomologous End Joining and Cause Different Dysmyelinating Phenotypes in Males and Females.” *American Journal of Human Genetics* 71 (4). Elsevier: 838–53. doi:10.1086/342728.

Ira, Grzegorz, and James E Haber. 2002. “Characterization of RAD51-Independent Break-Induced Replication That Acts Preferentially with Short Homologous Sequences.” *Molecular and Cellular Biology* 22 (18). American Society for Microbiology (ASM): 6384–92. doi:10.1128/MCB.22.18.6384-6392.2002.

Itoh, T, and J Tomizawa. 1980. “Formation of an RNA Primer for Initiation of Replication of ColE1 DNA by Ribonuclease H.” *Proceedings of the National Academy of Sciences of the United States of America* 77 (5): 2450–54. http://www.ncbi.nlm.nih.gov/pubmed/6156450.

Jones, R M, O Mortusewicz, I Afzal, M Lorvellec, P García, T Helleday, and E Petermann. 2013. “Increased Replication Initiation and Conflicts with Transcription Underlie Cyclin E-

Induced Replication Stress." *Oncogene* 32 (32): 3744–53. doi:10.1038/onc.2012.387.
Joo, Young Hoon, Sung Won Park, Seung Hyun Jung, Yeon Soo Lee, In Chul Nam, Kwang Jae Cho, Jun Ook Park, Yeun Jun Chung, and Min Sik Kim. 2013. "Recurrent Loss of the FHIT Gene and Its Impact on Lymphatic Metastasis in Early Oral Squamous Cell Carcinoma." *Acta Oto-Laryngologica* 133 (9): 992–99. doi:10.3109/00016489.2013.795289.
Jossen, Rachel, and Rodrigo Bermejo. 2013. "The DNA Damage Checkpoint Response to Replication Stress: A Game of Forks." *Frontiers in Genetics* 4 (March). Frontiers: 26. doi:10.3389/fgene.2013.00026.
Karras, Jenna R., Morgan S. Schrock, Bahadir Batar, Jie Zhang, Krista La Perle, Teresa Druck, and Kay Huebner. 2016. "Fhit Loss-Associated Initiation and Progression of Neoplasia in Vitro." *Cancer Science* 107 (11): 1590–98. doi:10.1111/cas.13032.
Kasai, Fumio, Noriko Hirayama, Midori Ozawa, Masashi Iemura, and Arihiro Kohara. 2016. "Changes of Heterogeneous Cell Populations in the Ishikawa Cell Line during Long-Term Culture: Proposal for an in Vitro Clonal Evolution Model of Tumor Cells." *Genomics* 107 (6). Academic Press: 259–66. doi:10.1016/J.YGENO.2016.04.003.
Kim, Nayun, and Sue Jinks-Robertson. 2012. "Transcription as a Source of Genome Instability." *Nature Reviews. Genetics* 13 (3). Nature Publishing Group: 204–14. doi:10.1038/nrg3152.
Kim, Philip M, Hugo Y K Lam, Alexander E Urban, Jan O Korbel, Jason Affourtit, Fabian Grubert, Xueying Chen, Sherman Weissman, Michael Snyder, and Mark B Gerstein. 2008. "Analysis of Copy Number Variants and Segmental Duplications in the Human Genome: Evidence for a Change in the Process of Formation in Recent Evolutionary History." *Genome Research* 18 (12). Cold Spring Harbor Laboratory Press: 1865–74. doi:10.1101/gr.081422.108.
Köhler, Carsten, Dennis Koalick, Anja Fabricius, Ann Christin Parplys, Kerstin Borgmann, Helmut Pospiech, and Frank Grosse. 2016. "Cdc45 Is Limiting for Replication Initiation in Humans." *Cell Cycle* 15 (7). Taylor & Francis: 974–85. doi:10.1080/15384101.2016.1152424.
Korbel, J. O., A. E. Urban, J. P. Affourtit, B. Godwin, F. Grubert, J. F. Simons, P. M. Kim, et al. 2007. "Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome." *Science* 318 (5849): 420–26. doi:10.1126/science.1149504.
Kosicki, Michael, Kärt Tomberg, and Allan Bradley. 2018. "Repair of Double-Strand Breaks Induced by CRISPR–Cas9 Leads to Large Deletions and Complex Rearrangements." *Nature Biotechnology* 36 (8). Nature Publishing Group: 765. doi:10.1038/nbt.4192.
Kotsantis, Panagiotis, Lara Marques Silva, Sarah Irmscher, Rebecca M. Jones, Lisa Folkes, Natalia Gromak, and Eva Petermann. 2016. "Increased Global Transcription Activity as a Mechanism of Replication Stress in Cancer." *Nature Communications* 7 (October). Nature Publishing Group: 13087. doi:10.1038/ncomms13087.
Kremer, E. J., M. Pritchard, M. Lynch, S. Yu, K. Holman, E. Baker, S. T. Warren, D. Schlessinger, G. R. Sutherland, and R. I. Richards. 1991. "Mapping of DNA Instability at the Fragile X to a Trinucleotide Repeat Sequence p(CCG)N." *Science* 252 (5013): 1711–14. doi:10.1126/science.1675488.
Krishnamoorthy, K., and Jessica Thomson. 2002. "A More Powerful Test for Comparing Two Poisson Means." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3142708.
Kuroki, Tamotsu, Francesco Trapasso, Takeshi Shiraishi, Hansjuerg Alder, Koshi Mimori, Masaki Mori, and Carlo M Croce. 2002. "Genetic Alterations of the Tumor Suppressor Gene WWOX in Esophageal Squamous Cell Carcinoma." *Cancer Research* 62 (8): 2258–

60. http://www.ncbi.nlm.nih.gov/pubmed/11956080.
Labun, Kornel, Tessa G. Montague, James A. Gagnon, Summer B. Thyme, and Eivind Valen. 2016. “CHOPCHOP v2: A Web Tool for the next Generation of CRISPR Genome Engineering.” *Nucleic Acids Research* 44 (W1): W272–76. doi:10.1093/nar/gkw398.
Lai, Lisa A., Rumen Kostadinov, Michael T. Barrett, Daniel A. Peiffer, Dimitry Pokholok, Robert Odze, Carissa A. Sanchez, et al. 2010. “Deletion at Fragile Sites Is a Common and Early Event in Barrett’s Esophagus.” *Molecular Cancer Research : MCR* 8 (8): 1084–94. doi:10.1158/1541-7786.MCR-09-0529.
Landgraf, Ralf, Chi-Hong B Chen, and David S Sigman. 1995. “R-Loop Stability as a Function of RNA Structure and Size.” *Nucleic Acids Research*. Vol. 23. https://pdfs.semanticscholar.org/b37a/2dac7d0500f9619e83573d0c08382247703f.pdf.
Lavin, Martin F, Abrey J Yeo, and Olivier J Becherel. 2013. “Senataxin Protects the Genome: Implications for Neurodegeneration and Other Abnormalities.” *Rare Diseases (Austin, Tex.)* 1. Taylor & Francis: e25230. doi:10.4161/rdis.25230.
Le Beau, M. 1998. “Replication of a Common Fragile Site, FRA3B, Occurs Late in S Phase and Is Delayed Further upon Induction: Implications for the Mechanism of Fragile Site Induction.” *Human Molecular Genetics* 7 (4): 755–61. doi:10.1093/hmg/7.4.755.
Le Beau, M M, H Drabkin, T W Glover, R Gemmill, F V Rassool, T W McKeithan, and D I Smith. 1998. “An FHIT Tumor Suppressor Gene?” *Genes, Chromosomes & Cancer* 21 (4): 281–89. http://www.ncbi.nlm.nih.gov/pubmed/9559339.
Le Tallec, Benoît, Gaël Armel Millot, Marion Esther Blin, Olivier Brison, Bernard Dutrillaux, and Michelle Debatisse. 2013. “Common Fragile Site Profiling in Epithelial and Erythroid Cells Reveals That Most Recurrent Cancer Deletions Lie in Fragile Sites Hosting Large Genes.” *Cell Reports*, July, 1–9. doi:10.1016/j.celrep.2013.07.003.
Lee-Kirsch, M A, C Wolf, and C Günther. 2014. “Aicardi-Goutières Syndrome: A Model Disease for Systemic Autoimmunity.” *Clinical and Experimental Immunology* 175 (1). Wiley-Blackwell: 17–24. doi:10.1111/cei.12160.
Lee, Jennifer a, Claudia M B Carvalho, and James R Lupski. 2007. “A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders.” *Cell* 131 (7): 1235–47. doi:10.1016/j.cell.2007.11.037.
Lei, M, and B K Tye. 2001. “Initiating DNA Synthesis: From Recruiting to Activating the MCM Complex.” *Journal of Cell Science* 114 (Pt 8): 1447–54. http://www.ncbi.nlm.nih.gov/pubmed/11282021.
Lesnik, Elena A., and Susan M. Freier. 1995. “Relative Thermodynamic Stability of DNA, RNA, and DNA:RNA Hybrid Duplexes: Relationship with Base Composition and Structure.” *Biochemistry* 34 (34). American Chemical Society: 10807–15. doi:10.1021/bi00034a013.
Letessier, Anne, Gaël A. Millot, Stéphane Koundrioukoff, Anne-Marie Lachagès, Nicolas Vogt, R. Scott Hansen, Bernard Malfoy, Olivier Brison, and Michelle Debatisse. 2011. “Cell-Type-Specific Replication Initiation Programs Set Fragility of the FRA3B Fragile Site.” *Nature* 470 (7332): 120–23. doi:10.1038/nature09745.
Li, Joachim J. 1995. “DNA Replication: Once, and Only Once.” *Current Biology* 5 (5). Cell Press: 472–75. doi:10.1016/S0960-9822(95)00094-7.
Lima, Walt F, Heather M Murray, Sagar S Damle, Christopher E Hart, Gene Hung, Cheryl Li De Hoyos, Xue-Hai Liang, and Stanley T Crooke. 2016. “Viable RNaseH1 Knockout Mice Show RNaseH1 Is Essential for R Loop Processing, Mitochondrial and Liver Function.” *Nucleic Acids Research* 44 (11). Oxford University Press: 5299–5312.

doi:10.1093/nar/gkw350.
Lõoke, Marko, Jüri Reimand, Tiina Sedman, Juhan Sedman, Lari Järvinen, Signe Värv, Kadri Peil, Kersti Kristjuhan, Jaak Vilo, and Arnold Kristjuhan. 2010. "Relicensing of Transcriptionally Inactivated Replication Origins in Budding Yeast." *The Journal of Biological Chemistry* 285 (51): 40004–11. doi:10.1074/jbc.M110.148924.
Ma, Jia-Lin, Eun Mi Kim, James E Haber, and Sang Eun Lee. 2003. "Yeast Mre11 and Rad1 Proteins Define a Ku-Independent Mechanism to Repair Double-Strand Breaks Lacking Overlapping End Sequences." *Molecular and Cellular Biology* 23 (23): 8820–28. http://www.ncbi.nlm.nih.gov/pubmed/14612421.
Machida, Yuichi J., Joyce L. Hamlin, and Anindya Dutta. 2005. "Right Place, Right Time, and Only Once: Replication Initiation in Metazoans." *Cell* 123 (1): 13–24. doi:10.1016/j.cell.2005.09.019.
Maestrini, E., A. T. Pagnamenta, J. A. Lamb, E. Bacchelli, N. H. Sykes, I. Sousa, C. Toma, et al. 2010. "High-Density SNP Association Study and Copy Number Variation Analysis of the AUTS1 and AUTS5 Loci Implicate the IMMP2L-DOCK4 Gene Region in Autism Susceptibility." *Molecular Psychiatry* 15 (9): 954–68. doi:10.1038/mp.2009.34.
Mai, Sabine, Stefan Imreh, David I. Smith, Sarah McAvoy, Yu Zhu, and Damon S. Perez. 2007. "Large Common Fragile Site Genes and Cancer." *Seminars in Cancer Biology* 17 (1): 31–41. doi:10.1016/j.semcancer.2006.10.003.
Malkova, Anna, and Grzegorz Ira. 2013. "Break-Induced Replication: Functions and Molecular Mechanism." *Current Opinion in Genetics & Development* 23 (3). NIH Public Access: 271–79. doi:10.1016/j.gde.2013.05.007.
Mantiero, Davide, Amanda Mackenzie, Anne Donaldson, and Philip Zegerman. 2011. "Limiting Replication Initiation Factors Execute the Temporal Programme of Origin Firing in Budding Yeast." *The EMBO Journal* 30 (23). Nature Publishing Group: 4805–14. doi:10.1038/emboj.2011.404.
McGuffee, Sean R., Duncan J. Smith, and Iestyn Whitehouse. 2013. "Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination." *Molecular Cell* 50 (1): 123–35. doi:10.1016/j.molcel.2013.03.004.
McIntosh, D., and J. J. Blow. 2012. "Dormant Origins, the Licensing Checkpoint, and the Response to Replicative Stresses." *Cold Spring Harbor Perspectives in Biology* 4 (10): a012955–a012955. doi:10.1101/cshperspect.a012955.
McIntosh, Debbie, and J. Julian Blow. 2012. "Dormant Origins, the Licensing Checkpoint, and the Response to Replicative Stresses." *Cold Spring Harbor Perspectives in Biology* 4 (10). Cold Spring Harbor Laboratory Press: a012955–a012955. doi:10.1101/cshperspect.a012955.
McVey, Mitch, and Sang Eun Lee. 2008. "MMEJ Repair of Double-Strand Breaks (Director's Cut): Deleted Sequences and Alternative Endings." *Trends in Genetics : TIG* 24 (11). Elsevier: 529–38. doi:10.1016/j.tig.2008.08.007.
Mefford, Heather C., Hiltrud Muhle, Philipp Ostertag, Sarah von Spiczak, Karen Buysse, Carl Baker, Andre Franke, et al. 2010. "Genome-Wide Copy Number Variation in Epilepsy: Novel Susceptibility Loci in Idiopathic Generalized and Focal Epilepsies." *PLoS Genetics* 6 (5). doi:10.1371/journal.pgen.1000962.
Mehta, Anuja, Annette Beach, and James E Haber. 2017. "Homology Requirements and Competition between Gene Conversion and Break-Induced Replication during Double-Strand Break Repair." *Molecular Cell* 65 (3). NIH Public Access: 515–526.e3.

doi:10.1016/j.molcel.2016.12.003.
Mikhail, Fady M., Edward J. Lose, Nathaniel H. Robin, Maria D. Descartes, Katherine D. Rutledge, S. Lane Rutledge, Bruce R. Korf, and Andrew J. Carroll. 2011. "Clinically Relevant Single Gene or Intragenic Deletions Encompassing Critical Neurodevelopmental Genes in Patients with Developmental Delay, Mental Retardation, and/or Autism Spectrum Disorders." *American Journal of Medical Genetics, Part A* 155 (10): 2386–96. doi:10.1002/ajmg.a.34177.
Mills, Ryan E., Klaudia Walter, Chip Stewart, Robert E. Handsaker, Ken Chen, Can Alkan, Alexej Abyzov, et al. 2011. "Mapping Copy Number Variation by Population-Scale Genome Sequencing." *Nature* 470 (7332): 59–65. doi:10.1038/nature09708.
Minocherhomji, Sheroy, Songmin Ying, Victoria A. Bjerregaard, Sara Bursomanno, Aiste Aleliunaite, Wei Wu, Hocine W. Mankouri, Huahao Shen, Ying Liu, and Ian D. Hickson. 2015. "Replication Stress Activates DNA Repair Synthesis in Mitosis." *Nature* 528 (7581): 286–90. doi:10.1038/nature16139.
Mirkin, Ekaterina V, and Sergei M Mirkin. 2005a. "Mechanisms of Transcription-Replication Collisions in Bacteria Mechanisms of Transcription-Replication Collisions in Bacteria" 25 (3). doi:10.1128/MCB.25.3.888.
———. 2005b. "Mechanisms of Transcription-Replication Collisions in Bacteria." *Molecular and Cellular Biology* 25 (3): 888–95. doi:10.1128/MCB.25.3.888-895.2005.
———. 2007. "Replication Fork Stalling at Natural Impediments." *Microbiology and Molecular Biology Reviews : MMBR* 71 (1). American Society for Microbiology (ASM): 13–35. doi:10.1128/MMBR.00030-06.
Miron, Karin, Tamar Golan-Lev, Raz Dvir, Eyal Ben-David, and Batsheva Kerem. 2015. "Oncogenes Create a Unique Landscape of Fragile Sites." *Nature Communications* 6 (1): 7094. doi:10.1038/ncomms8094.
Mishmar, D, A Rahat, S W Scherer, G Nyakatura, B Hinzmann, Y Kohwi, Y Mandel-Gutfroind, et al. 1998. "Molecular Characterization of a Common Fragile Site (FRA7H) on Human Chromosome 7 by the Cloning of a Simian Virus 40 Integration Site." *Proceedings of the National Academy of Sciences of the United States of America* 95 (14). National Academy of Sciences: 8141–46. doi:10.1073/PNAS.95.14.8141.
Miyawaki, Yutaka, Hiroshi Kawachi, Akishi Ooi, Yoshinobu Eishi, Tatsuyuki Kawano, Johji Inazawa, and Issei Imoto. 2012. "Genomic Copy-Number Alterations of MYC and FHIT Genes Are Associated with Survival in Esophageal Squamous-Cell Carcinoma." *Cancer Science* 103 (8): 1558–66. doi:10.1111/j.1349-7006.2012.02329.x.
Molès, Jean-Pierre, Anthony Griez, Jean-Jacques Guilhou, Céline Girard, Nicolas Nagot, Philippe Van de Perre, and Pierre Dujols. 2017. "Cytosolic RNA:DNA Duplexes Generated by Endogenous Reverse Transcriptase Activity as Autonomous Inducers of Skin Inflammation in Psoriasis." *PloS One* 12 (1). Public Library of Science: e0169879. doi:10.1371/journal.pone.0169879.
Mone, M. J., T. Bernas, C. Dinant, F. A. Goedvree, E. M. M. Manders, M. Volker, A. B. Houtsmuller, J. H. J. Hoeijmakers, W. Vermeulen, and R. van Driel. 2004. "In Vivo Dynamics of Chromatin-Associated Complex Formation in Mammalian Nucleotide Excision Repair." *Proceedings of the National Academy of Sciences* 101 (45): 15933–37. doi:10.1073/pnas.0403664101.
Montague, Tessa G., José M. Cruz, James A. Gagnon, George M. Church, and Eivind Valen. 2014. "CHOPCHOP: A CRISPR/Cas9 and TALEN Web Tool for Genome Editing."

Nucleic Acids Research 42 (W1). doi:10.1093/nar/gku410.
Muff, Roman, Prisni Rath, Ram Mohan Ram Kumar, Knut Husmann, Walter Born, Michael Baudis, and Bruno Fuchs. 2015. "Genomic Instability of Osteosarcoma Cell Lines in Culture: Impact on the Prediction of Metastasis Relevant Genes." *PloS One* 10 (5). Public Library of Science: e0125611. doi:10.1371/journal.pone.0125611.
Muramatsu, M, K Kinoshita, S Fagarasan, S Yamada, Y Shinkai, and T Honjo. 2000. "Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme." *Cell* 102 (5): 553–63. http://www.ncbi.nlm.nih.gov/pubmed/11007474.
Murano, I., A. Kuwano, and T. Kajii. 1989. "Fibroblast-Specific Common Fragile Sites Induced by Aphidicolin." *Human Genetics* 83 (1). Springer-Verlag: 45–48. doi:10.1007/BF00274145.
Nagamani, Sandesh C.S. S, Ayelet Erez, Bruria Ben-Zeev, Moshe Frydman, Susan Winter, Robert Zeller, Dima El-Khechen, et al. 2013. "Detection of Copy-Number Variation in AUTS2 Gene by Targeted Exonic Array CGH in Patients with Developmental Delay and Autistic Spectrum Disorders." *European Journal of Human Genetics : EJHG* 21 (3). Macmillan Publishers Limited: 343–46. doi:10.1038/ejhg.2012.157.
Naito, Yuki, Kimihiro Hino, Hidemasa Bono, and Kumiko Ui-Tei. 2015. "CRISPRdirect: Software for Designing CRISPR/Cas Guide RNA with Reduced off-Target Sites." *Bioinformatics* 31 (7): 1120–23. doi:10.1093/bioinformatics/btu743.
Newman, Timothy J., Mohammed A. Mamun, Conrad A. Nieduszynski, and J. Julian Blow. 2013. "Replisome Stall Events Have Shaped the Distribution of Replication Origins in the Genomes of Yeasts." *Nucleic Acids Research* 41 (21): 9705–18. doi:10.1093/nar/gkt728.
Ning, Y., M. Lovell, L. Taylor, and O.M. Pereira-Smith. 1992. "Isolation of Monochromosomal Hybrids Following Fusion of Human Diploid Fibroblast-Derived Microcells with Mouse A9 Cells." *Cytogenetic and Genome Research* 60 (1): 79–80. doi:10.1159/000133300.
Nishitani, Hideo, and Zoi Lygerou. 2018. "Control of DNA Replication Licensing in a Cell Cycle." Accessed August 21. https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2443.2002.00544.x.
Oberlé, I., F. Rousseau, D. Heitz, C. Kretz, D. Devys, A. Hanauer, J. Boué, M. F. Bertheas, and J. L. Mandel. 1991. "Instability of a 550-Base Pair DNA Segment and Abnormal Methylation in Fragile X Syndrome." *Science* 252 (5009): 1097–1102. doi:10.1126/science.252.5009.1097.
Orth, Kim, Jaclyn Hungt, Adi Gazdart, Anne Bowcock, J Michael Mathist, and Joseph Sambrookt. 1994. "Genetic Instability in Human Ovarian Cancer Cell Lines." *Proc. Natd. Acad. Sci. USA*. Vol. 91. http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC44839&blobtype=pdf.
Paige, A J, K J Taylor, C Taylor, S G Hillier, S Farrington, D Scott, D J Porteous, J F Smyth, H Gabra, and J E Watson. 2001. "WWOX: A Candidate Tumor Suppressor Gene Involved in Multiple Tumor Types." *Proceedings of the National Academy of Sciences of the United States of America* 98 (20). National Academy of Sciences: 11417–22. doi:10.1073/pnas.191175898.
Palakodeti, Aparna, Yu Han, Yanwen Jiang, and Michelle M Le Beau. 2004. "The Role of Late/Slow Replication of the FRA16D in Common Fragile Site Induction." *Genes, Chromosomes & Cancer* 39 (1): 71–76. doi:10.1002/gcc.10290.
Palumbo, Elisa, Laura Matricardi, Elena Tosoni, Aaron Bensimon, and Antonella Russo. 2010.

"Replication Dynamics at Common Fragile Site FRA6E." *Chromosoma* 119 (6): 575–87. doi:10.1007/s00412-010-0279-4.

Pâques, F, and J E Haber. 1999. "Multiple Pathways of Recombination Induced by Double-Strand Breaks in Saccharomyces Cerevisiae." *Microbiology and Molecular Biology Reviews : MMBR* 63 (2): 349–404. http://www.ncbi.nlm.nih.gov/pubmed/10357855.

Paulsen, Michelle T., Artur Veloso, Jayendra Prasad, Karan Bedi, Emily A. Ljungman, Brian Magnuson, Thomas E. Wilson, and Mats Ljungman. 2014. "Use of Bru-Seq and BruChase-Seq for Genome-Wide Assessment of the Synthesis and Stability of RNA." *Methods* 67 (1): 45–54. doi:10.1016/j.ymeth.2013.08.015.

Paulsen, Michelle T, Artur Veloso, Jayendra Prasad, Karan Bedi, Emily A Ljungman, Ya-Chun Tsan, Ching-Wei Chang, et al. 2013. "Coordinated Regulation of Synthesis and Stability of RNA during the Acute TNF-Induced Proinflammatory Response." *Proceedings of the National Academy of Sciences of the United States of America* 110 (6): 2240–45. doi:10.1073/pnas.1219192110.

Pelliccia, F., N. Bosco, A. Curatolo, and A. Rocchi. 2008. "Replication Timing of Two Human Common Fragile Sites: FRA1H and FRA2G." *Cytogenetic and Genome Research* 121 (3–4): 196–200. doi:10.1159/000138885.

Pérez-Ortín, José E, Paula M Alepuz, and Joaquín Moreno. 2007. "Genomics and Gene Transcription Kinetics in Yeast." *Trends in Genetics : TIG* 23 (5): 250–57. doi:10.1016/j.tig.2007.03.006.

Petersen-Mahrt, Svend K., Reuben S. Harris, and Michael S. Neuberger. 2002. "AID Mutates E. Coli Suggesting a DNA Deamination Mechanism for Antibody Diversification." *Nature* 418 (6893). Nature Publishing Group: 99–104. doi:10.1038/nature00862.

Petryk, Nataliya, Malik Kahli, Yves d'Aubenton-Carafa, Yan Jaszczyszyn, Yimin Shen, Maud Silvain, Claude Thermes, Chun-Long Chen, and Olivier Hyrien. 2016. "Replication Landscape of the Human Genome." *Nature Communications* 7 (January). Nature Publishing Group: 10208. doi:10.1038/ncomms10208.

Pfister, Sophia X, Enni Markkanen, Yanyan Jiang, Sovan Sarkar, Mick Woodcock, Giulia Orlando, Ioanna Mavrommati, et al. 2015. "Inhibiting WEE1 Selectively Kills Histone H3K36me3-Deficient Cancers by DNTP Starvation." *Cancer Cell* 28 (5). Elsevier: 557–68. doi:10.1016/j.ccell.2015.09.015.

Pham, Phuong, Ronda Bransteitter, John Petruska, and Myron F. Goodman. 2003. "Processive AID-Catalysed Cytosine Deamination on Single-Stranded DNA Simulates Somatic Hypermutation." *Nature* 424 (6944): 103–7. doi:10.1038/nature01760.

Pizzi, Sara, Sarah Sertic, Simona Orcesi, Cristina Cereda, Marika Bianchi, Andrew P. Jackson, Federico Lazzaro, Paolo Plevani, and Marco Muzi-Falconi. 2015. "Reduction of HRNase H2 Activity in Aicardi–Goutières Syndrome Cells Leads to Replication Stress and Genome Instability." *Human Molecular Genetics* 24 (3): 649–58. doi:10.1093/hmg/ddu485.

Pohjoismäki, Jaakko L.O., J. Bradley Holmes, Stuart R. Wood, Ming Yao Yang, Takehiro Yasukawa, Aurelio Reyes, Laura J. Bailey, et al. 2010. "Mammalian Mitochondrial DNA Replication Intermediates Are Essentially Duplex but Contain Extensive Tracts of RNA/DNA Hybrid." *Journal of Molecular Biology* 397 (5): 1144–55. doi:10.1016/j.jmb.2010.02.029.

Poli, Jérôme, Olga Tsaponina, Laure Crabbé, Andrea Keszthelyi, Véronique Pantesco, Andrei Chabes, Armelle Lengronne, and Philippe Pasero. 2012. "DNTP Pools Determine Fork Progression and Origin Usage under Replication Stress." *The EMBO Journal* 31 (4): 883–

94. doi:10.1038/emboj.2011.470.
Pollok, S., C. Bauerschmidt, J. Sänger, H.-P. Nasheuer, and F. Grosse. 2007. "Human Cdc45 Is a Proliferation-Associated Antigen." *The FEBS Journal* 274 (14). Wiley/Blackwell (10.1111): 3669–84. doi:10.1111/j.1742-4658.2007.05900.x.
Pomerantz, Richard T., and Mike O'Donnell. 2010a. "Direct Restart of a Replication Fork Stalled by a Head-on RNA Polymerase." *Science* 327 (5965): 590–92. doi:10.1126/science.1179595.
Pomerantz, Richard T, and Mike O'Donnell. 2010b. "What Happens When Replication and Transcription Complexes Collide?" *Cell Cycle (Georgetown, Tex.)* 9 (13). Taylor & Francis: 2537–43. doi:10.4161/cc.9.13.12122.
Postow, L., N. J. Crisona, B. J. Peter, C. D. Hardy, and N. R. Cozzarelli. 2001. "Topological Challenges to DNA Replication: Conformations at the Fork." *Proceedings of the National Academy of Sciences* 98 (15): 8219–26. doi:10.1073/pnas.111006998.
Postow, Lisa, Chris Ullsperger, Rebecca W. Keller, Carlos Bustamante, Alexander V. Vologodskii, and Nicholas R. Cozzarelli. 2001. "Positive Torsional Strain Causes the Formation of a Four-Way Junction at Replication Forks." *Journal of Biological Chemistry* 276 (4): 2790–96. doi:10.1074/jbc.M006736200.
Prado, Félix, and Andrés Aguilera. 2005. "Impairment of Replication Fork Progression Mediates RNA PolII Transcription-Associated Recombination." *The EMBO Journal* 24 (6). EMBO Press: 1267–76. doi:10.1038/sj.emboj.7600602.
Rajaram, Megha, Jianping Zhang, Tim Wang, Jinyu Li, Cem Kuscu, Huan Qi, Mamoru Kato, et al. 2013. "Two Distinct Categories of Focal Deletions in Cancer Genomes." *PLoS ONE* 8 (6). doi:10.1371/journal.pone.0066264.
Ray Chaudhuri, Arnab, Yoshitami Hashimoto, Raquel Herrador, Kai J Neelsen, Daniele Fachinetti, Rodrigo Bermejo, Andrea Cocito, Vincenzo Costanzo, and Massimo Lopes. 2012. "Topoisomerase I Poisoning Results in PARP-Mediated Replication Fork Reversal." *Nature Structural & Molecular Biology* 19 (4): 417–23. doi:10.1038/nsmb.2258.
Reaban, Mary E., and Johanna A. Griffin. 1990. "Induction of RNA-Stabilized DMA Conformers by Transcription of an Immunoglobulin Switch Region." *Nature* 348 (6299): 342–44. doi:10.1038/348342a0.
Revy, P, T Muto, Y Levy, F Geissmann, A Plebani, O Sanal, N Catalan, et al. 2000. "Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2)." *Cell* 102 (5): 565–75. http://www.ncbi.nlm.nih.gov/pubmed/11007475.
Roberts, R. W, and D. M Crothers. 1992. "Stability and Properties of Double and Triple Helices: Dramatic Effects of RNA or DNA Backbone Composition." *Science* 258 (5087): 1463–66. doi:10.1126/science.1279808.
Saldivar, Joshua C., Satoshi Miuma, Jessica Bene, Seyed Ali Hosseini, Hidetaka Shibata, Jin Sun, Linda J. Wheeler, Christopher K. Mathews, and Kay Huebner. 2012. "Initiation of Genome Instability and Preneoplastic Processes through Loss of Fhit Expression." Edited by Marshall S. Horwitz. *PLoS Genetics* 8 (11): e1003077. doi:10.1371/journal.pgen.1003077.
Saldivar, Joshua C, Hidetaka Shibata, and Kay Huebner. 2010. "Pathology and Biology Associated with the Fragile FHIT Gene and Gene Product." *Journal of Cellular Biochemistry* 109 (5): 858–65. doi:10.1002/jcb.22481.
Sankar, Natesan, Ravi-Kumar Kadeppagari, and Bayar Thimmapaya. 2009. "C-Myc-Induced

Aberrant DNA Synthesis and Activation of DNA Damage Response in P300 Knockdown Cells." *The Journal of Biological Chemistry* 284 (22). American Society for Biochemistry and Molecular Biology: 15193–205. doi:10.1074/jbc.M900776200.

Sanz, Lionel A, Stella R Hartono, Yoong Wearn Lim, Sandra Steyaert, Aparna Rajpurkar, Paul A Ginno, Xiaoqin Xu, and Frédéric Chédin. 2016. "Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals." *Molecular Cell* 63 (1). NIH Public Access: 167–78. doi:10.1016/j.molcel.2016.05.032.

Sarni, Dan, and Batsheva Kerem. 2017. "Oncogene-Induced Replication Stress Drives Genome Instability and Tumorigenesis." *International Journal of Molecular Sciences* 18 (7). Multidisciplinary Digital Publishing Institute (MDPI): 1339. doi:10.3390/ijms18071339.

Schwartz, Michal, Eitan Zlotorynski, and Batsheva Kerem. 2006. "The Molecular Basis of Common and Rare Fragile Sites." *Cancer Letters* 232 (1). Elsevier: 13–26. doi:10.1016/J.CANLET.2005.07.039.

Sclafani, R A, and T M Holzen. 2007. "Cell Cycle Regulation of DNA Replication." *Annual Review of Genetics* 41. NIH Public Access: 237–80. doi:10.1146/annurev.genet.41.110306.130308.

Sfeir, Agnel, and Lorraine S Symington. 2015. "Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?" *Trends in Biochemical Sciences* 40: 701–14. doi:10.1016/j.tibs.2015.08.006.

Shah, Sandeep N, Patricia L Opresko, Xiao Meng, Marietta Y W T Lee, and Kristin A Eckert. 2010. "DNA Structure and the Werner Protein Modulate Human DNA Polymerase Delta-Dependent Replication Dynamics within the Common Fragile Site FRA16D." *Nucleic Acids Research* 38 (4). Oxford University Press: 1149–62. doi:10.1093/nar/gkp1131.

Sharp, Andrew J., Devin P. Locke, Sean D. McGrath, Ze Cheng, Jeffrey A. Bailey, Rhea U. Vallente, Lisa M. Pertz, et al. 2005. "Segmental Duplications and Copy-Number Variation in the Human Genome." *The American Journal of Human Genetics* 77 (1): 78–88. doi:10.1086/431652.

Shen, Wen, Hong Sun, Cheryl L De Hoyos, Jeffrey K Bailey, Xue-Hai Liang, and Stanley T Crooke. 2017. "Dynamic Nucleoplasmic and Nucleolar Localization of Mammalian RNase H1 in Response to RNAP I Transcriptional R-Loops." *Nucleic Acids Research* 45 (18): 10672–92. doi:10.1093/nar/gkx710.

Silva, Jose M, Mamie Z Li, Ken Chang, Wei Ge, Michael C Golding, Richard J Rickles, Despina Siolas, et al. 2005. "Second-Generation ShRNA Libraries Covering the Mouse and Human Genomes." *Nature Genetics* 37 (11): 1281–88. doi:10.1038/ng1650.

Singh, Jarnail, and Richard A Padgett. 2009. "Rates of in Situ Transcription and Splicing in Large Human Genes." *Nature Structural & Molecular Biology* 16 (11): 1128–33. doi:10.1038/nsmb.1666.

Siprashvili, Z, G Sozzi, L D Barnes, P McCue, A K Robinson, V Eryomin, L Sard, et al. 1997. "Replacement of Fhit in Cancer Cells Suppresses Tumorigenicity." *Proceedings of the National Academy of Sciences of the United States of America* 94 (25). National Academy of Sciences: 13771–76. doi:10.1073/PNAS.94.25.13771.

Skoog, Lambert, and Bo NordenskjOLD. 1971. "Effects of Hydroxyurea and 1-Beta-D-Arabinofuranosyl-Cytosine on Deoxyribonucleotide Pools in Mouse Embryo Cells." *European Journal of Biochemistry* 19 (1). Wiley/Blackwell (10.1111): 81–89. doi:10.1111/j.1432-1033.1971.tb01290.x.

Skourti-Stathaki, Konstantina, Kinga Kamieniarz-Gdula, and Nicholas J. Proudfoot. 2014. "R-

Loops Induce Repressive Chromatin Marks over Mammalian Gene Terminators." *Nature* 516 (7531): 436–39. doi:10.1038/nature13787.
Slack, Andrew, P. C. Thornton, Daniel B. Magner, Susan M. Rosenberg, and P. J. Hastings. 2006. "On the Mechanism of Gene Amplification Induced under Stress in Escherichia Coli." *PLoS Genetics* 2 (4). Public Library of Science: e48. doi:10.1371/journal.pgen.0020048.
Smith, Cassandra L, Andrew Bolton, and Giang Nguyen. 2010. "Genomic and Epigenomic Instability, Fragile Sites, Schizophrenia and Autism." *Current Genomics* 11 (6): 447–69. doi:10.2174/138920210793176001.
Smith, David I, Yu Zhu, Sarah McAvoy, and Robert Kuhn. 2006. "Common Fragile Sites, Extremely Large Genes, Neural Development and Cancer." *Cancer Letters* 232 (1): 48–57. doi:10.1016/j.canlet.2005.06.049.
Snyder, M, R J Sapolsky, and R W Davis. 1988. "Transcription Interferes with Elements Important for Chromosome Maintenance in Saccharomyces Cerevisiae." *Mol. Cell. Biol.* 8 (5): 2184–94. doi:10.1128/MCB.8.5.2184.
Sollier, Julie, and Karlene A. Cimprich. 2015. "Breaking Bad: R-Loops and Genome Integrity." *Trends in Cell Biology*. doi:10.1016/j.tcb.2015.05.003.
Sollier, Julie, Caroline Townsend Stork, María L. García-Rubio, Renee D. Paulsen, Andrés Aguilera, and Karlene A. Cimprich. 2014. "Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability." *Molecular Cell*, 1–9. doi:10.1016/j.molcel.2014.10.020.
Sotiriou, Sotirios K, Irene Kamileri, Natalia Lugli, Konstantinos Evangelou, Caterina Da-Ré, Florian Huber, Laura Padayachy, et al. 2016. "Mammalian RAD52 Functions in Break-Induced Replication Repair of Collapsed DNA Replication Forks." *Molecular Cell* 64 (6). Elsevier: 1127–34. doi:10.1016/j.molcel.2016.10.038.
Su'etsugu, Masayuki, and Jeff Errington. 2011. "The Replicase Sliding Clamp Dynamically Accumulates behind Progressing Replication Forks in Bacillus Subtilis Cells." *Molecular Cell* 41 (6): 720–32. doi:10.1016/j.molcel.2011.02.024.
Sutherland, G.R. 2003. "Rare Fragile Sites." *Cytogenetic and Genome Research* 100 (1–4): 77–84. doi:10.1159/000072840.
Suzuki, Y., J. B. Holmes, S. M. Cerritelli, K. Sakhuja, M. Minczuk, I. J. Holt, and R. J. Crouch. 2010. "An Upstream Open Reading Frame and the Context of the Two AUG Codons Affect the Abundance of Mitochondrial and Nuclear RNase H1." *Molecular and Cellular Biology* 30 (21): 5123–34. doi:10.1128/MCB.00619-10.
Tanaka, Seiji, Ryuichiro Nakato, Yuki Katou, Katsuhiko Shirahige, and Hiroyuki Araki. 2011. "Origin Association of Sld3, Sld7, and Cdc45 Proteins Is a Key Step for Determination of Origin-Firing Timing." *Current Biology : CB* 21 (24). Elsevier Ltd: 2055–63. doi:10.1016/j.cub.2011.11.038.
Thys, Ryan G, Christine E Lehman, Levi C T Pierce, and Yuh-Hwa Wang. 2015. "DNA Secondary Structure at Chromosomal Fragile Sites in Human Disease." *Current Genomics* 16 (1). Bentham Science Publishers: 60–70. doi:10.2174/1389202916666150114223205.
Tian, Ming, and Frederick W. Alt. 2000. "Transcription-Induced Cleavage of Immunoglobulin Switch Regions by Nucleotide Excision Repair Nucleases *in Vitro*." *Journal of Biological Chemistry* 275 (31): 24163–72. doi:10.1074/jbc.M003343200.
Truong, Lan N, and Xiaohua Wu. 2011. "Prevention of DNA Re-Replication in Eukaryotic Cells," 13–22. doi:10.1093/jmcb/mjq052.

Tsakraklides, Vasiliki, and Stephen P Bell. 2010. “Dynamics of Pre-Replicative Complex Assembly.” *The Journal of Biological Chemistry* 285 (13). American Society for Biochemistry and Molecular Biology: 9437–43. doi:10.1074/jbc.M109.072504.

Tubbs, Anthony, and André Nussenzweig. 2017. “Leading Edge Review Endogenous DNA Damage as a Source of Genomic Instability in Cancer.” doi:10.1016/j.cell.2017.01.002.

Tucker, Tracy, Farah R. Zahir, Malachi Griffith, Allen Delaney, David Chai, Erica Tsang, Emmanuelle Lemyre, et al. 2014. “Single Exon-Resolution Targeted Chromosomal Microarray Analysis of Known and Candidate Intellectual Disability Genes.” *European Journal of Human Genetics* 22 (6): 792–800. doi:10.1038/ejhg.2013.248.

Tuduri, Sandie, Laure Crabbé, Chiara Conti, Hélène Tourrière, Heidi Holtgreve-Grez, Anna Jauch, Véronique Pantesco, et al. 2009. “Topoisomerase I Suppresses Genomic Instability by Preventing Interference between Replication and Transcription.” *Nature Cell Biology* 11 (11). Nature Publishing Group: 1315–24. doi:10.1038/ncb1984.

Tumini, Emanuela, Sonia Barroso, Carmen Pérez -Calero, and Andrés Aguilera. 2016. “Roles of Human POLD1 and POLD3 in Genome Stability.” *Scientific Reports* 6 (1). Nature Publishing Group: 38873. doi:10.1038/srep38873.

Valton, Anne-Laure, and Marie-Noëlle Prioleau. 2016. “G-Quadruplexes in DNA Replication: A Problem or a Necessity?” doi:10.1016/j.tig.2016.09.004.

VanHulle, K., F. J. Lemoine, V. Narayanan, B. Downing, K. Hull, C. McCullough, M. Bellinger, K. Lobachev, T. D. Petes, and A. Malkova. 2007. “Inverted DNA Repeats Channel Repair of Distant Double-Strand Breaks into Chromatid Fusions and Chromosomal Rearrangements.” *Molecular and Cellular Biology* 27 (7): 2601–14. doi:10.1128/MCB.01740-06.

Veerappa, Avinash M., Marita Saldanha, Prakash Padakannaya, and Nallur B. Ramachandra. 2013. “Family-Based Genome-Wide Copy Number Scan Identifies Five New Genes of Dyslexia Involved in Dendritic Spinal Plasticity.” *Journal of Human Genetics* 58 (8): 539–47. doi:10.1038/jhg.2013.47.

Veloso, Artur, Killeen S Kirkconnell, Brian Magnuson, Benjamin Biewen, Michelle T Paulsen, Thomas E Wilson, and Mats Ljungman. 2014. “Rate of Elongation by RNA Polymerase II Is Associated with Specific Gene Features and Epigenetic Modifications.” *Genome Research* 24 (6): 896–905. doi:10.1101/gr.171405.113.

Verdin, Hannah, Barbara D’haene, Diane Beysen, Yana Novikova, Björn Menten, Tom Sante, Pablo Lapunzina, et al. 2013. “Microhomology-Mediated Mechanisms Underlie Non-Recurrent Disease-Causing Microdeletions of the FOXL2 Gene or Its Regulatory Domain.” Edited by Nancy B. Spinner. *PLoS Genetics* 9 (3). Public Library of Science: e1003358. doi:10.1371/journal.pgen.1003358.

Verkerk, Annemiske J.M.H., Maura Pieretti, James S. Sutcliffe, Ying Hui Fu, Derek P.A. Kuhl, Antonio Pizzuti, Orly Reiner, et al. 1991. “Identification of a Gene (FMR-1) Containing a CGG Repeat Coincident with a Breakpoint Cluster Region Exhibiting Length Variation in Fragile X Syndrome.” *Cell* 65 (5): 905–14. doi:10.1016/0092-8674(91)90397-H.

Vilette, D, S D Ehrlich, and B Michel. 1995. “Transcription-Induced Deletions in Escherichia Coli Plasmids.” *Molecular Microbiology* 17 (3): 493–504. http://www.ncbi.nlm.nih.gov/pubmed/8559068.

Vissers, Lisenka E.L.M. L M, Samarth S. Bhatt, Irene M. Janssen, Zhilian Xia, Seema R. Lalani, Rolph Pfundt, Katarzyna Derwinska, et al. 2009. “Rare Pathogenic Microdeletions and Tandem Duplications Are Microhomology-Mediated and Stimulated by Local Genomic

Architecture." *Human Molecular Genetics* 18 (19): 3579–93. doi:10.1093/hmg/ddp306.
Wahba, Lamia, Jeremy D. Amon, Douglas Koshland, and Milena Vuica-Ross. 2011. "RNase H and Multiple RNA Biogenesis Factors Cooperate to Prevent RNA:DNA Hybrids from Generating Genome Instability." *Molecular Cell* 44 (6). Cell Press: 978–88. doi:10.1016/J.MOLCEL.2011.10.017.
Wang, Hailong, and Xingzhi Xu. 2017. "Microhomology-Mediated End Joining: New Players Join the Team." *Cell & Bioscience* 7 (1). BioMed Central: 6. doi:10.1186/s13578-017-0136-8.
Wang, James C. 2002. "Cellular Roles of DNA Topoisomerases: A Molecular Perspective." *Nature Reviews Molecular Cell Biology* 3 (6). Nature Publishing Group: 430–40. doi:10.1038/nrm831.
Wang, L, J Darling, J S Zhang, H Huang, W Liu, and D I Smith. 1999. "Allele-Specific Late Replication and Fragility of the Most Active Common Fragile Site, FRA3B." *Human Molecular Genetics* 8 (3): 431–37. http://www.ncbi.nlm.nih.gov/pubmed/9949202.
Wang, Liming, Hui Shen, Bei Feng, Da Zhu, Lan Yu, Xun Tian, Ci Ren, et al. 2017. "Reduction in the Copy Number and Expression Level of the Recurrent Human Papillomavirus Integration Gene Fragile Histidine Triad (FHIT) Predicts the Transition of Cervical Lesions." In *PLoS ONE*. Vol. 12. doi:10.1371/journal.pone.0175520.
Wang, Xianlong, Jinwei Zhou, Chunwei Cao, Jiaojiao Huang, Tang Hai, Yanfang Wang, Qiantao Zheng, et al. 2015. "Efficient CRISPR/Cas9-Mediated Biallelic Gene Disruption and Site-Specific Knockin after Rapid Selection of Highly Active SgRNAs in Pigs." *Scientific Reports* 5 (August). Nature Publishing Group: 13348. doi:10.1038/srep13348.
Wang, Zhibin, Chongzhi Zang, Jeffrey A Rosenfeld, Dustin E Schones, Artem Barski, Suresh Cuddapah, Kairong Cui, et al. 2008. "Combinatorial Patterns of Histone Acetylations and Methylations in the Human Genome." *Nature Genetics* 40 (7). NIH Public Access: 897–903. doi:10.1038/ng.154.
Wansink, D G, E E Manders, I van der Kraan, J A Aten, R van Driel, and L de Jong. 1994. "RNA Polymerase II Transcription Is Concentrated Outside Replication Domains throughout S-Phase." *Journal of Cell Science* 107 (Pt 6 (June): 1449–56. http://www.ncbi.nlm.nih.gov/pubmed/7962188.
Wei, Pei-Chi, Amelia N. Chang, Jennifer Kao, Zhou Du, Robin M. Meyers, Frederick W. Alt, and Bjoern Schwer. 2016. "Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells." *Cell* 164 (4). Cell Press: 644–55. doi:10.1016/J.CELL.2015.12.039.
Wei, X, J Samarabandu, R S Devdhar, A J Siegel, R Acharya, and R Berezney. 1998. "Segregation of Transcription and Replication Sites into Higher Order Domains." *Science (New York, N.Y.)* 281 (5382): 1502–6. http://www.ncbi.nlm.nih.gov/pubmed/9727975.
White, S.J., A. Aartsma-Rus, K.M. Flanigan, R.B. Weiss, A.L.J. Kneppers, T. Lalic, A.A.M. Janson, H.B. Ginjaar, M.H. Breuning, and J.T. den Dunnen. 2006. "Duplications in TheDMD Gene." *Human Mutation* 27 (9): 938–45. doi:10.1002/humu.20367.
Wilke, Charles M., Bryan K. Hall, Ann Hoge, William Paradee, David I. Smith, and Thomas W. Glover. 1996. "FRA3B Extends Over a Broad Region and Contains a Spontaneous HPV16 Integration Site: Direct Evidence for the Coincidence of Viral Integration Sites and Fragile Sites." *Human Molecular Genetics* 5 (2). Oxford University Press: 187–95. doi:10.1093/hmg/5.2.187.
Wilson, Thomas E, Martin F Arlt, So Hae Park, Sountharia Rajendran, Michelle Paulsen, Mats

Ljungman, and Thomas W Glover. 2015. “Large Transcription Units Unify Copy Number Variants and Common Fragile Sites Arising under Replication Stress.” *Genome Res.* 25 (2): 189–200. doi:10.1101/gr.177121.114.

Wong, Philip G., Sherry L. Winter, Elena Zaika, Thinh V. Cao, Umut Oguz, John M. Koomen, Joyce L. Hamlin, and Mark G. Alexandrow. 2011. “Cdc45 Limits Replicon Usage from a Low Density of PreRCs in Mammalian Cells.” Edited by Anja-Katrin Bielinsky. *PloS One* 6 (3). Public Library of Science: e17533. doi:10.1371/journal.pone.0017533.

Woodward, Anna M, Thomas Göhler, M Gloria Luciani, Maren Oehlmann, Xinquan Ge, Anton Gartner, Dean a Jackson, and J Julian Blow. 2006. “Excess Mcm2-7 License Dormant Origins of Replication That Can Be Used under Conditions of Replicative Stress.” *The Journal of Cell Biology* 173 (5): 673–83. doi:10.1083/jcb.200602108.

Wu, Hongjiang, Hong Sun, Xuehai Liang, Walt F. Lima, and Stanley T. Crooke. 2013. “Human RNase H1 Is Associated with Protein P32 and Is Involved in Mitochondrial Pre-RRNA Processing.” Edited by Janine Santos. *PLoS ONE* 8 (8): e71006. doi:10.1371/journal.pone.0071006.

Wu, Yun, Kun Xu, Chonghua Ren, Xinyi Li, Huijiao Lv, Furong Han, Zehui Wei, Xin Wang, and Zhiying Zhang. 2017. “Enhanced CRISPR/Cas9-Mediated Biallelic Genome Targeting with Dual Surrogate Reporter-Integrated Donors.” *FEBS Letters* 591 (6). Wiley-Blackwell: 903–13. doi:10.1002/1873-3468.12599.

Xu, Baoji, and David A Clayton. 1996. “RNA-DNA Hybrid Formation at the Human Mitochondrial Heavy-Strand Origin Ceases at Replication Start Sites: An Implication for RNA-DNA Hybrids Serving as Primers.” *The EMBO Journal* 15 (12): 3135–43.

Yang, Fan, Sheila S. Teves, Christopher J. Kemp, and Steven Henikoff. 2014. “Doxorubicin, DNA Torsion, and Chromatin Dynamics.” *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1845 (1). Elsevier: 84–89. doi:10.1016/J.BBCAN.2013.12.002.

Yu, Kefei, Frederic Chedin, Chih Lin Hsieh, Thomas E. Wilson, and Michael R. Lieber. 2003. “R-Loops at Immunoglobulin Class Switch Regions in the Chromosomes of Stimulated B Cells.” *Nature Immunology* 4 (5): 442–51. doi:10.1038/ni919.

Yu, Kefei, Deepankar Roy, Feng-Ting Huang, and Michael R. Lieber. 2006. “Detection and Structural Analysis of R-Loops.” In *Methods in Enzymology*, 409:316–29. doi:10.1016/S0076-6879(05)09018-X.

Yu, S., M. Pritchard, E. Kremer, M. Lynch, J. Nancarrow, E. Baker, K. Holman, et al. 1991. “Fragile X Genotype Characterized by an Unstable Region of DNA.” *Science* 252 (5009): 1179–81. doi:10.1126/science.252.5009.1179.

Zack, Travis I, Steven E Schumacher, Scott L Carter, Andrew D Cherniack, Gordon Saksena, Barbara Tabak, Michael S Lawrence, et al. 2013. “Pan-Cancer Patterns of Somatic Copy Number Alteration.” *Nature Genetics* 45 (10). Nature Publishing Group: 1134–40. doi:10.1038/ng.2760.

Zeman, Michelle K, and Karlene A Cimprich. 2013. “Causes and Consequences of Replication Stress.” *Nature Cell Biology* 16 (1). NIH Public Access: 2–9. doi:10.1038/ncb2897.

Zhang, Feng, Pavel Seeman, Pengfei Liu, Marian A.J. Weterman, Claudia Gonzaga-Jauregui, Charles F. Towne, Sat Dev Batish, et al. 2010. “Mechanisms for Nonrecurrent Genomic Rearrangements Associated with CMT1A or HNPP: Rare CNVs as a Cause for Missing Heritability.” *The American Journal of Human Genetics* 86 (6). Cell Press: 892–903. doi:10.1016/J.AJHG.2010.05.001.

Zhang, Haihua, and Catherine H Freudenreich. 2007. “An AT-Rich Sequence in Human

Common Fragile Site FRA16D Causes Fork Stalling and Chromosome Breakage in S. Cerevisiae." *Molecular Cell* 27 (3). NIH Public Access: 367–79. doi:10.1016/j.molcel.2007.06.012.

Zhao, Yongping, Congsheng Zhang, Wenwen Liu, Wei Gao, Changlin Liu, Gaoyuan Song, Wen Xue Li, et al. 2016. "An Alternative Strategy for Targeted Gene Replacement in Plants Using a Dual-SgRNA/Cas9 Design." *Scientific Reports* 6. doi:10.1038/srep23890.

Zheng, Qiupeng, Xiaohong Cai, Meng How Tan, Steven Schaffert, Christopher P. Arnold, Xue Gong, Chang Zheng Chen, and Shenglin Huang. 2014. "Precise Gene Deletion and Replacement Using the CRISPR/Cas9 System in Human Cells." *BioTechniques* 57 (3): 115–24. doi:10.2144/000114196.

Zhou, Yi, Pierre Caron, Gaëlle Legube, and Tanya T Paull. 2014. "Quantitation of DNA Double-Strand Break Resection Intermediates in Human Cells." *Nucleic Acids Research* 42 (3). Oxford University Press: e19. doi:10.1093/nar/gkt1309.

Zlotorynski, Eitan, Ayelet Rahat, Jennifer Skaug, Neta Ben-Porat, Efrat Ozeri, Ruth Hershberg, Ayala Levi, Stephen W Scherer, Hanah Margalit, and Batsheva Kerem. 2003. "Molecular Basis for Expression of Common and Rare Fragile Sites." *Molecular and Cellular Biology* 23 (20). American Society for Microbiology (ASM): 7143–51. doi:10.1128/MCB.23.20.7143-7151.2003.

Zotter, Angelika, Martijn S Luijsterburg, † Daniël, O Warmerdam, Shehu Ibrahim, Alex Nigg, Wiggert A Van Cappellen, et al. 2006. "Recruitment of the Nucleotide Excision Repair Endonuclease XPG to Sites of UV-Induced DNA Damage Depends on Functional TFIIH." *MOLECULAR AND CELLULAR BIOLOGY* 26 (23): 8868–79. doi:10.1128/MCB.00695-06.