SMARCAL1 Maintains Telomere Integrity During DNA Replication

By

Lisa A. Poole

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

December 16, 2017

Nashville, Tennessee

Approved:

David Cortez, Ph.D.

John York, Ph.D.

Scott Hiebert, Ph.D.

Katherine Friedman, Ph.D.

Emily Hodges, Ph.D.

To my parents, I love you both

and to my brother for always being amazing

ACKNOWLEDGEMENTS

A big thank you goes to Nancy Zhao and Gloria Glick who have both contributed to my project. To Nancy, I appreciate all the work you did to keep my project going and also keep the lab up and running efficiently. Thank you, Gloria, for your kind words throughout my time in the Cortez lab when I felt down and all your gifts of fruit that really brightened my days.

I also would like to thank current and previous students of the Cortez lab who have made this journey enjoyable. Thank you to Kami, who joined the same year as me, for all her insight and guidance both with science and Bollywood dancing. To Kareem, who is brilliant and sometimes challenging, it has been a privilege to talk science and look at cute pictures of puppies with you. Huzefa, I have appreciated you responding all my questions about iPOND without frustration and opening my eyes to a world of different cakes. To Thomas, I have always appreciated your infectious optimism about both science and other things every day. Sarah, it was great to get to know you in the last year and I have really appreciated all your advice, especially when it comes to matters concerning job searches and baking. Vaughn and Petria, you both are wonderful labmates who are not afraid to speak up in lab meetings. Keep that up. To Taha, Archana, and Aditya - the newest members of the lab – it was a pleasure to get to know you and I foresee great things in your futures in the Cortez lab. I would also like to thank Bianca Sirbu – former Cortez lab member and my rotation mentor – for being an inspiring mentor and for continuing to check on me even after she has moved on. To Gina Kavanaugh, a fellow Hoosier and former Cortez member, thank you for your chats about science and life and delicious Korean food. I would also like to thank Akosua Badu-Nkansah, whose late-night pep talks were instrumental in staying upbeat about science. I also need to extend my gratitude to Courtney Lovejoy, a former member of the lab that jump started the telomere work presented in this dissertation and has continued to

grateful we were able to become friends after all these years since high school and I have enjoyed our fun times together, especially hanging out with the Predators. Lauren Salay, even though we never met in undergrad, I'm glad we became friends at Vandy. Knowing you has been a pleasure and thank you for my introduction to the magic that is Dr. Who. Finally, to my dear friend Alex Trevisan, thank you for taking pity on my that first day in IMPACT and sparking a friendship that would last all these years. These years have been filled with Korean food, dramas, and wonderful times that I would not change for anything.

Finally, I need to thank my family for being the source of my strength and success. I have never wanted for anything thanks to the love and support from my mother and father. I have tried to be as hard working and dedicated to excellence as they are in all things they do in life. Mom, you have always taught me to keep smiling because even the worst in life eventually gets better – and you are right, as always. Dad, I only hope to be as clever and hard working as you have been all my life. To my brother, Nicholas, you have always been a source of support and love. Thank you for always being there and pushing me forward when I wanted to quit. Finally, James Pino, you have become family to me in these last few years pushing me forward when I was down and stuck in a rut. You have helped me to take on what I thought would be the impossible and helped me learn Python. I love and appreciate all that you are and all that you do.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

4OHT	4-hydroxytamoxofin
53BP1	p53 Binding Protein 1
γH2AX	H2AX Phosphoserine 139
ABP	ALT-associated PML Body
AH2	Annealing Helicase 2
ALT	Alternative Lengthening of Telomeres
APIM	AlkB homology 2 PCNA interaction motif
ASF1	Anti-silencing Function 1
BLESS	Break Labeling, Enrichment on Streptavidin, and Next-Generation Sequencing
BLM	Bloom Syndrome Protein
BRCA1	Breast Cancer Type 1 Susceptibility Protein
BRCA2	Breast Cancer Type 2 Susceptibility Protein
C	Cytidine
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation with DNA sequencing
CO-FISH	Chromosome Orientation-Fluorescence *In Situ* Hybridization
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CST complex	CDC13, STN1, TEN1 complex
DDR	DNA Damage Response
D-loop	Displacement Loop
dsDNA	Double Stranded DNA

END-seq	DNA End Sequencing
EtOH	Ethanol
FBS	Fetal Bovine Serum
GFP	Green Fluorescent Protein
HARP	HepA Related Protein
HDR	Homology Directed Repair
HTGTS	High Throughput, Genome-wide Translocation Sequencing
HLTF	Helicase-like Transcription Factor
HU	Hydroxyurea
ISWI	Imitation SWI
LIGIV	Ligase IV
LINE	Long Interspersed Nuclear Element
MEF	Mouse Embryonic Fibroblast
MMS	Methyl Methanesulfonate
MMS2	Methyl-methanesulphonate sensitivity 2
MSI	Microsatellite Instability
MTS	Multiple Telomere Signals
MUS81	MUS81 Structure-specific Endonuclease Subunit
NGS	Next-Generation Sequencing
NHEJ	Non-homologous End Joining
NZF	NPL4 Zinc-finger
PBS	Phosphate Buffer Solution
PCNA	Proliferating Cell Nuclear Antigen

PCR	Polymerase Chain Reaction
PEI	Polyethyleneimine
PIP-box	PCNA Interacting Protein Box
PML	Promyelocytic Leukemia
POT1	Protection of Telomeres Protein 1
PTIP	Pax Transactivation Domain-Interacting Protein
RAD5	Radiation Sensitivity Protein 5
RAD6	Radiation Sensitivity Protein 6
RAD18	Radiation Sensitivity Protein 18
RBD	RPA Binding Domain
RECQ1	RecQ Like Helicase 1
RECQ5	RecQ Like Helicase 5
RING	Really interesting New Gene (E3 ligase domain)
RPA	Replication Protein A
RTEL1	Regulator of Telomere Elongation Helicase 1
SCE	Sister Chromatid Exchange
SINE	Short Interspersed Nuclear Element
SIOD	Schimke Immune-osseus Dysplasia
SLX4	SLX4 Structure Specific Endonuclease Subunit
SMARCAL1	SWI/SNF Related, Matrix Associated Regulator of Chromatin Subfamily A-like 1
SNF2	Sucrose Non-Fermenting 2
SRD	Substrate Recognition Domain

ssDNA	Single Stranded DNA
T	Thymidine
TIF	Telomere Dysfunction-Induced Foci
TMPyP4	Tetra-(N-methyl-4-pyridyl)porphyrin
t-SCE	Telomere Sister Chromatid Exchange
TSS	Transcription Start Site
UBC13	E2 Ubiquitin-conjugating Protein UBC13
UV	Ultraviolet Radiation
WRN	Werner Syndrome Protein
WT	Wild-type
ZRANB3	Zinc finger RANBP2-type containing 3

CHAPTER I[†]

INTRODUCTION

Replication stress and the DNA damage response

DNA replication is an essential biological process where the genetic information in a cell is duplicated to generate two copies of the genome. These two copies are then segregated into separate daughter cells during mitosis, or cell division. In humans, nearly 7 billion base pairs of DNA have to be accurately duplicated. This essential process is constantly challenged by various sources of replication stress that can impede the accurate and timely completion of genome duplication (Zeman and Cimprich, 2014). These challenges include defects in the DNA template such as base damage and backbone breaks, collisions between the replication and transcription machineries, limiting precursors caused by oncogene-driven aberrant cell cycles, and inherently difficult to replicate sequences that are prone to forming polymerase stalling structures.

Avoiding mutations or chromosome aberrations requires a replication stress response that involves hundreds of proteins acting in multiple pathways. These pathways coordinate to perform diverse function in response to replication stress including direct repair of DNA lesions, activation of cell cycle checkpoints, and regulation of transcription and metabolism (Ciccia and Elledge, 2010). If the damage is too great, there are also proteins that promote exit from the cell cycle through senescence or apoptosis (Ciccia and Elledge, 2010). These functions ensure the integrity of the genome and prevent the development of several human diseases.

Replication fork remodeling

Several proteins in the DNA damage response (DDR) promote the stability of replication forks through fork remodeling. When a replication fork encounters stress, the fork can stall resulting in the uncoupling of the DNA polymerase and the helicase (Figure 1.1) (Byun et al., 2005). This uncoupling, in turn, generates long stretches of single stranded DNA (ssDNA) that can be cleaved by structure specific nucleases to generate double stranded DNA (dsDNA) breaks and, subsequently, collapse of the replication fork (Cortez, 2015). Therefore, stalled forks must be stabilized to prevent this nuclease cleavage and genome instability. In situations when there is no converging replication fork, the stalled fork must also restart to complete DNA synthesis (Figure 1.1).

To stabilize the stalled fork, there are several enzymes in the DDR capable of performing replication fork reversal (also called fork regression) (Fugger et al., 2015; Neelsen and Lopes, 2015; Zellweger et al., 2015). This process involves a concerted reannealing of the parental template strands to reverse the direction of the replication fork and generate a four-way junction, also termed the "chicken foot" (Figure 1.2) (Neelsen and Lopes, 2015). Fork reversal is thought to stabilize the fork, promote repair of the parental DNA template, or serve as a template switching mechanism to bypass the source of replication stress. Once the source of stress has been addressed, several enzymes can also catalyze the reverse reaction (called fork restoration), to reset the replication fork to a three-way junction from the reversed chicken foot structure (Figure 1.2) (Neelsen and Lopes, 2015). The importance of this process is highlighted by the abundance of human diseases associated with mutations in proteins that catalyze fork remodeling.

Fork reversal has been observed in human cells treated with replication stress inducing agents using electron microscopy (Zellweger et al., 2015). Interestingly, the type of damage had

2
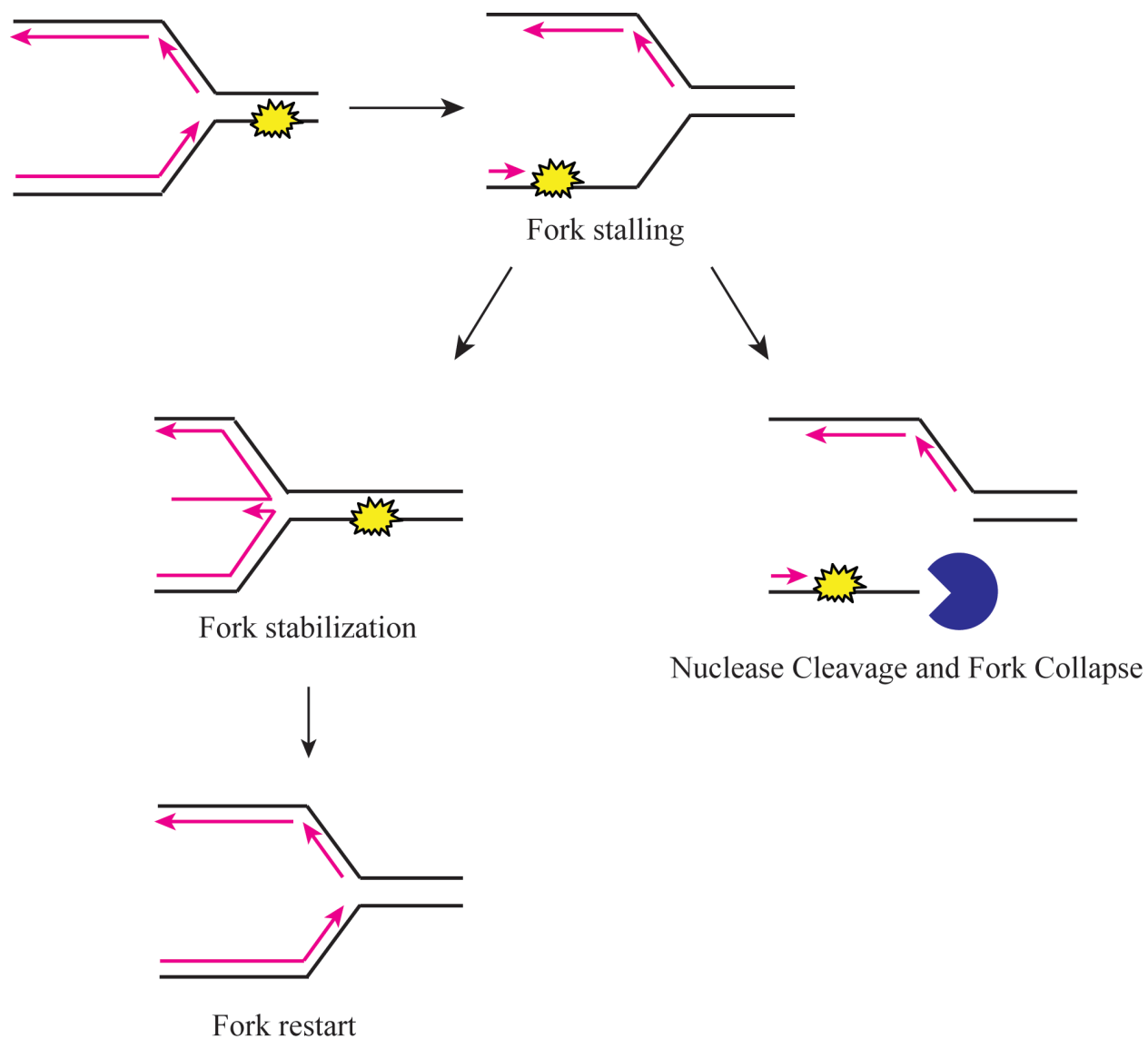
Figure 1.1. Model of replication fork stalling in response to replication stress. Uncoupling of the DNA polymerase and helicase in response to a lesion on the leading strand leads to long stretches of ssDNA. Stalled forks can then be stabilized to permit repair or bypass of the source of replication stress to promote fork restart (left) or nucleases cleave the stalled fork to generate dsDNA breaks and collapsed replication forks (right).
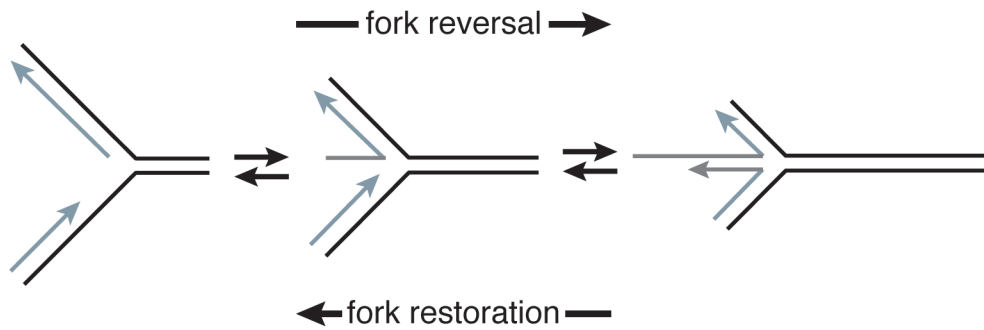
Figure 1.2. Diagram depicting fork reversal and fork restoration. Fork reversal anneals the parental (black) and nascent (silver) DNA strands in a concerted reaction generating a chicken foot structure. Migrating the four-way junction in the opposite direction yields fork restoration. A ssDNA gap is diagramed on the leading strand template in this example.

no effect on the prevalence of reversed forks indicating this process could be a universal response to any type of replication fork stalling (Zellweger et al., 2015). However, it is unclear if this phenomenon is also true for fork stalling caused by endogenous sources of replication stress.

The SNF2 family of DNA translocases

One subset of the DNA damage response utilizes enzymes in the SNF2 family of DNA translocases to remodel stalled DNA replication forks. SNF2 proteins are ATP-dependent motors that act in a multiple aspects of nucleic acid metabolism. These proteins share a bi-lobed helicase-like domain and have been classified by sequence differences into multiple subfamilies (Figure 1.3) (Flaus et al., 2006; Flaus and Owen-Hughes, 2011). Highly studied members of this family include the catalytic subunits of the SWI/SNF, ISWI, and INO80 chromatin remodeling complexes that act during transcription and DNA repair to alter nucleosome position or composition.

Some of the more distant subfamilies of SNF2 enzymes contain proteins that act in response to replication stress but do not affect chromatin structure. SMARCAL1 (SWI/SNF related, matrix associated regulator of chromatin subfamily A-like 1) and ZRANB3 (Zinc finger RANBP2-type containing 3) are part of the most distant subgroup while HLTF (helicase-like transcription factor) is a member of the RAD5/16-like group named for the yeast RAD5 protein that acts in a post-replicative repair pathway (Figure 1.3) (Flaus and Owen-Hughes, 2011; Unk et al., 2010). Despite sequence conservation with several chromatin remodeling proteins, none of these enzymes assemble into larger complexes, interact with histones, or bind DNA in a sequence-specific manner – all characteristics of chromatin remodeling enzymes. Instead, SMARCAL1, ZRANB3, and HLTF all function in the DDR to stabilize replication forks through several mechanisms including fork remodeling.

5
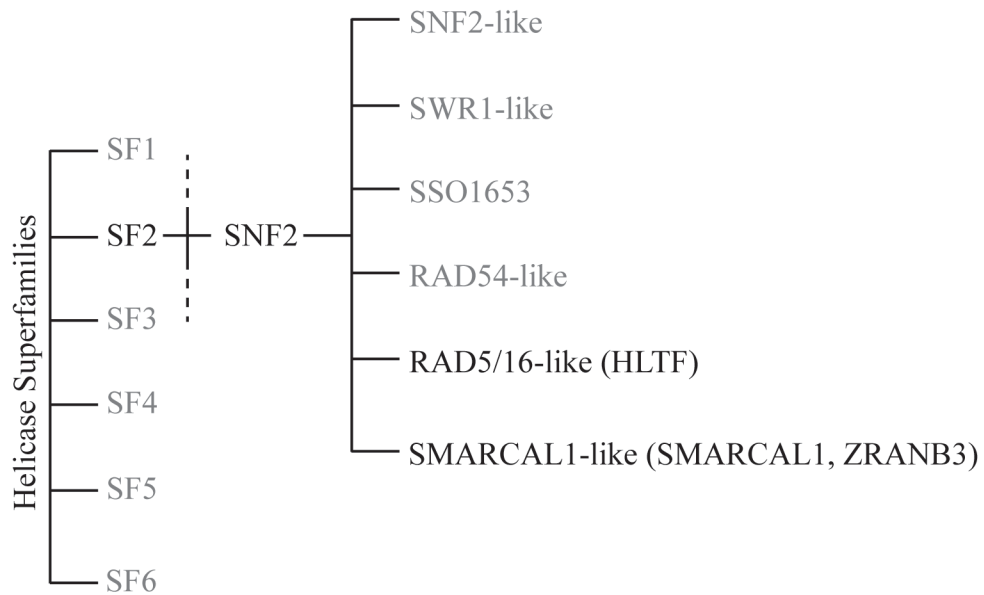
Figure 1.3. Classification tree of SNF2 subfamilies. Subfamilies of SNF2 proteins are shown clustered based on the sequence similarities of the ATPase domains present in these enzymes. My work focuses on the proteins SMARCAL1 and ZRANB3, found in the most distant subfamily, and HLTF, found in the RAD5/16-like subfamily. Other domains shown in grey contain various proteins involved in chromatin remodeling and transcription regulation. This figure was adapted from Flaus et al., 2006.

SMARCAL1 remodels replication forks to promote genome stability

SMARCAL1 (also called HARP) is a 954-amino acid protein containing a replication protein A (RPA) binding domain at the N-terminus followed by two HARP domains that are important for DNA binding and conferring substrate specificity. The two lobes of the separated ATPase domain are found in the C-terminal half of the protein (Figure 1.4). SMARCAL1 functions as a DNA-dependent ATPase that translocates on DNA (Coleman et al., 2000). Defects in this activity have severe consequences on genome integrity and human health. Bi-allelic loss-of-function mutations in *SMARCAL1* result in the human disease Schimke immuno-osseous dysplasia (SIOD), a rare human disease where only a few hundred patients have been described (Boerkoel et al., 2002). SIOD manifests with diverse phenotypes including renal dysfunction, immune deficiencies, microcephaly, and growth defects (Boerkoel et al., 2002). Most patients do not survive past childhood due to infections, although a few have developed cancer suggesting a mild cancer predisposition (Carroll et al., 2013). *SMARCAL1* mutations linked to SIOD typically truncate the protein or are found in critical functional domains highlighting the importance of the enzymatic activity of this protein (Boerkoel et al., 2002; Carroll et al., 2013; Clewing et al., 2007).

The first biochemical activity identified for SMARCAL1, other than ATP hydrolysis, was the reannealing of complementary DNA sequences (Yusufzai and Kadonaga, 2008). In a plasmid-based assay where complementary ssDNA is bound by the ssDNA binding protein RPA, SMARCAL1 is able to reanneal the complementary strands, evicting RPA, to generate an entirely dsDNA plasmid (Figure 1.5) (Yusufzai and Kadonaga, 2008). Importantly, this function is different from the unwinding action of canonical DNA helicases that does not require complementary DNA substrates (Yusufzai and Kadonaga, 2008). DNA binding and ATP hydrolysis are required for this complementary annealing function of SMARCAL1 *in vitro*
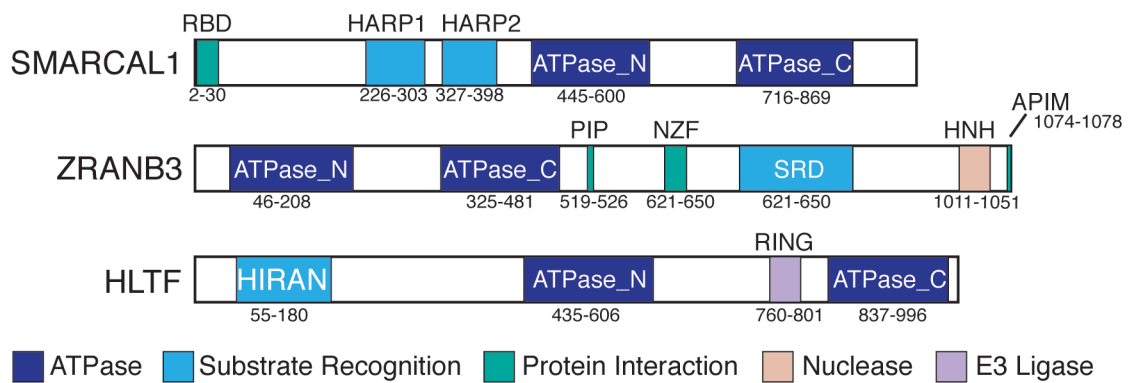
Figure 1.4. Domain structures of SMARCAL1, ZRANB3, and HLTF. The ATPase, substrate recognition, protein interaction, and other enzymatic domains are depicted. RBD, RPA binding domain; SRD, substrate recognition domain.
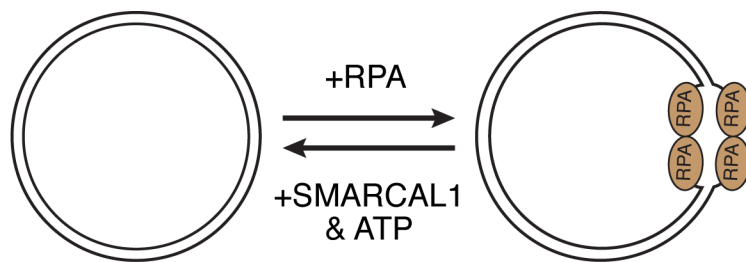
Figure 1.5. Annealing helicase activity assay. At high concentrations, RPA will induce and stabilize a single-stranded DNA bubble in supercoiled plasmid DNA. SMARCAL1 uses the energy of ATP hydrolysis to re-anneal the complementary DNA strands and displace RPA.

(Betous et al., 2012; Yusufzai and Kadonaga, 2008). SMARCAL1 was the first mammalian enzyme identified with this "annealing helicase" activity. SMARCAL1 uses its annealing activity to catalyze fork reversal and fork restoration on DNA substrates *in vitro* and in cells (Betous et al., 2013a; Betous et al., 2012; Bhat et al., 2015; Kolinjivadi et al., 2017).

SMARCAL1 localizes to sites of replication stress to catalyze fork remodeling in cells by directly contacting RPA through a conserved N-terminal motif (Figure 1.4) (Bansbach et al., 2009; Ciccia et al., 2009; Postow et al., 2009; Yuan et al., 2009; Yusufzai et al., 2009). This interaction is required for SMARCAL1 to prevent fork damage (Bansbach et al., 2009; Bansbach et al., 2010; Ciccia et al., 2009; Yuan et al., 2009). RPA also directly regulates the fork remodeling functions of SMARCAL1. RPA confers substrate specificity to SMARCAL1 reactions by inhibiting it in some cases while activating it in others (discussed below and (Betous et al., 2013a; Bhat et al., 2015)). Further, RPA also stimulates the rate of fork reversal and the processive power of SMARCAL1 when bound to DNA substrates that mimic stalled replication forks (Betous et al., 2013a). Importantly, this interaction with RPA is not shared by the other SNF2 fork remodeling proteins.

ZRANB3 is an annealing helicase, fork remodeler, and structure-specific nuclease

ZRANB3 is a 1079-amino acid protein and, based on sequence homology, is the most closely related SNF2 protein to SMARCAL1 (Flaus and Owen-Hughes, 2011). The separated lobes of the conserved ATPase domain are located at the N-terminal portion of the protein followed by several protein interaction domains, a substrate recognition domain (SRD), and an HNH endonuclease domain (Figure 1.4). Like SMARCAL1, defects in ZRANB3 function cause genome instability. For example, ZRANB3-deficient cells display higher rates of replication fork stalling

and increases in sister chromatid exchanges (SCEs), a marker of hyper-recombination (Ciccia et al., 2012). ZRANB3 deficiency also causes hyper-sensitivity to diverse DNA damaging agents (Ciccia et al., 2012; Weston et al., 2012; Yuan et al., 2012). Despite these striking phenotypes, ZRANB3 deficiency has not been directly associated with any human diseases; however, reports of *ZRANB3* mutations in endometrial cancers suggest it may function as a tumor suppressor (Lawrence et al., 2014).

Like SMARCAL1, ZRANB3 lacks helicase activity but can act to anneal complementary DNA using the plasmid assay containing an RPA-induced bubble in the DNA duplex (Yusufzai and Kadonaga, 2010). As it was the second protein identified with this "annealing helicase" function, it is also known as annealing helicase 2 (AH2). Furthermore, like SMARCAL1, ZRANB3 catalyzes both fork reversal and restoration *in vitro* (Betous et al., 2013a; Ciccia et al., 2012). In the absence of any additional proteins, ZRANB3 can catalyze fork reversal of DNA substrates containing a gap on the leading or lagging strand with equal efficiency (Betous et al., 2013a). When RPA is added, a situation mirroring what is happening in cells during DNA replication, fork reversal on DNA substrates containing a lagging strand gap remains unchanged; whereas the presence of RPA on replication forks with a leading strand gap results in a strong inhibition of fork reversal (Betous et al., 2013a). RPA also inhibits ZRANB3-mediated restoration of a replication fork with a gap on the lagging strand. Thus, while it is able to catalyze these same reactions *in vitro*, ZRANB3 displays different substrate preferences compared to SMARCAL1 (discussed below (Betous et al., 2013a)).

In cells, ZRANB3 localizes to sites of replication stress using several domains that bind polyubiquitinated PCNA (Ciccia et al., 2012; Weston et al., 2012). When the replisome encounters an obstacle to replication, PCNA is polyubiquitinated on lysine 164 (K164) which acts as a signal

for fork restart through error-free methods such as template switching (Mailand et al., 2013). ZRANB3 contains two motifs that directly contact PCNA – a conserved canonical PCNA-interacting protein (PIP) box and an APIM (AlkB homology 2 PCNA interaction motif) (Figure 1.4) (Ciccia et al., 2012; Sebesta et al., 2017; Weston et al., 2012; Yuan et al., 2012). ZRANB3 is also the first human protein known to contain an NZF (NPL4 zinc-finger) domain that binds poly-ubiquitin chains (Ciccia et al., 2012; Weston et al., 2012). All three domains are important for facilitating ZRANB3 recruitment to sites of damage (Ciccia et al., 2012; Sebesta et al., 2017; Weston et al., 2012; Yuan et al., 2012).

In contrast to other SNF2 family members, ZRANB3 possesses endonuclease activity in addition to its other enzymatic functions (Badu-Nkansah et al., 2016b; Sebesta et al., 2017; Weston et al., 2012). The endonuclease activity depends on ATP hydrolysis by the intact motor domain as well as a C-terminal HNH nuclease domain (Figure 1.4). The HNH domain, named for the conserved amino acids that compose this domain, is characteristic of numerous bacterial and fungal nucleases including the RNA guided nuclease Cas9 (Jinek et al., 2012; Yusufzai and Kadonaga, 2010). Interestingly, ZRANB3 is the only protein present in vertebrates described to contain this domain. Further, structure comparisons of the HNH of ZRANB3 with other HNH domains from lower organisms indicate that ZRANB3 contains an evolutionarily conserved unique sequence insert in the HNH domain that is absent in most other organisms (Sebesta et al., 2017).

DNA cleavage by ZRANB3 happens on one of the two strands of a DNA duplex, but it requires adjacent ssDNA. The preferred DNA substrate is a splayed arm with a minimum of 20 nucleotides of ssDNA (Weston et al., 2012). ZRANB3 will also cleave replication fork structures as long as ssDNA present at the fork junction (Weston et al., 2012). On these substrates, ZRANB3 generates a nick two nucleotides into the DNA duplex on the leading strand template (Figure 1.6).
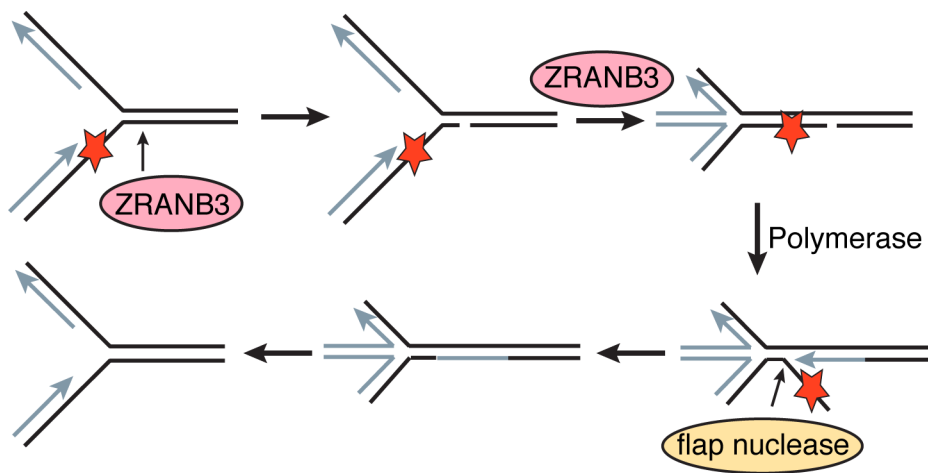
Figure 1.6. Model for how ZRANB3 nuclease and fork remodeling activities could be coordinated to repair a leading strand template lesion (red star). ZRANB3 endonuclease cuts two nucleotides into the parental duplex. Fork reversal could then stabilize the fork and allow for strand displacement DNA synthesis. A flap endonuclease could then remove the damaged DNA and permit fork restoration to restart the fork. This model is adapted from Weston et al., Genes and Development (Weston et al., 2012).

Importantly, the nuclease activity conferred by the ZRANB3 HNH domain is dependent on an intact motor domain, ATP hydrolysis, and an intact SRD domain (Badu-Nkansah et al., 2016b; Sebesta et al., 2017; Weston et al., 2012). How this linkage is achieved is not known. One proposed model is that ZRANB3 nicks the leading strand duplex and catalyzes fork reversal to prevent the formation of a double-strand break (Weston et al., 2012). The nicking activity leaves a 3'OH group that could be extended by a polymerase. If the polymerase displaces the damaged DNA, then flap cleavage could yield a religatable nick and successful repair of the lesion (Figure 1.6).

Some cancer-associated mutations in ZRANB3 inactivate its nuclease activity without affecting its ATPase activity (Sebesta et al., 2017), and the nuclease domain contributes to ZRANB3 localization to damaged forks (Weston et al., 2012). Thus, nuclease activity may be important for its genome protection functions, but further studies will be needed to determine if nuclease inactivation actually drives tumorigenesis.

HLTF is a fork remodeler and E3 ubiquitin ligase

HLTF is a 1009-amino acid protein that functions primarily in the replication stress response. Although named as a helicase-like transcription factor, it is unlikely to directly regulate gene expression. Defects in HLTF function without exogenous stress do not result in strong genome instability phenotypes in cell culture (Blastyak et al., 2010). However, HLTF is commonly silenced in colorectal cancers indicating it may function as a tumor suppressor like ZRANB3 (Moinova et al., 2002; Sandhu et al., 2012). Like SMARCAL1 and ZRANB3, HLTF is an ATP-dependent dsDNA translocase (Blastyak et al., 2010). The function of HLTF in the replication stress response is multi-faceted and dependent on both its motor domain and an associated ubiquitin ligase activity conferred by the RING domain (Figure 1.4). Failure to perform any of

these functions results in increases in replication fork collapse and decreased cell viability after treatment with UV radiation or alkylating agent methyl methanesulfonate (MMS) (Blastyak et al., 2010; Motegi et al., 2008; Unk et al., 2008).

HLTF shares a high degree of sequence conservation with the yeast protein Rad5, a ubiquitin ligase that promotes post-replication repair of DNA damage through an error-free pathway (Unk et al., 2008). When a replication fork stalls in yeast, PCNA is monoubiquitinated through the concerted efforts of Rad6, a ubiquitin conjugating enzyme, and Rad18, a ubiquitin ligase that modifies PCNA on K164 (Figure 1.7) (Unk et al., 2010). Rad5 then transfers K63 polyubiquitin chains to PCNA that were assembled by the E2 ubiquitin conjugating enzyme complex Mms2/Ubc13 (Hoege et al., 2002). Yeast cells lacking Rad5 are hyper-sensitive to UV radiation, similar to HLTF-deficiency in human cells. Initial studies suggested human HLTF could compensate for Rad5 in *S. cerevisiae*; however, this result has not been reproduced by other groups (MacKay et al., 2009; Motegi et al., 2008; Unk et al., 2008). Nonetheless, HLTF is able to polyubiquitinate PCNA through the action of its RING domain, indicating it may share some functions with Rad5 (Lin et al., 2011; Masuda et al., 2012; Motegi et al., 2008; Unk et al., 2008). This ubiquitin ligase function of HLTF is required for genome stability as mutations in the responsible RING domain result in increases in replication fork collapse after treatment DNA damaging agents (Blastyak et al., 2010).

In addition to its ubiquitin ligase functions, HLTF is also able to catalyze fork reversal like SMARCAL1 and ZRANB3 (Achar et al., 2011; Blastyak et al., 2010). HLTF reverses replication forks either lacking any ssDNA or containing a gap on the leading strand in a process dependent on the ATPase domain and ATP hydrolysis (Achar et al., 2011; Achar et al., 2015; Blastyak et al., 2010; Kile et al., 2015). However, studies with other DNA substrates, such as lagging strand
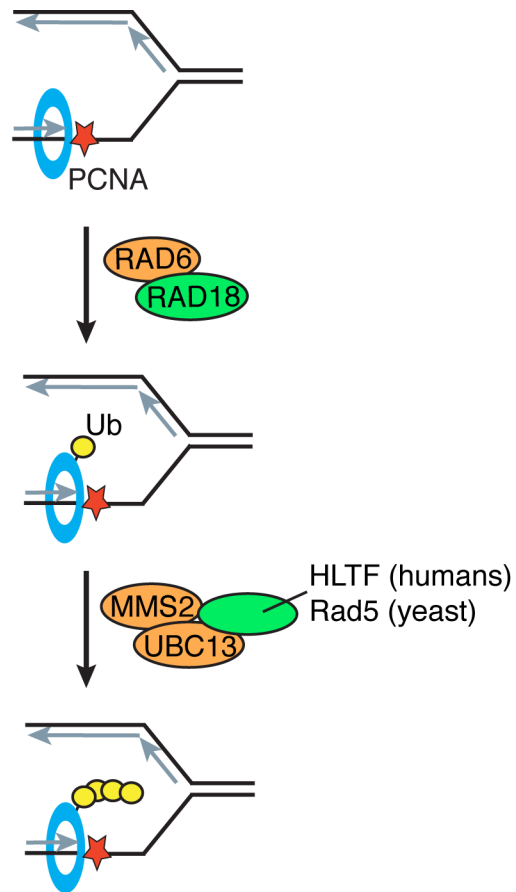
Figure 1.7. PCNA polyubiquitination. PCNA is monoubiquitinated by the RAD6/RAD18 complex after the replication fork stalls. Polyubiquitin chains are added by MMS2/UBC13 with the ubiquitin ligase Rad5 (yeast) or HLTF (humans). Ub, ubiquitin.

gapped substrates, have not been reported. Further, it has not yet been determined if HTLF can catalyze the restoration of a replication fork like ZRANB3 and SMARCAL1.

Interestingly, replication elongation actually proceeds faster in HLTF-deficient cells exposed to replication stress induced by low concentrations of HU compared to control cells (Kile et al., 2015). The authors of this study attribute this phenotype to a lack of HLTF-mediated fork reversal in response to replication stress since inactivation of the ATPase domain yields the same phenotype. Decreasing the frequency of fork reversal could yield faster overall elongation rates although it might come at the expense of genome stability.

Why so many fork remodeling enzymes?

SMARCAL1, ZRANB3, and HLTF are able to catalyze similar reactions *in vitro* suggesting they have some biochemical redundancies (Table 1.1). However, genetic studies indicate these enzymes do not function redundantly in cells. Co-depleting SMARCAL1 and ZRANB3 further sensitizes cells to replication stress compared to individual depletion (Ciccia et al., 2012). Also, cellular defects associated with depletion of each protein individually did not phenocopy each other (Table 1.1). SMARCAL1 loss of function resulted in increases in DNA breaks while loss of ZRANB3 or HLTF did not (Achar et al., 2015; Betous et al., 2012; Dungrawala et al., 2017). ZRANB3 deficiency in mammalian cells resulted in increased frequencies of SCEs while this phenotype was not observed with loss of SMARCAL1 or HLTF function (Ciccia et al., 2012). Additionally, loss of each protein individually yielded sensitivity to a different subset of DNA damaging agents (Table 1.1).

Several groups have conducted *in vitro* studies to begin differentiating the functions of these enzymes. The SRDs of each of these proteins confer different DNA preferences. The HIRAN

		SMARCAL1	ZRANB3	HLTF
Biochemistry	Annealing helicase?	Yes	Yes	?
	Fork remodeling activities	Fork reversal, Fork restoration	Fork reversal, Fork restoration	Fork reversal
	Other enzymatic activities		ATP-dependent endonuclease	Ubiquitin ligase
Protein Structure	Substrate recognition domain (SRD)	HARP	HARP-like?	HIRAN
	SRD DNA preference	Splayed arm junction	Splayed arm junction	ssDNA with 3'-OH
	Replication fork recruitment mechanism	RPA	Polyubiquitinated PCNA	?
	Evolutionary conservation	Metazoans, Bacteriophage	Metazoans	Metazoans to yeast
In cells	Reported locations of function	Telomeres	?	?
	Loss of function replication phenotypes	dsDNA breaks, fork restart deficiency	SCEs, fork restart deficiency	faster fork elongation with mild stress
	Drug sensitivity	HU, CPT, MMC, IR, aphidicolin	HU, CPT, MMS, MMC, IR, cisplatin	UV, MMS
	Human disease	Schimke Immunooseous Dysplasia	endometrial cancer?	colorectal cancer

Table 1. Properties of SMARCAL1, ZRANB3, and HLTF.

domain of HLTF, located at the N-terminus of the protein, specifically recognizes a 3'-hydroxyl group on the DNA substrate (Figure 1.8, Table 1.1) (Hishiki et al., 2015; Kile et al., 2015). Conversely, the DNA binding preferences of SMARCAL1 appear to be dependent on DNA structure rather than specific functional groups on the DNA (Kile et al., 2015). SMARCAL1 binds ssDNA/dsDNA junctions with much higher affinity compared to substrates composed entirely of ssDNA or dsDNA (Figure 1.8) (Betous et al., 2012; Ghosal et al., 2011; Mason et al., 2014; Muthuswami et al., 2000; Yusufzai and Kadonaga, 2008). In addition to the simple junction DNA structures, SMARCAL1 also binds a variety of other DNA substrates commonly found during DNA replication and repair including 3-way DNA junctions, 4-way Holliday junctions, a splayed arm, and dsDNA structures with internal ssDNA gaps (Betous et al., 2012). Much of these preferences are conferred through the HARP domains located in the first half of the protein. Based on functional similarities with the HARP domains of SMARCAL1, the SRD of ZRANB3 was initially classified as "HARP-like" (Yuan et al., 2012). ZRANB3, like SMARCAL1, preferentially binds DNA junctions over substrates composed of ssDNA or dsDNA, but it can also bind 3-way junctions and a splayed arm substrate (Figure 1.8) (Badu-Nkansah et al., 2016b; Ciccia et al., 2012; Yusufzai and Kadonaga, 2010). However, a crystal structure of the ZRANB3 SRD has not yet been determined so it is not clear if the HARP domains of SMARCAL1 and the SRD of ZRANB3 adopt a similar three-dimensional structure.

Despite these differences in DNA binding, all three enzymes are capable of catalyzing fork reversal *in vitro* (Achar et al., 2011; Achar et al., 2015; Betous et al., 2013a; Betous et al., 2012; Bhat et al., 2015; Blastyak et al., 2010; Kile et al., 2015; Kolinjivadi et al., 2017). When supplemental proteins, such as RPA, are added to the reaction, differences in fork remodeling preferences start to emerge. SMARCAL1 preferentially reverses replication forks containing RPA
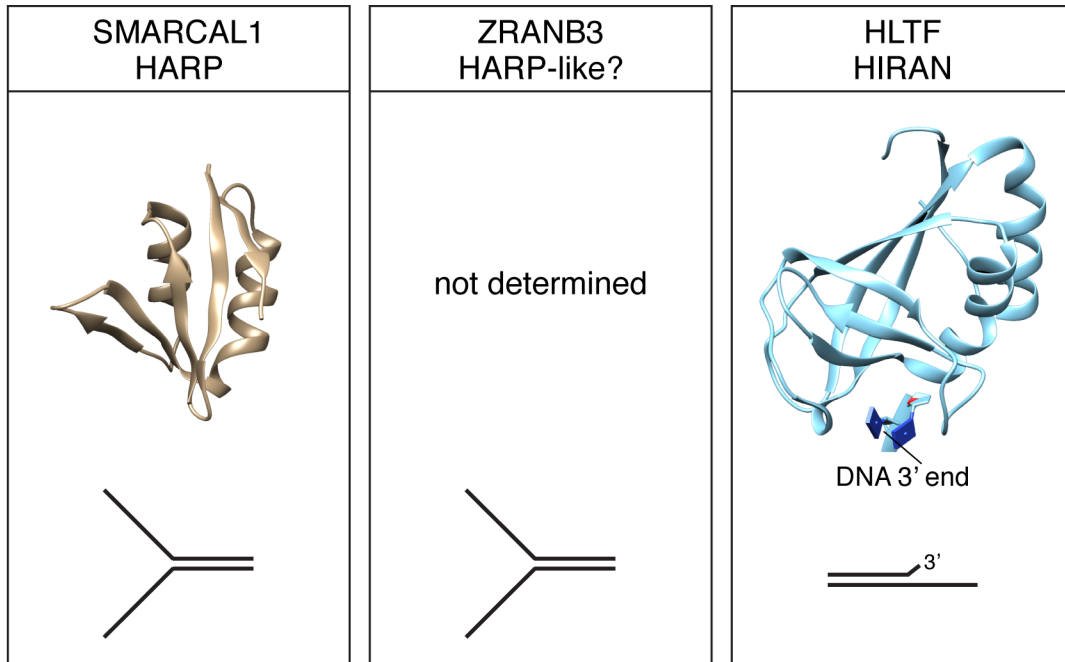
Figure 1.8. Substrate recognition domains of SMARCAL1, ZRANB3, and HLTF. Structures of the SMARCAL1 HARP and HLTF HIRAN domains were derived from PDB 4S0N and 4XZF respectively. The DNA binding specificity of the domains is illustrated. The structure of the ZRANB3 SRD domain has not been determined.

bound to a leading strand gap, a substrate mimicking a stalled replication fork, while inhibiting reversal on RPA bound to the lagging strand, a substrate mirroring a normal replication intermediate (Betous et al., 2013a; Bhat et al., 2015). In the case of ZRANB3, RPA does not stimulate any function but inhibits fork reversal when bound to leading strand gaps. Whether RPA affects fork reversal by HLTF has not been tested yet. These differences might explain the requirement for multiple enzymes capable of catalyzing fork reversal.

Much of the *in vitro* work to identify the DNA binding and substrate preferences has been done with naked DNA or in the presence of one additional protein, like RPA. The biochemical studies have not yet incorporated other proteins that might be expected to alter the enzymatic activities of the DNA translocases, such as PCNA. Also, while nucleosomes are removed from the immediate vicinity of the fork, their deposition on the newly synthesized DNA would be expected to limit the distances that replication forks can reverse and could have implications on the rates of fork restoration. Even the DNA itself is not adequately modeled biochemically since things like torsional stress are not accounted for at least in the solution biochemistry experiments. The fork is also a very crowded place since the replisome does not disassemble when the fork is stalled (Dungrawala et al., 2015). Recapitulating these complex biological conditions *in vitro* would be difficult. Therefore, it is critical to extend these comparative studies to include work in cells to address the caveats associated with *in vitro* work.

In this study, I present my work characterizing the functions of the SNF2 fork remodeling proteins SMARCAL1, ZRANB3, and HLTF in cells. I was able to identify telomeres as the first source of replication stress that requires the function of SMARCAL1 but not HLTF and ZRANB3 (Chapter III). I also adapted several unbiased DNA sequencing methods to begin identifying other regions of the genome that require the function of SMARCAL1 with the expectation of extending

these studies to HLTF and ZRANB3 (Chapter IV). These findings are significant steps in understanding why cells employ several biochemically similar enzymes. This work also provides information about the process of fork reversal – a process heavily studied with the addition of exogenous drugs – in the context of endogenous replication stress.

CHAPTER II

MATERIALS AND METHODS

Cloning

Plasmids were generated using the gateway cloning system. Validated plasmids were then transformed by adding 2μL of DNA to competent cells, incubating on ice for 10 minutes, heat shocking at 42°C for 45 seconds, and recovering in 1mL of SOC media at 37°C for 1 hour. Bacteria were then plated onto appropriate antibiotic-containing agar plates and incubated at 37°C.

Cell Culture

Cells were cultured in appropriate media listed in Table 2.1. All cells were incubated at 37°C with 5% CO_2.

Transfections

Transfections were performed according to conditions listed in Table 2.2.

Cell Viability Assays

For clonogenic assays, cells were plated at multiple dilutions depending on cell type on 6cm dishes. Cells were then treated with the appropriate drug for the indicated time and allowed to form colonies. Colonies were stained with methylene blue and counted.

For absorbance-based assays, cells were plated at different dilutions depending on cell type in 96-well dishes. Cells were treated with the indicated drug for the indicated time. Cells were then

	Cell Line	Media	Source	Notes
Human Cell Lines	293T	DMEM + 7.5% FBS		
	U2OS	DMEM + 7.5% FBS		
	AsiSI U2OS	DMEM + 10% FBS + 1mM Sodium Pyruvate + 1x GlutaMax	Gaelle Legube	Induce DSBs with 300µM 4OHT
	RPE	DMEM F12 + 7.5% FBS + 0.24% Na Bicarbonate		
	HeLa1.3	DMEM + 10% FBS	Titia de Lange	HeLa cell clone with long telomeres
	HeLa1.3 + WT SM1	DMEM + 10% FBS	Retrovirus infection with pRB84	
	HeLa1.3 + R764Q SM1	DMEM + 10% FBS	Retrovirus infection with pCC09	
	HeLa1.3 + ΔN SM1	DMEM + 10% FBS	Retrovirus infection with pRB85	
Mouse Cell Lines	SM1$^{+/+}$ p53$^{+/+}$ MEF	DMEM + 10% FBS + 2mM L-glutamine + 1x NEAA	Christine Eischen	
	SM1$^{+/+}$ p53$^{-/-}$ MEF	DMEM + 10% FBS + 2mM L-glutamine + 1x NEAA	Christine Eischen	
	SM1$^{+/+}$ p53$^{+/+}$ SV40 immortalized MEF	DMEM + 10% FBS + 2mM L-glutamine + 1x NEAA	Infection with virus from Fanning lab plasmid	
	SM1$^{\Delta/\Delta}$ p53$^{+/+}$ MEF	DMEM + 10% FBS + 2mM L-glutamine + 1x NEAA	Christine Eischen	
	SM1$^{\Delta/\Delta}$ p53$^{-/-}$ MEF	DMEM + 10% FBS + 2mM L-glutamine + 1x NEAA	Christine Eischen	
	SM1$^{\Delta/\Delta}$ p53$^{+/+}$ SV40 immortalized MEF	DMEM + 10% FBS + 2mM L-glutamine + 1x NEAA	Infection with virus from Fanning lab plasmid	
	SM1$^{\Delta/\Delta}$ MEF + WT SM1	DMEM + 10% FBS + 2mM L-glutamine + 1x NEAA	Retrovirus infection with pRB84	Made for p53-null and SV40 immortalized MEFs
	SM1$^{\Delta/\Delta}$ MEF + R764Q SM1	DMEM + 10% FBS + 2mM L-glutamine + 1x NEAA	Retrovirus infection with pCC09	Made for p53-null and SV40 immortalized MEFs
	SM1$^{\Delta/\Delta}$ MEF + ΔN SM1	DMEM + 10% FBS + 2mM L-glutamine + 1x NEAA	Retrovirus infection with pRB85	Made for p53-null and SV40 immortalized MEFs

Table 2.1. Summary of cell lines and culture conditions used for this work. DSB, double strand breaks; 4OHT, 4-hydroxy tamoxofin

	Cell Line	Forward/ Reverse	Number of cells	Transfection Protocol
siRNA Transfection	U2OS	Reverse	$3x10^5$ (6cm, 3mLs)	1.25μL Dharmafect 1 50pmol siRNA 500μL OptiMEM --incubate 15 minutes at room temperature and add dropwise to cells
	HeLa1.3	Reverse	$1x10^6$ (6-well, 2mLs)	5μL Lipofectamine RNAiMax 25pmol siRNA 500μL OptiMEM --incubate 15 minutes at room temperature and add dropwise to cells
	RPE	Reverse	$3x10^5$ (6cm, 3mLs)	10μL Dharmafect 1 40pmol siRNA 1mL OptiMEM --incubate 15 minutes at room temperature and add dropwise to cells
Plasmid Transfection	U2OS	Forward	$2x10^5$ (6-well, 2mLs)	12μL Fugene HD 3.3μg plasmid 150μL OptiMEM --incubate 15 minutes at room temperature and add dropwise to cells
	RPE	Forward	$1x10^5$ (6cm, 3mLs)	12μL Lipofectamine 2000 1μg plasmid 150μL OptiMEM --incubate 15 minutes at room temperature and add dropwise to cells
	293T (lentivirus production)	Forward	$2.5-3x10^6$ (10cm, 10mLs)	24μL PEI 3μg plasmid + 0.5μg psPAX2 + 0.5ug pMD2.G 100μL OptiMEM --incubate 15 minutes at room temperature and add dropwise to cells
	GP293 (retrovirus production)	Forward	$3-4x10^6$ (10cm, 10mLs)	24μL PEI 4μg plasmid + 2μg pVSV-G 100μL OptiMEM --incubate 15 minutes at room temperature and add dropwise to cells

Table 2.2. Summary of transfection conditions used for this work. PEI, polyethyleneimine.

released into fresh media (if indicated) and either WST or Alamar Blue was added and analyzed according to the manufacturer's instruction.

Western Blotting

Cells were lysed in NP-40 lysis buffer (1% NP-40, 5mM Tris pH 8, 200mM NaCl) supplemented with 9.1mM NaF, 1mM Na_3VO_4, 1mM DTT, 5μg/mL leupeptin, 5μg/mL aprotinin, 20mM β-glycerophosphate, and 1mM PMSF on ice for 30 minutes. Insoluble portions were removed by centrifugation at maximum speed for 10 minutes at 4°C and soluble protein concentration was determined using the Bradford Assay (Bio-rad). 2x SDS loading buffer (50μg/mL SDS, 25% glycerol, 156mM Tris pH 6.8, 12.5mg/mL bromophenol blue) was added to sample and boiled for 5 minutes. Samples were separated by gel electrophoresis on polyacrylamide gels and protein was transferred to nitrocellulose membrane at 4°C with constant current at 0.2mA for between 4-8 hours. Antibodies used for protein detection are detailed in Table 2.3. All antibodies were blocked with 5% milk diluted in 1x TBST and diluted in 1% milk in TBST. Blots processed by Reliablot were blocked and antibodies were diluted according to manufacturer's instructions.

Nuclear Extract Preparation and Immunoprecipitation (IP)

Cells were harvested by scraping and pelleted by centrifugation at 1200rpm at 4°C for 5 minutes. Cell pellets were resuspended with 5 times the cell pellet volume of hypotonic buffer (10mM HEPES pH 7.9, 1.5mM $MgCl_2$, 10mM KCl, 0.2mM PMSF, 0.5mM DTT) and cells were pelleted by centrifugation at 3000rpm. Cells were then resuspended with 3 times the cell pellet volume of hypotonic buffer, incubated on ice for 10 minutes, then homogenized with ten up-and

	Antibody	Species	Dilution	Company	Catalog Number
Western Blot	SMARCAL1 909	Rabbit	1:1000	OPEN	CB custom
	SMARCAL1 193	Rabbit	1:1000	Covance	CB custom
	GFP	Mouse	1:1000	Clontech	632569
	WRN	Rabbit	1:1000	Novus	NB100-471
	ZRANB3	Rabbit	1:1000	Bethyl	A303-033A
	GAPDH	Mouse	1:10,000	Millipore	MAB374
	HA	Rabbit	1:1000	Covance	PRB101P
	HLTF	Rabbit	1:1000	Karlene Cimprich	
	SLX4	Rabbit	1:1000	Bethyl	A302-369A-1
	RPA70	Rabbit	1:1000	Cell Signaling	2267
	RPA32	Rabbit	1:10,000	Bethyl	A300-244A
	Flag M2	Mouse	1:1000	Sigma	F-3165
	STN1	Mouse	1:2000	Abcam	Ab89250
	RAP1	Rabbit	1:200 (Reliablot)	Bethyl	A300-306A
IF & Telomere IF-FISH	GFP	Rabbit	1:2000	Abcam	Ab290
	TRF2	Mouse	1:800	Imgenex	IMG-124A
	PML	Mouse	1:250	Santa Cruz	SC-966
	RPA32	Mouse	1:1000	Abcam	ab2175
	53BP1	Mouse	1:500	Millipore	MAB3802
ChIP	HA	Rabbit	3µg per 2 15cm dishes	Abcam	Ab9110
	RPA32	Mouse	10µg per $10x10^6$ cells	Millipore	NA19L
	IgG2a	Mouse	10µg per $10x10^6$ cells	Abcam	ab18413
IP	Flag M2	Mouse	3µg/IP	Sigma	F-3165
	IgG Mouse Control	Mouse	3µg/IP	Jackson Labs	015-000-003
	SMARCAL1 909	Rabbit	3µg/IP	OPEN	CB custom
	RAP1	Rabbit	3µg/IP	Bethyl	A300-306A
	IgG Rabbit Control	Rabbit	3µg/IP	Santa Cruz	SC-2027

Table 2.3. Summary of antibodies used as part of this work.

down strokes with a glass Dounce homogenizer. Nuclei were isolated by centrifugation at 3300xg for 15 minutes at 4°C. Supernatant from this step is reserved for cytoplasmic extract sample for western blotting. Nuclei were then resuspended in 0.5 times the packed nuclei volume of low salt buffer (20mM HEPES pH 7.9, 25% glycerol, 1.5mM $MgCl_2$, 20mM KCl, 0.2mM EDTA, 0.2mM PMSF, 0.5mM DTT). 0.5 times the packed nuclei volume of high salt buffer (20mM HEPES pH 7.9, 25% glycerol, 1.5mM $MgCl_2$, 1.2M KCl, 0.2mM EDTA, 0.2mM PMSF, 0.5mM DTT) was added dropwise to isolated nuclei. Nuclei rotated at 4°C for 30 minutes before centrifugation at 13,200rpm at 4°C for 30 minutes. Nuclear extracts were dialyzed at 4°C for 1 hour in dialysis buffer (20mM HEPES pH 7.9, 20% glycerol, 10mM KCl, 0.2mM EDTA, 0.2mM PMSF, 0.5mM DTT) with constant agitation by a stir bar. Dialyzed nuclear extracts were either prepared for immunoprecipitation or flash frozen with liquid nitrogen and placed at -80°C.

Protein concentrations of nuclear extracts were determined by Bradford assay. Nuclear extracts were rotated with the appropriate antibodies (indicated in Table 2.3) for 1 hour at 4°C. Beads were then prepared by washing with 45μL of bead slurry with dialysis buffer 3 times at room temperature. Lysate-antibody solution was added to washed beads and rotated at 4°C for 30 minutes. Beads were then washed with dialysis buffer 3 times, transferred to a new tube, and eluted with either incubation with 0.25mg/mL of Flag peptide for 1-1.5 hours with agitation by gentle flicking or by boiling for 5 minutes.

C-circle Assay

Cells were transfected with the indicated siRNAs and harvested 72 hours later. Genomic DNA was harvested from the cells using Promega Wizard genomic DNA isolation kit. A maximum of 5μg of DNA was digested with 10U/mL of HinfI and 10U/mL of RsaI in digestion buffer (1x

Buffer 4, 1x BSA, 2ng/mL RNase A) for 3 hours at 37°C. Digested DNA was diluted to 3ng/μL. 10μL of diluted, digested DNA was combined with 10μL of C-circle assay mix (1x Φ29 Buffer, 0.2mg/mL BSA, 0.1% Tween-20, 1mM dATP, 1mM dGTP, 1mM dTTP, 10U/mL Φ29 DNA Polymerase) and incubated at 30°C for 8 hours then 65°C for 20 minutes. Samples were diluted to 200μL with 1x SSC buffer and loaded onto dot blot. DNA was crosslinked to nylon membrane with $1200 \times 100 \mu J/cm^2$ from a VWR UV crosslinker. Blot was pre-hybridized at 37°C with blotting buffer (5x SSC buffer, 5x Denhardt's Solution, 0.5mM $Na_4P_2O_7$, 10mM Na_2HPO_4) for 2 hours. Blot was then hybridized in blotting buffer supplemented with ^{32}P-end labeled $(AACCCT)_4$ telomere probe overnight at 37°C. Blot was washed 3 times with 4x SSC at 37°C for 30 minutes each and 1 time with 4x SSC with 0.1% SDS for 30 minutes. Blot was dried and exposed to phosphorimager screen overnight at room temperature. Image was analyzed using Quantity one and Biorad Phosphorimager.

Immunofluorescence

Cells were grown to sub-confluence on coverslips and washed with 1x PBS. Cells were extracted with 0.5% Triton-X100 in PBS for 5 minutes on ice and washed twice with 1x PBS for 5 minutes each. Cells were then fixed with 3% paraformaldehyde 2% sucrose solution for 10 minutes at room temperature and washed twice with 1x PBS for 5 minutes each. Cells were blocked with 5% BSA in PBS for 30 minutes at room temperature. Coverslips were then incubated with primary antibody diluted in 1% BSA in PBS for 30 minutes to 1 hour at room temperature. Coverslips were then washed with 1x PBS three times and incubated with secondary antibody diluted in 1% BSA in PBS for 30 minutes at room temperature protected from light. Samples were

then washed with 1x PBS three times before being mounted on slides with Prolong Gold (Thermo Fisher) containing DAPI.

IF-Telomere FISH

Cells were grown to sub-confluence on coverslips and washed with 1x PBS. Cells were extracted with 0.5% Triton-X100 in PBS for 5 minutes on ice and washed twice with 1x PBS for 5 minutes each. Cells were then fixed with 3% paraformaldehyde 2% sucrose solution for 10 minutes at room temperature and washed twice with 1x PBS for 5 minutes each. Coverslips were incubated with blocking solution (1mg/mL BSA, 3% v/v goat serum, 0.1% Triton-X100, 1mM EDTA pH 8) for 30 minutes at room temperature. Primary antibody diluted in blocking solution was added to each coverslip and incubated at room temperature for 1 hour. Coverslips were then washed three times with 1x PBS for 5 minutes each then incubated with secondary antibody diluted in blocking solution for 30 minutes at room temperature. Coverslips were then washed with 1x PBS three times before being fixed again with 3% paraformaldehyde 2% sucrose for 5 minutes at room temperature. Cells were washed twice with 1x PBS for 5 minutes each then dehydrated with 70%, 90%, and 100% ethanol, consecutively, for 5 minutes each. Coverslips were inverted on microscope slides, covered with 50μL of hybridization solution (70% v/v formamide, 0.5% Roche blocking reagent, 10mM Tris pH 7.2, 100nM Cy3-TelC probe), and denatured on 80°C heat block for 3 minutes. Slides hybridized overnight at room temperature. Coverslips were then washed in washing solution (70% v/v formamide, 10mM Tris pH 7.2) twice for 15 minute each and 1x PBS three times for 5 minutes each with DAPI supplemented in the second wash. Cells were dehydrated with 70%, 90%, and 100% ethanol, consecutively, for 5 minutes each then air dried at room temperature for 10 minutes. Coverslips were mounted on slides with Prolong Gold.

Samples pulsed with 10μM 5-ethynyl-2'-deoxyuridine (EdU) were conjugated to an anti-EdU antibody using click chemistry prior to incubation with the primary antibody by incubating coverslips with click reaction solution (5μM AlexaFluor-488/AlexaFluor-594 Azide, 2mg/mL sodium ascorbate, 2mM copper sulfate diluted in 1x PBS) for 30 minutes protected from light.

Chromosome Orientation FISH (CO-FISH)

Cells were incubated with 7.5μM BrdU/2.5μM BrdC for 16-20 hours. 1.5-2 hours before harvest, 0.5μg/mL of KaryoMAX colcemid was added to the media. Cells were harvested (including the media on cells) and resuspended in 0.075M KCl. Cells were incubated at 37°C for 15 minutes with frequent inversion to ensure cells remain in solution. Cells were pelleted at 1000rpm for 5 minutes and supernatant was removed and discarded. 10mL of fresh fixative was added to the cells in 0.5mL increments with constant gentle swirling and incubated at 4°C overnight. Cells were then pelleted and resuspended in 1mL of fixative. Cells were dropped onto wet slides and dried overnight. Cells were rehydrated in 1x PBS for 5 minutes then incubated with 0.5mg/mL of RNase A for 10 minutes at 37°C. Slides were stained with 0.5μg/mL of Hoechst 33258 for 15 minutes at room temperature then exposed to $5.4 \times 10^3 J/m^2$ of 365nm UV light. BrdU/BrdC-substituted DNA was digested away with 10U/μL of Exonuclease III for 10 minutes at room temperature. Slides were then rinsed in 1x PBS then dehydrated with 70%, 90%, and 100% EtOH for 5 minutes each at room temperature. Slides were then hybridized for 2 hours at room temperature protected from light with 100nM of Alexa488-TTAGGG probe after quick denaturation for 5 minutes at 90°C. Cells were rinsed twice for 30 minutes each in wash solution 1 (10mM Tris pH 7.2, 70% formamide, and 0.1% BSA) and 3 times for 5 minutes each with wash solution 2 (0.1M Tris pH 7.2, 0.15M NaCl, 0.08% Tween-20). Slides were hybridized for 2 hours

with 100nM of Cy3-CCCTAA probe protected from light at room temperature. Slides were then rinsed in wash solution 1 twice for 30 minutes each and 3 times with wash solution 2 supplemented with 170ng/μL of DAPI in the second wash. Slides were dehydrated with 70%, 90%, and 100% ethanol for 5 minutes each. Slides air-dried at room temperature protected from light and were mounted with Prolong Gold embedding medium.

Telomere Restriction Fragment Analysis

Genomic DNA was isolated from cells using Promega Wizard gDNA kit and digested with 10U/mL of HinfI and 10U/mL of RsaI in digestion buffer (1x Buffer 4, 1x BSA, 2ng/mL RNase A) for 3 hours at 37°C. DNA was then separated by pulsed field gel electrophoresis on a 1% pulsed field gel electrophoresis grade agarose. Pulsed field gel was run with the following conditions: 5-30kb run, calibration factor 1, 0.5x TBE, 14°C, 1% LE agarose; F voltage gradient 9V/cm, Int. Sw. Tm=0.11S, F Sw. Tm=0.46s; R Voltage gradient 6V/cm, In. Sw. Tm=0.11s, F in Sw. Tm=0.46s; run time 20 hours 30 minutes; F ramp a linear, R ramp a linear. Following the completion of the run, the gel was washed twice for 15 minutes each with 0.25M HCl and 0.5M NaOH 1.5M NaCl with a quick rinse in water after the second HCl wash. The gel was then washed once for 30 minutes in 0.5M Tris pH7.4 1.5M NaCl. DNA was transferred to a nylon membrane via capillary action overnight at room temperature. The following day, the membrane was quickly rinsed in 2x SSC buffer, air-dried, and DNA was crosslinked to the membrane with $1200x100\mu J/cm^2$ on a VWR UV crosslinker. Membrane was prehybridized with Church buffer (0.5M $NaPO_4$, 1mM EDTA, 7% SDS, 1% BSA) for 45 minutes at 65°C. Membrane hybridized to Sty11 telomere probed overnight rotating in a hybridization oven at 65°C. The following day, the membrane was

washed 3 times for 15 minutes each with wash buffer (0.1xSSC buffer, 0.1% SDS). Membrane was dried and exposed to phosphoimager screen and analyzed using Quantity One-1D.

Telomere probe was prepared by digesting the Sty11 plasmid (obtained from Dr. Titia De Lange Lab) with EcoR1-HF and gel purification of the small digested fragment. Sty11 probe was then radiolabeled by combining 200ng of probe with 5ng of $(CCTAAA)_3$ oligo, and 24μL of water. Solution was boiled at 100°C for 5 minutes and flash cooled on ice. OLB buffer (0.5M Tris pH 6.8, 0.1M MgOAc, 1mM DTT, 0.5mg/mL BSA) was added to the probe solution with 5U of Klenow polymerase and 0.6mM dATP, dGTP, and dTTP, and 50μCi of ^{32}P-alpha-dCTP. Reaction was incubated at room temperature for 90 minutes and purified using a G50 column. Probe was boiled for 5 minutes prior to be added to Church buffer for overnight hybridization with the membrane.

Radiolabeled ladder was prepared by incubating 5μL of DNA ladder (either High Range ladder from ThermoScientific or 1kb DNA ladder from NEB) with 1x Antarctic Phosphatase buffer and 5U of Antarctic Phosphatase for 3 minutes at 37°C and 65°C for 5 minutes. Reaction volume was adjusted to 100μL and DNA ladder was purified with phenol:chloroform, ethanol precipitated, and resuspended in 30μl of 10mM Tris 0.5mM EDTA. Ladder was labeled by incubating 50μCi ^{32}P-gamma-ATP, 1x T4 PNK buffer, 10U T4 PNK, and 5μL of the Antarctic phosphatase treated ladder at 37°C for 1 hour. Reaction was supplemented with 30μL of water and purified with a G23 column. 20μL of labeled high range ladder was combined with 5μL of 1kb ladder and diluted with 6x DNA loading dye.

Telomere Chromatin Immunoprecipitation (ChIP)

$50\text{-}60\text{x}10^6$ cells were grown to sub-confluence on 15cm dishes. Cells were then fixed with 15mL of 1% formaldehyde diluted in 1x PBS for 30 minutes at room temperature. Formaldehyde was quenched with 176.5mM glycine for 5 minutes at room temperature. Cells were washed with cold 1x PBS twice and harvested by scraping in 10mLs of 1x PBS. Cells were pelleted by centrifugation at 1000rpm for 5 minutes at 4°C twice. Cell pellets were flash frozen with liquid nitrogen and placed at -80°C until ready for use.

Frozen cell pellets were thawed on ice then resuspended in 2.2mLs of lysis buffer (0.01% SDS; 10mM EDTA, pH 8; 50mM Tris, pH 8) containing protease inhibitors (1mM PMSF, 1µg/mL of leupeptin, aprotinin). Cells were lysed on ice for 15 minutes then sonicated with the following parameters: amplitude 20, power 7W, 10 cycles of 20 seconds of sonication per sample. Samples were transferred to multiple 1.7mL Eppendorf tubes and insoluble fragments were removed by centrifugation at maximum speed for 10 minutes at 4°C. 1mL of soluble sample was added to 200µL of dilution buffer (0.1% SDS; 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris, pH 8.0; 150mM NaCl) and incubated on ice for 10 minutes. 100µL of diluted lysate was reserved for IP input.

Flag M2 beads were washed with IP dilution buffer. 15µL of beads were added to lysates and incubated overnight on a rotator at 4°C. Beads were then washed with the following buffers for 2-3 minutes on ice:

(1) 1mL Buffer A (0.1% SDS; 1% Triton X-100; 2mM EDTA, pH 8; 150mM NaCl) with protease inhibitors

(2) 1mL of Buffer B (0/1% SDS; 1% Triton X-100; 2mM EDTA, pH 8; 20mM Tris, pH 8; 500mM NaCl)

(3) 1mL of Buffer C (0.25M LiCl; 1% NP-40; 1% Na-Deoxycholate; 1mM EDTA, pH 8; 10mM Tris, pH 8)

(4) 1mL of TE buffer (10mM Tris, pH 8; 1mM EDTA, pH 8)

DNA was eluted by adding 250µL of 1% SDS, 0.1M NaHCO$_3$ and rotating at room temperature for 10 minutes. This was repeated twice to yield a total final volume of 500µL. Input samples were diluted to 500µL with 1% SDS, 0.1M NaHCO$_3$. Samples were supplemented with 20µL of 5M NaCl and incubated at 65°C overnight. Samples were supplemented with 10µL 0.5M EDTA, 20µL of 1mg/mL RNase A, and 20µL 1M Tris, pH 6.5 and incubated for 1 hour at 37°C. DNA was isolated from phenol:chloroform:isoamyl alcohol extraction and isolated through ethanol precipitation. DNA was rehydrated with 100µL of elution buffer from Qiagen.

Resuspended DNA was loaded onto nylon membrane. DNA on the membrane was denatured with 1.5M NaCl, 0.5M NaOH solution for 10 minutes at room temperature. Membrane was then neutralized with 1M NaCl, 0.5M Tris, pH 7. Membrane was dried and DNA was crosslinked with 1200x100µJ/cm^2. Membrane was probed in a manner similar to TRF analysis. Blot was washed four times with 2x SSC buffer for 5 minutes at room temperature. Membrane was dried and exposed to phosphoimager screen for 2 hours and analyzed using Quantity One-1D.

RPA ChIP

Cells were synchronized in early S phase with a double thymidine block (2mM thymidine overnight). Cells were then released into media containing either 300nM 4-hydroxy tamoxifen (4OHT) or an ethanol vehicle control for 4 hours. Cells were then rinsed once in 1x PBS and proteins were crosslinked to DNA with 1% formaldehyde in PBS for 10 minutes at room temperature. Formaldehyde was quenched with 0.12M glycine and cells were harvested by

scraping. Cells were washed twice with 1x PBS and processed for ChIP or flash frozen in liquid nitrogen and placed at -80°C.

$1.2x10^8$ cells were used for each ChIP. Cells were lysed with 6mL of cell lysis buffer (10mM Tris pH 7.5, 10mM NaCl, 0.2% NP-40) supplemented with 9.1mM Nave, 1mM Na_3VO_4, 5μg/mL leupeptin, 5μg/mL aprotinin, 20mM β-glycerophosphate, and 1mM PMSF on ice for 10 minutes. Cells were pelleted at 1000xg for 5 minutes at 4°C. Isolated nuclei were resuspended in 2mL cell resuspension buffer (35mM Tris pH 8.0, 6mM EDTA, 0.505% SDS, 150mM NaCl, 1% Triton). Cells were split into 2 aliquots of 1mL and sonicated in 15mL polystyrene conicals with 15mL metal adaptors on low for 3 cycles of 10 minutes (30 seconds on, 30 seconds off). Insoluble components were removed by centrifugation in 1.5mL Eppendorf tubes at 4°C at max speed. Lysates were then diluted with dilution buffer (35mM Tris pH 8.0, 6mM EDTA, 150mM NaCl, 1% Triton) to a final volume of 10mLs. 50μg of RPA32 antibody (Millipore, NA19L) was pre-bound to 200μL of Protein A Dynabeads by rotation for 1 hour at room temperature. Immunoprecipitations (pre-bound Dynabeads with lysates) were rotated overnight at 4°C in 15mL conicals.

Beads were washed the following day with the buffers below:

(1) Twice with 1mL of Wash Buffer 1 (35mM Tris pH 8.0, 6mM EDTA, 0.1% SDS, 150mM NaCl, 1% Triton)

(2) Twice with 1mL of Wash Buffer 2 (35mM Tris pH 8.0, 6mM EDTA, 0.1% SDS, 300mM NaCl, 1% Triton)

(3) Twice with 1mL of LiCl Buffer (250mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate)

(4) Once with 1mL of TE/Triton buffer (10mM Tris pH 8.0, 1mM EDTA pH 8.0, 0.2% Triton X-100)

(5) Once with 1mL of TE buffer (10mM Tris pH 8.0, 1mM EDTA pH 8.0)

Crosslinks between protein and DNA were reversed by incubation at 65°C overnight in 300mM NaCl followed by a 30 minute incubation at 37°C with 200μg/mL RNase A and a 1 hour incubation at 45°C in crosslink reversal solution (9.3mM EDTA, 37mM Tris pH 6.8, 92.6μg/mL Proteinase K) and DNA was isolated by phenol:cholorform:isoamyl alcohol extraction and ethanol precipitation overnight.

Quantitative PCR (qPCR) analyses were conducted using SYBR green (BioRad) according to manufacturer's instructions. Primer sequences are as follows:

Primer Set Name	Forward Primer Sequence	Reverse Primer Sequence
SMARCAL1	ACAGCATCAGAGGACTAGCTC	CACTGGCTTACAAGACTCCCT
ETAA1	TTCAGGTAATTATCTCTGGGAAAAA	GCATTTTAAAGCATTACCTTACTTCG
DSBIII	GGGACAGCGCGTACTTTG	TCGCTAGGCCCAGCAGTT
DSBVI	ACCCTTTGGCTGACCTTACC	TGAGGTTTTTGGTGGTGGTT
DSB2	ATCGGGCCAATCTCAGAGG	GCGACGCTAACGTTAAAGCA

For samples submitted for sequencing, libraries were prepared as follows:

1. DNA polishing: Ethanol precipitated ChIP DNA (8μL) was incubated with 1x NEB buffer 2, 50μg/mL of BSA, 150μM dNTPs, and 1.5U T4 DNA Polymerase for 20 minutes at 12°C followed by 20 minutes at 75°C.

2. DNA end phosphorylation: The polished DNA was then supplemented with 1x T4 DNA ligase buffer and 5U T4 PNK in a total volume of 20μL. The reaction was incubated for 30 minutes at 37°C followed by 20 minutes at 65°C.

3. A-tailing: Phosphorylated DNA was incubated with 1x NEB buffer 2, 100μM dATP, and 5U of Klenow DNA Polymerase in a total volume of 30μL for 30 minutes at 37°C followed by 20 minutes at 75°C.

4. Adaptor Ligation: A-tailed DNA fragments were supplemented with 1x T4 Ligase Buffer, 75μg/mL of BSA, 150pmol of pre-annealed TruSeq Adaptors, and 800U T4 DNA Ligase overnight at 16°C followed by 20 minutes at 75°C. Adaptors were pre-annealed by combining 5nmoles of each oligo diluted in STE buffer (10mM Tris pH 8.0, 50mM NaCl, 1mM EDTA pH 8.0), heating to 95°C for 4 minutes, immediate transfer to 65°C and addition of 1μL of 100mM $MgCl_2$ for 3 minutes, then transfer to room temperature for 1 hour.

5. AMPure Bead Clean-up: DNA fragments with ligated Adaptors were separated from adaptors using AMPure beads according to manufacturer's protocol using a ratio of 1:1 library-beads.

6. PCR Amplification: DNA library was then amplified using Q5 DNA polymerase according to manufacturer's protocol using the following TruSeq indexed Primers.

Forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Reverse 1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
Reverse 2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATC
Reverse 3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
Reverse 4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
Reverse 5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
Reverse 6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
Reverse 7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
Reverse 8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
Reverse 9	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

Reverse 10   CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

Reverse 11   CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

Reverse 12   CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

ChIP-seq data was analyzed as described in the "AutomatedDNASeqAnalysis" folder on github. Peaks were called with MACS2 using default parameters (-f BAM, --keep-dup auto, --qvalue 0.01, -g 2.7e9).

RNA Sequencing (RNAseq).

RNA was isolated using the Aurum total RNA mini-kit (Biorad). Messenger RNA was isolated using PolyA selection. RNAseq analysis was carried out as described in the "AutomatedRNASeqAnalysis" folder on github.

(https://github.com/lisatothepoole/cortez_lab_rnaseq_analyses.git)

DNA End Sequencing (END-seq).

Cells were harvested by trypsinization and pelleted by centrifugation at 400xg for 7 minutes and washed twice with 1x PBS. 2% low melting agarose (Biorad) was melted in 70°C water bath then equilibrated in 43°C water bath for at least 10 minutes prior to use. $3x10^6$ cells (U2OS specific) were resuspended in cell resuspension buffer (10mM Tris pH 7.2, 20mM NaCl, 50mM EDTA) such that the total volume of cells in buffer is 60-70μL (usually ~50μL of buffer added to cells). Cell suspension and agarose were equilibrated in 37°C water bath for 5 minutes. 37.5μL of melted agarose was added to cell suspension, mixed by pipetting up and down 15 times, then added to one well of an agarose plug cast. Plugs (1-3 per condition) were incubated at 4°C for 30 minutes then incubated with 170μL of Proteinase K (Qiagen, 158920) diluted in 2.5mL of

lysis buffer for 1 hour at 50°C with agitation every 10 minutes. Plugs were then transferred to 37°C waterbath overnight. Plugs were quickly rinsed with 1x wash buffer (50mM EDTA pH 8.0, 10mM Tris pH 8.0) 3 times and washed twice by gentle agitation on a horizontal platform shaker for 15 minutes at room temperature. Plugs were then incubated with 50μL of RNase A (Qiagen, 158924) diluted in 2.5mL of TE buffer (10mM EDTA pH 8.0, 10mM Tris pH 8.0) for 1 hour at 37°C with agitation every 10 minutes. Plugs were then quickly rinsed 3 times and washed 4 times with 1x wash buffer by gentle agitation on a horizontal platform shaker for 15 minutes at room temperature. Plugs were then shipped to collaborators for processing. END-seq analysis was carried out as described in the "AutomatedDNASeqAnalysis" folder on github.

(https://github.com/lisatothepoole/AutomatedDNAseqAnalysis.git) END-seq peaks were called using HOMER with default parameters.

Microsatellite Instability (MSI) Analysis

DNA was isolated from cells using the Promega Wizard genomic DNA isolation kit. PCR was performed with Standard Taq according to the manufacturer's protocol. Primers for MSI analysis are listed in Table 2.4.

SMARCAL1 editing by CRISPR/Cas9

SMARCAL1 was targeted for genome editing using a combination of two guide RNAs (gRNAs). Sequences for gRNAs and PCR checks are included in Table 2.5. Oligonucleotides to generate gRNAs were end phosphorylated with T4 PNK and annealed at 37°C for 30 minutes, 95°C for 5 minutes then ramped down incrementally to 25°C at 5°C/minute. Annealed gRNA oligonucleotides were diluted 1.250 the ligated into pSpCas9(BB)-2A-puro backbone using T4

Microsatellite Name	Repeat Type	Forward and Reverse Primers
Human Microsatellites		
Control		F: GTAGGTTCAAAGGGTGGGT R: GGCAACGTGACAAGGAATG
D2S123	CA/TG	F: AAACAGGATGCCTGCCTTTA R: GGACTTTCCACCTATGGGAC
D5S346	CA/TG	F: AGCAGATAAGACAGTATTACTAGTT R: ACTCACTCTAGTGATAAATCGGG
S17S250	CA/TG	F: GGAAGAATCAAATAGACAAT R: GCTGGCCATATATATATTTAAACC
SCA1	CTG/CAG	F: TGGAGGCCTATTCCACTCT R: GAGTCGGCGTATTGCATGA
ERDA1	CTG/CAG	F: GCCTCCAACCTGAATCTTGATAAC R: GTCCTTCCAGATTGCAACTTTAAATAC
NR-24	A/T	F: ATTGCTGAATTTTACCTC R: ATTGTGCCATTGCATTCCAA
BAT-25	A/T	F: TCGCCTCCAAGAATGTAAGT R: TCTGCATTTTAACTATGGCTC
Mouse Microsatellites		
GAPDH		F: ACCACAGTCCATGCCATCAC R: CCAAGTCACTGTCACACCAGA
D14mit102	CA/TG	F: CACAGAACTCCAGTCTAACTATCACA R: GAGGGTTATGAAAGTCAGCACC
D1mit36	CA/TG	F: TGAATAGATTAAGAGCCTGGAAGCT R: GAGAAATGTAGAGTCCAACCTGGTT
D3mit278	CA/TG	F: AACTACCATCTAAAACATCCTCTGTG R: AGATCCCTAGAGAAACAGAACTGG
D10mit266	CA/TG	F: ACCAACTACACAACATTGTCTTCTG R: CAGTTGCAAAGATGCCACC
D11mit258	CA/TG	F: AAACAGAGATAAACCACGGGG R: TGTGGAACTAACTCTCAGAAGGC
D14mit15	CA/TG	F: TTGGCTGCTCACTTGCAGCTCCTGT R: TTACCCTCCCCATAACTCCCAAAAG
D6mit236	TC/GA	F: ATCCTGCTCTGGCCTCTACA R: TTGTGTATACACACAGAGTGGGG
D17Nds3	TC/GA	F: TTCCTGTGGCGGCCTTATCAG R: AGACAATGGGTAACAGAGGCA
D5mit48	TA/AT	F: GACTATCATCCAAGCCAAGACC R: AAAAGACACTTTCCCTGACATAGC

Table 2.4. PCR primers to assay MSI.

DNA ligase at 37°C for 5 minutes, 23°C for 5 minutes for a total of 6 cycles. Plasmids were transformed into bacteria and DNA was harvested and checked by digestion with AgeI and BbsI. Cells were transfected according to transfection protocols in Table 2.2. 2μg/mL of puromycin was added to the transfected plates to select for transfected clones for 48 hours. Remaining cells were plated for single cell clones. Individual clones were then picked, propagated, and screened for editing by PCR and western blot.

Oligonucleotide Name	Forward and Reverse Primers
gRNA oligonucleotides		
SMARCAL1 gRNA1 (pTB14)	F: TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGCAGATTGCATCAACGTCG R: GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCGACGTTGATGCAATCTGC
SMARCAL1 gRNA2 (pLP05)	F: TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGTCTCAAACTCCCGATCTC R: GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGATCGGGAGTTTGAGAC
SMARCAL1 gRNA3 (pTB16)	F: TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGTTGACCTTCTTAGCAAGT R: GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACACTTGCTAAGAAGGTCAAC
PCR primers		
SMARCAL1 check 1	F: GTCTCACCACAACCTCCACT R: TGAGGGGAATGAATTTCCAG
SMARCAL1 check 2	F: TTCCGAGTAGCTGGTCCTTC R: GAGTCACGCTGGGAAAGAGA
SMARCAL1 check 3	F: TGAGACCCGAGGTTTCTCTC R: CTTGCCGTCCAGGTTGTAGT

Table 2.5. DNA sequences for primers needed for *SMARCAL1* **editing by CRISPR/Cas9.** gRNA seqencences for cloning into Cas9 containing vector are indicated in the top panel of the table. Final plasmids containing the indicated guide RNA are shown in parentheses. PCR primers used to check editing are listed in the lower half of the table. Expected product sizes for WT *SMARCAL1* are 755bp, 452bp, and 999bp for SMARCAL1 check primer set 1, 2, and 3, respectively.

SMARCAL1 MAINTAINS TELOMERE INTEGRITY DURING DNA REPLICATION

Introduction

Telomeres are a prime example of difficult-to-replicate sequences in the genome. Telomeres are composed of long stretches of a hexameric DNA repeat found at the ends of linear chromosomes. These regions of the genome pose several obstacles to DNA replication. The telomere repeats are prone to forming unusual DNA structures. The G-rich telomere strand, composed of the sequence TTAGGG, can form a G4-quadruplex structure that can affect progression of the replisome (Figure 3.1) (Bochman et al., 2012; Gilson and Geli, 2007). Further, the overhangs present at the ends of the telomeres loops back and invades the DNA duplex to form the telomere loop (t-loop) (Griffith et al., 1999). This structure prevents chromosome ends from being recognized as DNA breaks and being aberrantly repaired through chromosome fusions. Finally, telomeres are coated by the protein complex shelterin that binds telomeres and stabilizes the t-loop (Figure 3.1) (de Lange, 2005). Since they are also origin-poor regions, it is especially important to prevent replication fork inactivation within telomeric sequences (Sfeir et al., 2009). These obstacles at the telomere must be addressed to permit complete replication and prevent telomere instability.

Depleting SMARCAL1 from human cells causes an increased basal level of DNA damage in S-phase cells without the addition of exogenous genotoxic agents

---

[†] This chapter was adapted from Poole, L.A., Zhao, R., Glick, G.G., Lovejoy, C.A., Eischen, C.M., and Cortez, D. (2015). SMARCAL1 maintains telomere integrity during DNA replication. Proc Natl Acad Sci U S A *112*, 14864-14869.
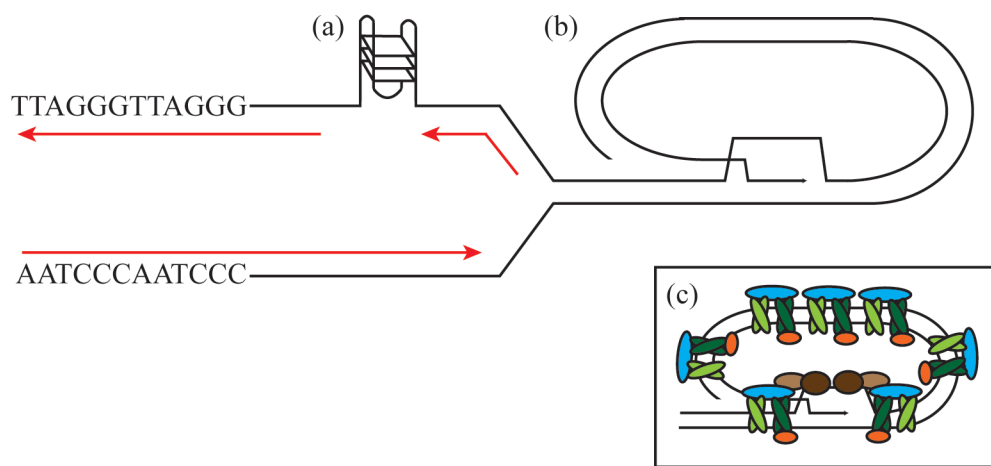
Figure 3.1. Telomeres are difficult-to-replicate DNA sequences. Several obstacles to replication are present at telomeres including (a) G4-quadruplex structures, (b) the telomere loop (T-loop), and (c) the protein complex shelterin coating the telomeres.

(Bansbach et al., 2009; Bansbach et al., 2010; Ciccia et al., 2009; Yuan et al., 2009). I reasoned that this replication-associated genome instability may reflect a requirement for SMARCAL1 to respond to endogenous forms of replication stress such as difficult-to-replicate sequences. Since telomeres present one of these replication stress sources, I tested whether SMARCAL1 is required to maintain genome integrity specifically within telomere sequences.

Results

SMARCAL1 is required to prevent accumulation of DNA damage at telomeres

First, I examined whether the increased DNA damage caused by SMARCAL1 depletion included telomere damage. Indeed, SMARCAL1 knockdown caused a significant increase in telomere dysfunction-induced foci (TIFs), as indicated by the co-localization of 53BP1 and telomeric DNA (Figure 3.2A-B), and a significant increase in co-localization of RPA with telomeres (Figure 3.2C-E). The amount of telomere damage is relatively small compared to that caused by inactivation of shelterin proteins (Sfeir et al., 2009; Takai et al., 2011), suggesting that the damage may either be transient or confined to a small subset of telomeres. These telomeric defects are not due to off-target effects of the siRNA since an siRNA-resistant wild-type (WT) SMARCAL1 cDNA is capable of complementing the siRNA-transfected cells (Figure 3.2C-E).

RPA regulates SMARCAL1 but is typically excluded from functional telomeres (Gong and de Lange, 2010). Therefore, I tested if an RPA-binding-deficient SMARCAL1 mutant (ΔN) could also rescue the increased incidence of RPA-telomere co-localization. This SMARCAL1 mutant lacks 32 amino acids at the N-terminus of SMARCAL1 that directly bind the 32C domain of RPA and are required for SMARCAL1 localization to replication forks stalled by the replication stress
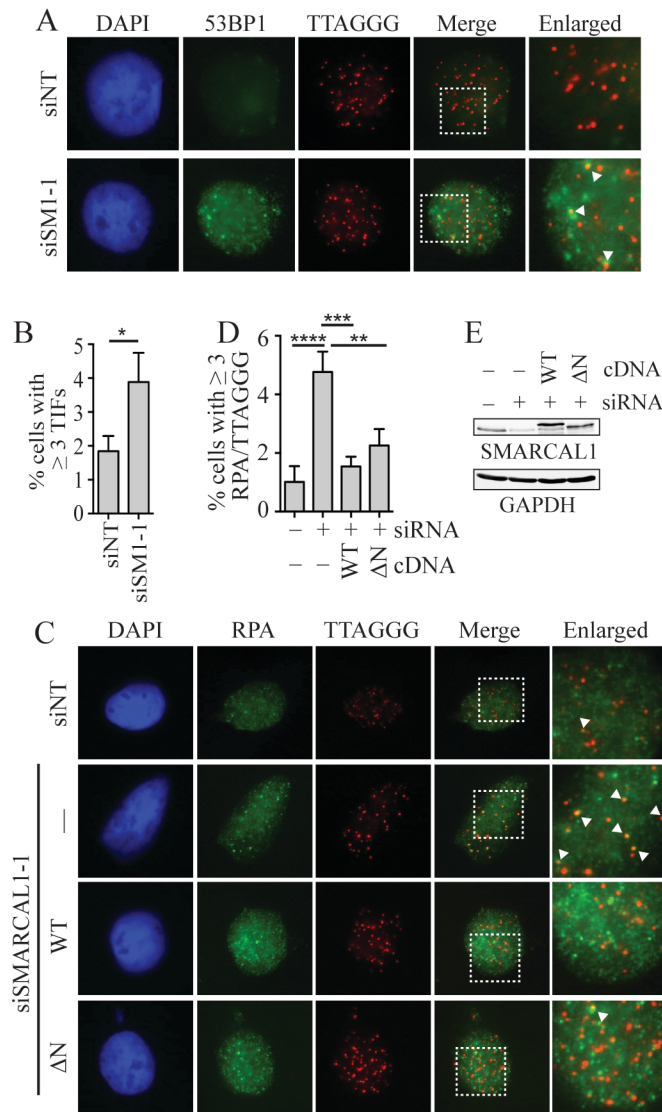
Figure 3.2. SMARCAL1 silencing causes telomere DNA damage. HeLa1.3 cells were transfected with non-targeting (NT) or SMARCAL1 (SM1-1) siRNA, fixed, and IF-FISH was performed with 53BP1 or RPA antibodies and a telomeric DNA probe, as indicated. (A) Representative images of telomere dysfunction induced foci (TIFs) in SMARCAL1-depleted HeLa1.3 cells. White arrows indicate co-localization between 53BP1 and telomeres (TIFs). (B) Quantification of cells with ≥3 TIFs. Samples were compared with a two-tailed student's t-test (p=0.02). (C-E) HeLa1.3 cells and cells expressing siRNA-resistant wild-type SMARCAL1 (WT) or RPA-interaction-deficient SMARCAL1 (ΔN) were transfected with the indicated siRNAs and imaged for RPA and telomere localization. (C) Representative images of RPA/telomere co-localization. (D) Quantification of cells with ≥3 RPA/telomere co-localized foci. Samples were compared with one-way ANOVA (p=0.0001). Bonferroni's multiple comparison test was used as a follow-up to compare siSMARCAL1 to siNT and complemented samples. (E) Immunoblot to monitor SMARCAL1 expression. Plots in (B) and (D) are mean +/- SD from three independent experiments in which approximately 100 and 500 nuclei were examined, respectively.

inducing agent hydroxyurea (HU) (Bansbach et al., 2009). Surprisingly, we also observed a significant reduction in the frequency of TIFs in cells expressing ΔN-SMARCAL1 (Figure 3.2C-E). The level of expression is approximately 2-fold higher than endogenous SMARCAL (Figure 3.2E). This level of ΔN-SMARCAL1 expression is not capable of rescuing other SMARCAL1 loss of function phenotypes (Bansbach et al., 2009; Bansbach et al., 2010; Ciccia et al., 2009; Yuan et al., 2009; Yusufzai et al., 2009). Thus, the function of SMARCAL1 at telomeres may be separable from its function in bulk chromosomal replication stress responses (see more below).

SMARCAL1 depletion causes accumulation of circular extrachromosomal telomere DNA

The increase in TIFs in SMARCAL1-deficient cells suggests that SMARCAL1 maintains telomere integrity. To further characterize this function, I examined whether SMARCAL1 deficiency caused additional telomere dysfunction phenotypes. Indeed, silencing SMARCAL1 in HeLa cells with long telomeres (HeLa1.3 cells) caused a significant increase in the abundance of extrachromosomal, partly duplexed, circular DNAs derived from telomere sequences (C-circles) (Figure 3.3A). The abundance of these C-circles correlated with the degree of SMARCAL1 knockdown although it was not as high as that seen in U2OS cells that utilize the recombination-based mechanism of telomere maintenance ALT (alternative lengthening of telomeres). The C-circles were resistant to degradation by exonuclease V, a nuclease that selectively digests linear DNA, and dependent on rolling circle amplification by phi29 DNA polymerase. I also observed an elevated level of C-circles in SMARCAL1-deficient mouse embryonic fibroblasts (*Smarcal1$^{\Delta/\Delta}$* MEFs) compared to MEFs derived from wild-type littermates (Figure 3.3B).
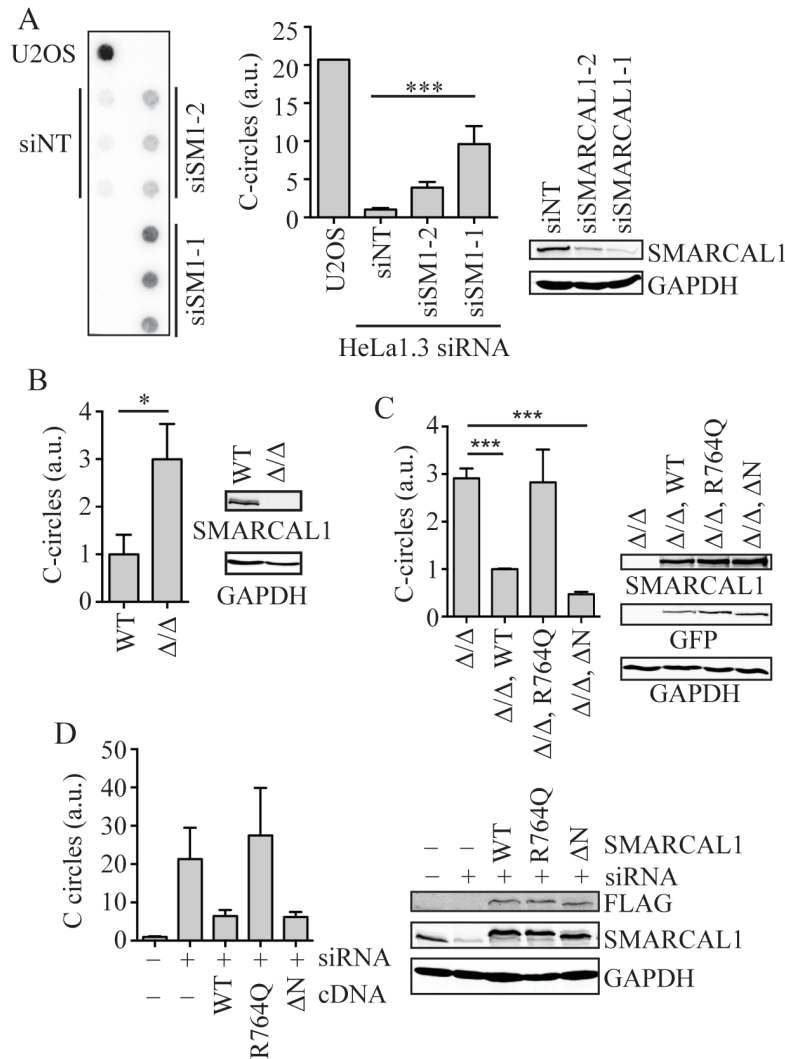
Figure 3.3. SMARCAL1 deficiency causes accumulation of extrachromosomal telomere circles. (A) C-circles were quantified from genomic DNA isolated from untransfected U2OS cells and HeLa1.3 cells transfected with the indicated siRNAs. A representative dot blot is shown. Values were normalized such that siNT was set at 1. Samples were compared with one-way ANOVA ($p<0.0001$). Bonferroni's multiple comparison test was used as a follow-up to compare siNT to siSMARCAL1 samples. Right, immunoblot analysis of SMARCAL1 protein expression in HeLa1.3 cells. (B) Quantification of C-circles from Smarcal1Δ/Δ vs WT MEFs. Values were normalized such that WT was set at 1. Samples were compared with two-tailed student's t-test ($p=0.015$). (C) Quantification of C-circle assay performed in Smarcal1Δ/Δ MEFs complemented with GFP-tagged wild-type (WT), ATPase dead (R764Q), and RPA-binding-deficient (ΔN) SMARCAL1. Values were normalized such that Δ/Δ,WT was set at 1. Error bars indicate +/- the SD from three experiments. Samples were compared with one-way ANOVA ($p<0.001$) followed by Bonferroni's multiple comparison test to compare Δ/Δ to complemented samples. Right, immunoblot analysis of SMARCAL1 expression. (D) Quantification of C-circles in complemented HeLa1.3 cells treated with SMARCAL1-1 siRNA. Samples were compared with one-way ANOVA ($p=0.01$). Error bars in panels indicate +/- the SEM from at least three experiments. Right, immunoblot analysis of SMARCAL1 expression.

Expression of human GFP-SMARCAL1 in the *Smarcal1$^{Δ/Δ}$* MEFs returned C-circle abundance to low levels (Figure 3.3C). However, an SIOD patient-derived SMARCAL1 mutant protein (R764Q) that lacks enzymatic activity did not decrease C-circle abundance (Figure 3.3C). These data indicate that, like the genome-wide function of SMARCAL1, its function at telomeres also requires hydrolysis of ATP. The N-terminal truncation mutant of SMARCAL1 that removes the RPA binding domain (ΔN) also prevented the accumulation of C-circles, again suggesting a separation of SMARCAL1 functions at telomeres from other sites of replication stress (Figure 3.3C). Similar results were also obtained with siRNA to SMARCAL1 in HeLa1.3 cells complemented WT-, R764Q-, and ΔN-SMARCAL1 (Figure 3.3D). Thus, I conclude that the function of SMARCAL1 at telomeres is dependent on its enzymatic activity but does not require an interaction with RPA.

SMARCAL1 functions during replication elongation to prevent SLX-dependent telomere processing

Previous studies investigating the origin of C-circles demonstrated that ongoing replication is a requirement for C-circle formation in ALT cells (O'Sullivan et al., 2014). I performed C-circle assays on samples treated with the replication stalling agents HU or aphidicolin to investigate the dependence on replication in a SMARCAL1-depleted setting. C-circle levels in SMARCAL1-deficient cells were reduced to baseline levels similar to that of samples treated with non-targeting siRNA after treatment with either replication stress agent for 48 hours (Figure 3.4A). Thus, while SMARCAL1 is recruited to stalled forks in response to HU and aphidicolin, it is not required to prevent extrachromosomal telomere circles when replication elongation is inhibited by these drugs.
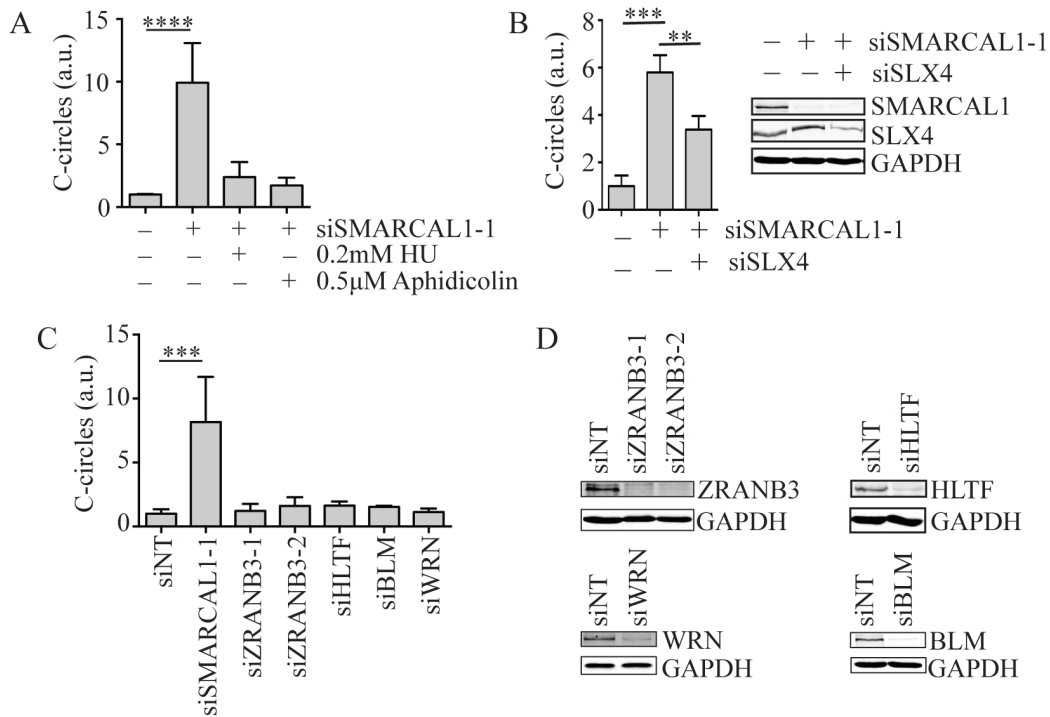
Figure 3.4. Replication-dependent C-circle formation is specific to SMARCAL1 deficiency. (A) HeLa1.3 cells were transfected with NT or SMARCAL1 siRNA and treated with 0.2mM hydroxyurea (HU) or 0.5μM aphidicolin for 48 hours prior to assessing C-circle abundance. (B and C) Quantification of C-circles after transfection of HeLa1.3 cells with the indicated siRNAs. Values were normalized such that siNT was set at 1. Error bars in all panels are SD from three experiments. Samples were compared with one-way ANOVA ((A) p<0.0001, (B) p=0.0002, and (C) p<0.001) followed by Bonferroni's multiple comparison test to compare siNT to other siRNA transfection samples. (D) Immunoblots to monitor ZRANB3, HLTF, BLM, and WRN expression.

In addition to replication, extrachromosomal telomere circle accumulation in at least some settings is also dependent on the function of the nuclease scaffold protein SLX4 (Vannier et al., 2012). To test whether SLX4 is required to generate the C-circles in SMARCAL1-deficient cells, I co-depleted SMARCAL1 and SLX4 in HeLa1.3 cells using siRNA. Although I was only able to partly silence SLX4 expression, I did observe a decrease in C-circles by approximately 50%, (Figure 3.4B). Therefore, I conclude that the C-circles generated in SMARCAL1-deficient cells are generated at least partly through telomere cleavage by an SLX4-dependent nuclease during DNA replication.

The SMARCAL1 function at telomeres is not shared by related DNA translocases

I next asked whether the telomere function of SMARCAL1 is shared by related DNA translocases, including HLTF and ZRANB3. In contrast to SMARCAL1, depletion of either ZRANB3 or HLTF from HeLa1.3 cells did not cause any change in C-circle levels (Figure 3.4C-D). We also did not observe any increase in C-circle abundance when we knocked down the RECQ helicases WRN or BLM (Figure 3.4C-D). Thus, although these enzymes are all recruited to stalled replication forks and are capable of catalyzing overlapping biochemical reactions such as fork reversal, only SMARCAL1 loss of function causes increased telomere C-circle abundance in the absence of any exogenous agents.

Telomere length and recombination in SMARCAL1-deficient settings

C-circles have previously been described as a marker of cells using the ALT pathway for telomere elongation which involves a recombination mechanism (Cesare and Reddel, 2010). In addition to C-circles and damage at telomeres, ALT cells also display other changes in telomere

integrity including increased rates of telomere recombination and dramatic changes in telomere length.

As I saw damage at telomeres and C-circles, I next examined if there is a change in telomere length in SMARCAL1-deficient MEFs. By telomere restriction fragment analysis, I saw a small but not statistically significant difference in telomere length in $Smarcal1^{\Delta/\Delta}$ mouse embryonic fibroblasts (MEFs) when compared to wild-type littermates (Figure 3.5A-B). There was also no significant difference in telomere lengths in HeLa1.3 cells depleted of SMARCAL1 with siRNA for three weeks (Figure 3.5C).

I also examined if SMARCAL1 depletion affected telomere recombination rates, another hallmark of ALT. We did not observe a significant increase in the frequency of telomere sister chromatid exchanges (t-SCEs) using chromosome orientation fluorescence *in situ* hybridization (CO-FISH) on metaphase spreads from HeLa1.3 cells depleted of SMARCAL1 (Figure 3.5D). As a control, I did find that depletion of ASF1 caused a significant increase in t-SCEs, as reported previously (Figure 3.5D) (O'Sullivan et al., 2014). Additionally, I did not observe a significant change in the percentage of telomeres with multiple telomere signals (MTS) or telomere signal free ends in SMARCAL1-depleted cells (Figure 3.5D). Similarly, no significant differences in t-SCE, MTS, or missing telomere frequencies were observed between WT and $Smarcal1^{\Delta/\Delta}$ MEFs (Figure 3.5E).

Another recombination marker of ALT cells is an increased frequency of promyelocytic leukemia (PML) co-localization with telomeric DNA. These ALT-associated PML bodies (APBs) also include homologous recombination proteins, which may coordinate these components to facilitate recombination events at telomeres (Chung et al., 2012). SMARCAL1 depletion in HeLa1.3 cells does not cause a significant increase in the co-localization of telomeric DNA and

A

WT   Δ/Δ

24.5
20.6
17
12.1
10.2

D

| | siNT | siSMARCAL1-1 | siASF1 |

HeLa1.3 siRNA	t-SCEs	MTS	Missing
siNT (n=23,860)	0.06%	0.12%	0.85%
siSM1-1 (n=24,680)	0.04%	0.15%	0.79%
siASF1 (n=26,186)	1.67%	0.05%	0.50%

B   Telomere length (kB)

n.s.

80
60
40
20
0

WT   Δ/Δ

C   Telomere Length (kB)

n.s.

30
20
10
0

siNT   siSM1

E

| | WT | Δ/Δ |

	t-SCEs	MTS	Missing
WT (n=9,992)	0.04%	0.27%	0.39%
Δ/Δ (n=7,760)	0.08%	0.17%	0.52%
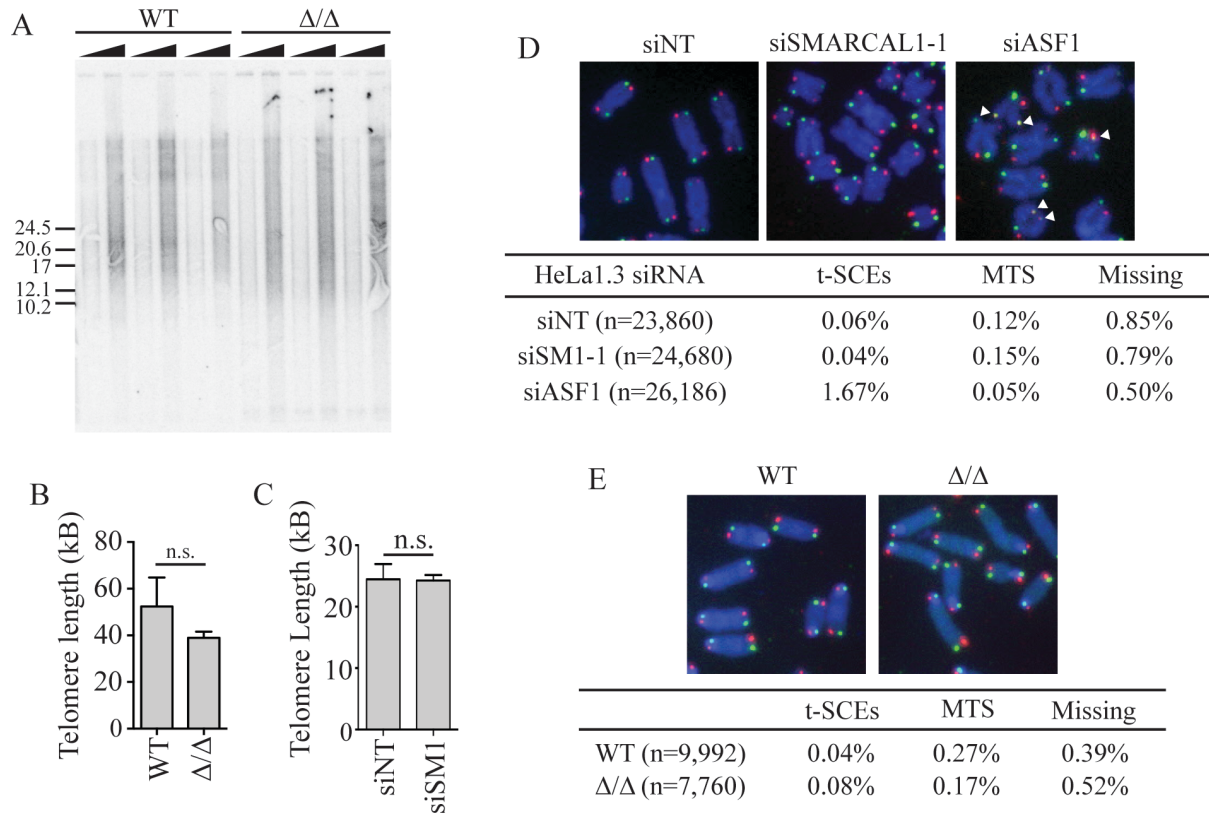
Figure 3.5. Analysis of telomere length and recombination in SMARCAL1-deficient cells. (A-C) Telomere length was analyzed in SMARCAL1 WT and Δ/Δ MEFs or HeLa1.3 cells transfected with non-targeting or SMARCAL1 siRNA every three days for a total of 7 transfections by telomere restriction fragment analysis. Two quantities of telomere DNA for each of three independent samples was separated by pulse-field gel electrophoresis prior to southern blotting. Error bars indicate +/- the SD. Significance (MEFs, p=0.14; HeLa1.3, p=0.9) was determined using two-tailed student's t-test. (D and E) Metaphase spreads from (D) HeLa1.3 cells transfected with NT, SMARCAL1, or ASF1 siRNA or (E) SMARCAL1 WT and Δ/Δ MEFs were fixed and stained with telomere probes to the G rich repeat (red) and the C rich repeat (green). Representative images are shown. White arrows indicate telomere sister chromatid exchange (t-SCE) events. Tables depict mean frequency of t-SCEs, multiple telomere signals (MTS), and signal free ends (Missing) (n= number of telomere ends analyzed).

PML protein, despite an increase in total number of PML foci (Figure 3.6). Therefore, SMARCAL1 loss is not sufficient to induce telomere recruitment to these proposed recombination centers. In contrast, almost 100% of the ALT$^+$ U2OS cells display co-localization of TTAGGG sequences with PML (Figure 3.6).

Given the lack of many ALT phenotypes in telomerase-positive HeLa cells following SMARCAL1 depletion, we proceeded to examine whether SMARCAL1-deficiency altered telomere integrity in an ALT$^+$ cell line. Following SMARCAL1 knockdown in ALT$^+$ U2OS cells, we quantified the frequency of TIFs, APBs, and RPA co-localization with telomeres in SMARCAL1 proficient and deficient settings. In all cases, loss of SMARCAL1 in U2OS cells had no significant effect on the frequency of occurrence of each marker of telomere instability (Figure 3.7A-F). Additionally, there was no difference in the abundance of C-circles between U2OS cells treated with non-targeting or SMARCAL1 siRNA (Figure 3.7G).

SMARCAL1 localization to telomeres

SMARCAL1 function in genome-wide replication is thought to be due to a direct activity at replication forks, requiring RPA-dependent localization to site of replication stress (Bansbach et al., 2009; Betous et al., 2012; Ciccia et al., 2009; Postow et al., 2009; Yusufzai et al., 2009). To test whether SMARCAL1 also localizes to telomeres and directly functions to promote telomere integrity during DNA replication, I performed telomere IF-FISH and telomere chromatin immunoprecipitation (ChIP). When GFP-SMARCAL1 is highly over-expressed in U2OS cells, I am able to see a subset of SMARCAL1 foci co-localize with the telomere probe (Figure 3.8A). However, in these circumstances, high levels of overexpressed SMARCAL1 induces DNA damage during DNA replication (Bansbach et al., 2009). When expressed at lower levels from an
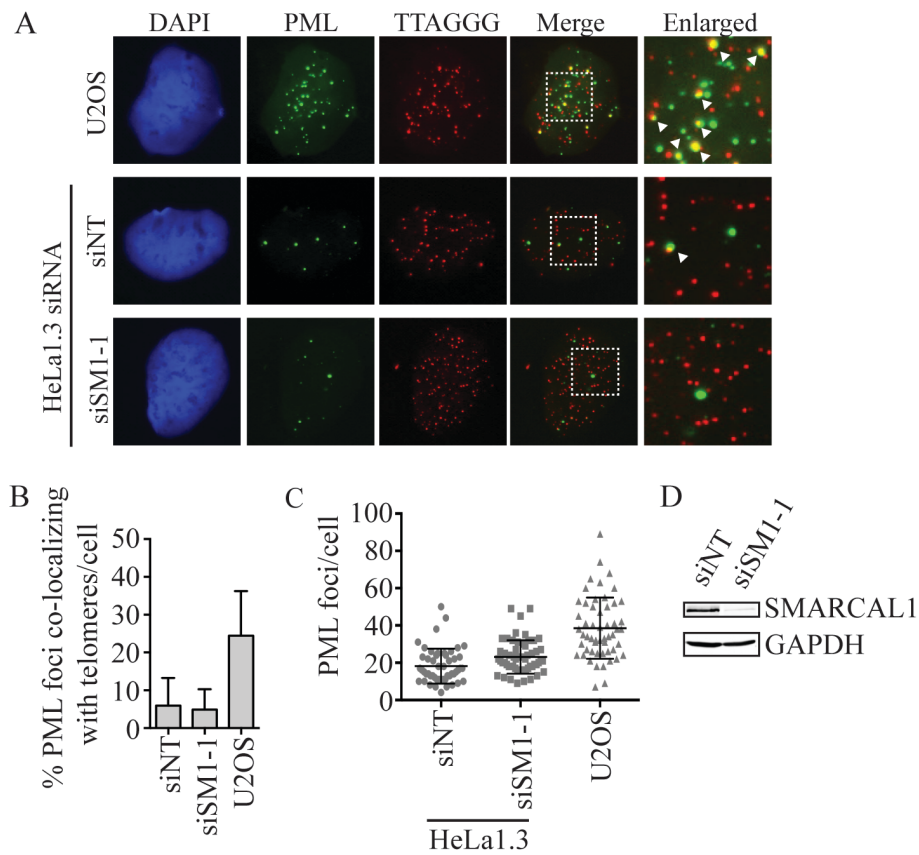
Figure 3.6. SMARCAL1 depletion does not increase PML-telomere co-localization. Non-targeting (NT) or SMARCAL1 (siSM1-1) siRNA treated HeLa1.3 cells and non-transfected U2OS cells were fixed and stained with antibodies to PML and a telomere probe as indicated. (A) Representative images of PML-telomere co-localization in SMARCAL1-depleted HeLa1.3 cells and ALT+ U2OS cells. White arrows indicate APBs. (B) Quantification of the percent of PML foci that co-localize with telomeres. Error bars indicate SD from three experiments in which approximately 100 cells were examined. (C) Dot plot of total number of PML foci per cell. (D) Immunoblot monitoring SMARCAL1 expression.
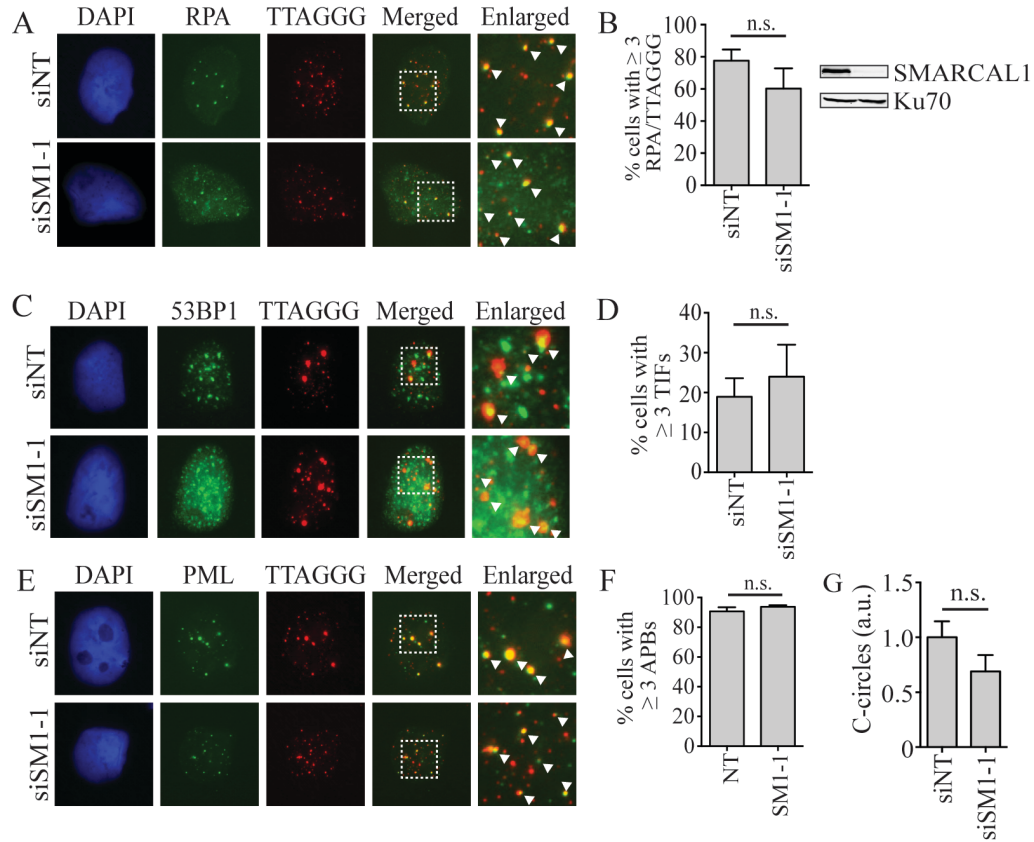
Figure 3.7. SMARCAL1 depletion in U2OS cells does not change ALT-related phenotypes. U2OS cells were transfected with non-targeting (siNT) or SMARCAL1 (siSM1-1) siRNA, fixed, and IF-FISH was performed with 53BP1, RPA, or PML antibodies and a telomeric DNA probe, as indicated. (A) Representative images of RPA/telomere co-localization in SMARCAL1-depleted U2OS cells. White arrows indicate co-localization between RPA and telomeres. (B) Quantification of cells with ≥3 RPA/telomere co-localized foci. Right, immunoblot to monitor SMARCAL1 expression. (C) Representative images of telomere dysfunction induced foci (TIFs). (D) Quantification of cells with ≥3 TIFs. (E) Representative images of ALT-associated PML bodies (APBs). (F) Quantification of cells with ≥3 APBs. (G) Quantification of C-circles in U2OS cells treated with NT or SM1-1 siRNA. Plots in (B), (D), and (F) are mean +/- SD from three independent experiments in which approximately 100 cells were counted. Plot in (G) is the mean +/- SD from three independent experiments. All statistical analyses were done with a two-tailed student's t-test and yielded p values greater than 0.05.
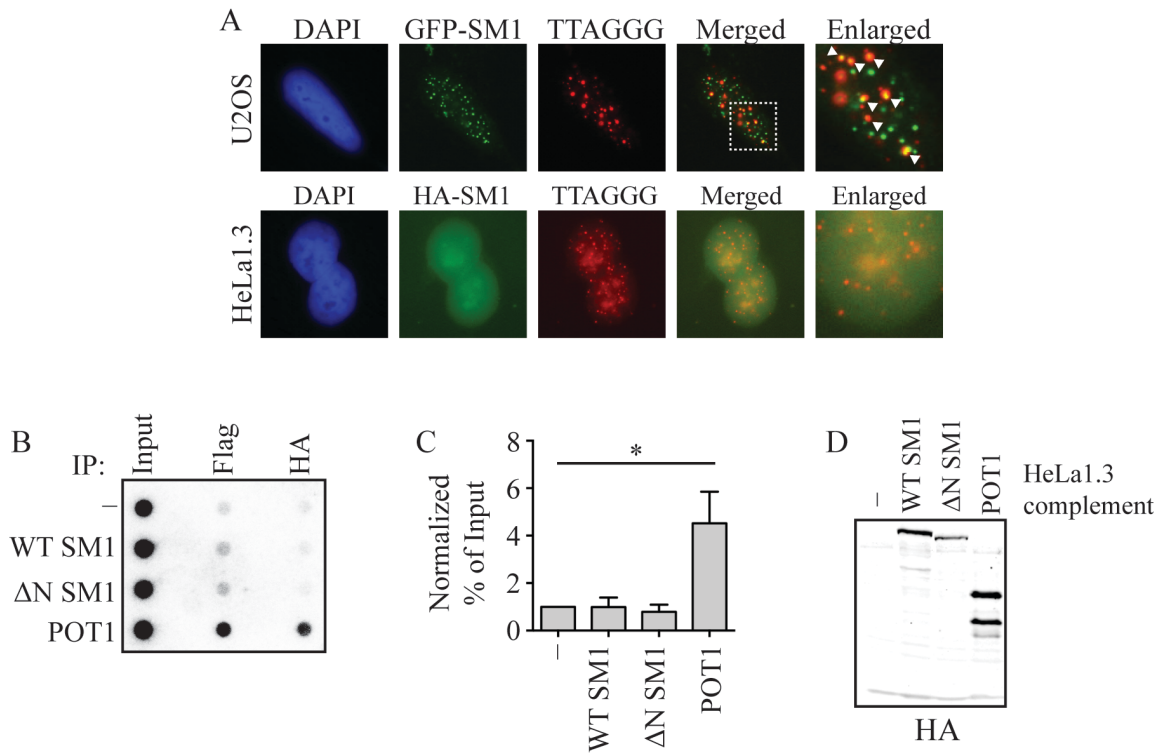
Figure 3.8. Overexpressed SMARCAL1 localizes to a subset of telomeres. (A) U2OS cells were transiently transfected with a vector that highly overexpresses GFP-SMARCAL1 prior to performing IF-FISH with GFP antibodies and a telomere DNA probe. (B) HeLa1.3 cells expressing Flag-HA-SMARCAL1 at closer to endogenous levels were fixed and IF-FISH was performed using HA antibodies and telomere DNA probe. (C) Representative dot blot for telomeric DNA from ChIP of cells expressing Flag-HA-SMARCAL1, Flag-HA-ΔN-SMARCAL1, Flag-HA-POT1, or non-complemented HeLa1.3 cells. Both Flag and HA antibodies were utilized as indicated. (D) Quantification of HA-ChIP samples. Plot indicates the mean +/- SEM from three separate experiments. Significance was determined using a one-way ANOVA (p=0.016) followed by Bonferroni's multiple comparison test to compare non-complemented HeLa1.3 to other complemented samples. (D) Immunoblot to monitor expression of Flag-HA-complemented HeLa1.3 cells.

58

alternative vector that does not cause DNA damage, no specific SMARCAL1 localization at telomeres was evident in either U2OS or Hela1.3 cells (Figure 3.8A). Furthermore, when I performed telomere ChIP in HeLa1.3 cells stably expressing Flag-HA-WT SMARCAL1, I was unable to see an enrichment of telomere DNA immunoprecipitated with SMARCAL1 compared to a control ChIP sample (Figure 3.8B-D). Conversely, the shelterin protein POT1 was significantly enriched (Figure 3.8C-E). Thus, SMARCAL1 can associate with telomeres; however, it is not detectable when expressed at close to endogenous levels suggesting a transient association perhaps with only a subset of telomeres within the cell population undergoing replication.

Discussion

Despite the utility of genotoxic agents to study the replication stress response, these drugs often fail to differentiate the functions of related fork repair enzymes and are presumed to model some undefined, endogenous source of replication stress. In the current study, I show that SMARCAL1 has a unique function to promote replication through an endogenous source of replication stress—telomeric sequences. This is the first identification of an endogenous source of replication stress that is resolved by SMARCAL1 (or any of the related enzymes).

Following SMARCAL1 inactivation, cells display markers of telomere instability including TIF formation and the accumulation of extrachromosomal, circular telomeric DNA (C-circles). Notably I did not observe the generation of C-circles after inactivation of the related enzymes ZRANB3 and HLTF. Although all three enzymes possess similar DNA remodeling capabilities *in vitro*, our studies provide a distinction among these DNA translocases in cells (see Chapter V for further discussion).

There are many additional repair enzymes including helicases like WRN, BLM, and RTEL1 that work at damaged replication forks to promote genome stability. WRN likely has a function at telomeres and can form a complex with SMARCAL1 (Betous et al., 2013b); however, WRN knockdown did not cause C-circle formation. BLM also has reported functions at telomeres but knockdown in the absence of exogenous stress did not induce C-circle accumulation (Barefield and Karlseder, 2012; Drosopoulos et al., 2015). RTEL1 has at least two essential telomere functions although we have not detected any interaction between RTEL1 and SMARCAL1 (Popuri et al., 2014; Vannier et al., 2012). It will be important to understand how these helicases and SMARCAL1 work cooperatively to achieve successful telomere replication.

Our discovery that SMARCAL1 has an important telomere stability function provides the first evidence for a specific requirement of SMARCAL1 at a site of endogenous replication stress. However, the specific source of replication stress that requires the function of SMARCAL1 remains undetermined. Biochemical data suggest it may be specific to leading strand replication problems (Betous et al., 2013a; Bhat et al., 2015). SMARCAL1-deficient MEFs and HeLa1.3 cells do not display sensitivity to the G4 quadruplex stabilizing agent TMPyP4 suggesting SMARCAL1 may be dispensable for replication through these complex DNA structures (Figure 3.9). Further characterization the telomere function of SMARCAL1 would be informative for understanding the cooperativity between SMARCAL1 and other enzymes that function at telomeres such as RTEL1 and WRN. Further, the damage that occurs following knockdown of SMARCAL1 does not occur exclusively at telomeres. Identifying the source of replication stress at telomeres would be beneficial in characterizing where SMARCAL1 functions in replication through other regions of the genome.
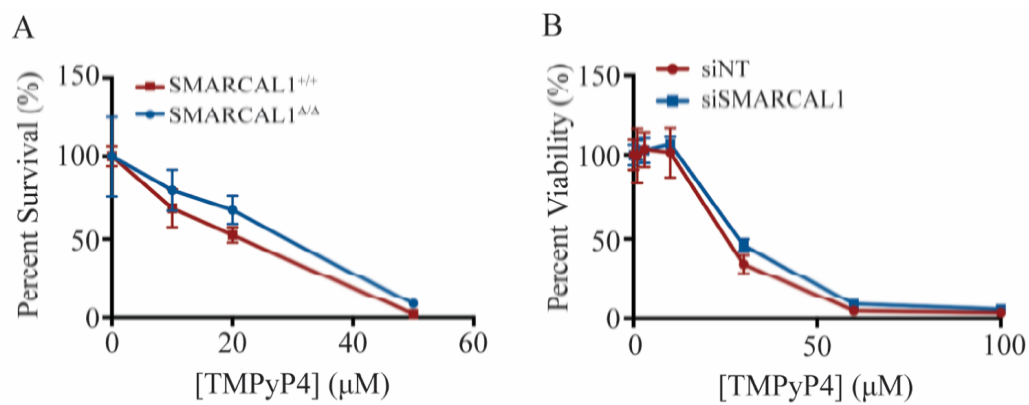
Figure 3.9. SMARCAL1-deficient cells aren't sensitive to G4-quadruplex stabilizing agents. (A) Percent of surviving litter matched SMARCAL1 WT (+/+) and mutant (Δ/Δ) MEFs treated with varying concentrations of TMPyP4 for 7 days. (B) Quantification of cell viability by Alamar Blue staining of HeLa1.3 cells treated with non-targeting or SMARCAL1 siRNA after 72 hour treatment with varying concentrations of TMPyP4.

SMARCAL1-deficient cells display a telomere instability phenotype partially reminiscent of the

ALT pathway

C-circles have previously been described as a marker of cells using the ALT pathway for telomere maintenance (Henson et al., 2009). Surprisingly, although SMARCAL1 depletion causes C-circle formation, I did not observe other ALT-related phenotypes. Specifically, I did not observe increased rates of inter-telomere recombination, co-localization of telomeric DNA with PML, or dramatic changes in telomere length in SMARCAL1-deficient cells (Figures 3.5A-E, Figure 3.6). C-circles have also been seen following loss of the ASF1 histone chaperone, but in that case other ALT-like phenotypes were also observed (O'Sullivan et al., 2014). SMARCAL1-deficiency is insufficient to generate all ALT phenotypes indicating that C-circle formation is not a sufficient indicator of ALT utilization. Likewise, SMARCAL1 depletion in the context of an ALT positive cell line has no significant effect on telomere stability (TIFs and RPA/telomere co-localization), APB frequency, or C-circle abundance (Figure 3.6) indicating SMARCAL1 is also not necessary to maintain an ALT phenotype.

After this work was published, another study identified SMARCAL1 as a telomere replication protein; however, this study limited the function of SMARCAL1 to ALT cells (Cox et al., 2016). In this report, SMARCAL1 knockdown cells also display severe telomere defects such as C-circles and DNA damage markers co-localizing with telomeres. However, the ALT-positive SMARCAL1-deficent cells also display increased frequencies of chromosome fusions and APBs and increases in telomere size that was attributed to telomere clustering (Cox et al., 2016). However, this doesn't appear to be a phenotype specific to all cells undergoing ALT as I was unable to see any of these phenotypes in U2OS cells, another ALT-positive cell line.

The authors offered no mechanism to suggest why the defects in the function of SMARCAL1 would be selective to cells using ALT. They cite that there is a higher level of undefined replication stress present at ALT telomeres that requires the activities of SMARCAL1 (Cox et al., 2016). However, a third study observed C-circle accumulation in SMARCAL1-deficient embryonic stem cells overexpressing telomerase (Rivera et al., 2017). Thus, SMARCAL1 appears to be required for replication through telomeres, independent of the telomerase maintenance mechanism.

The differences between my results and the work published in Cox, et.al. likely lies with the use of telomerase-positive cell lines with average telomere length in the latter study rather than long telomeres utilized in my research. The use of shorter telomere cell lines was problematic as some assays, specifically the C-circle assay, have lower detection limits and require the use of cells with long telomeres (Henson et al., 2009; O'Sullivan et al., 2014). Thus, the use of a cell line with shorter telomeres may have impeded their ability to see the effects of SMARCAL1 deficiency in their telomerase-positive cell lines.

Although C-circles are a major hallmark of cells utilizing the ALT pathway, it is unclear whether they are required for the recombination-based mechanism of telomere maintenance. As yet, it is unclear how C-circles are actually generated either in ALT or SMARCAL1-deficient cells although an SLX4-dependent nuclease is involved in both. SLX4 scaffolds several endonucleases that can cleave stalled or damaged replication forks (Svendsen et al., 2009). Thus, it is likely that replication forks stalled within telomeric sequences undergo aberrant processing when SMARCAL1 is inactivated. Since we do not observe an increased rate of inter-telomere recombination, it is possible the C-circles are formed due to intra-telomere processing events.

A recent study suggests C-circles result from replication stress associated with uncontrolled telomere length regulation (Rivera et al., 2017). This study proposed that C-circles form from processing of reversed replication forks that occurs in the telomere repeats in times of endogenous replication stress. As SMARCAL1 can perform this function *in vitro*, this remains a viable model for C-circle formation in SMARCAL1-deficient conditions.

Despite these recent advances in the field of C-circle biology, several other questions still remain. In addition to C-circles, there are also other telomere circles that have been identified in cells, most notably T-circles – entirely double stranded telomere circles. As yet, studies to compare the relationship between these DNA structures has not yet been completed. Knockdown of proteins such as ASF1 and SMARCAL1 yield C-circles, but the presence of T-circles has not yet been tested. Conversely, T-circles have been identified with knockdown of several proteins with known telomere functions including the telomere helicase RTEL1 (Vannier et al., 2012); however, the accumulation of C-circles in these conditions has not yet been assayed. A comparison of the proteins needed for each structure would be informative for determining the origins of these extrachromosomal circles and the cooperativity between multiple enzymes at telomeres during DNA replication.

SMARCAL1 bulk chromatin and telomere functions are separable

Although RPA is required for SMARCAL1 function in bulk chromatin replication following the addition of a genotoxic agent, RPA binding is dispensable for SMARCAL1 function at telomeres. This result is intriguing but not entirely unexpected as RPA is typically excluded from functional telomeres to prevent aberrant DNA damage response signaling (Gong and de Lange, 2010).

This RPA-independence raises the question of how SMARCAL1 localizes to sites of replication stress at telomeres. It is unclear if another protein is required to recruit SMARCAL1 to telomeres although I have failed to find any direct interaction with known telomere binding proteins. Taking a candidate approach, I was unable to observe any interaction with SMARCAL1 and the CST complex, a ssDNA binding protein complex often considered the telomeric RPA, or components of shelterin. However, all these studies were performed with WT SMARCAL1 protein that primarily identified RPA as an interaction partner (Betous et al., 2013b). Future studies to identify telomere binding proteins might be better achieved using the RPA binding mutant of SMARCAL1.

As an alternative to protein-mediated recruitment, SMARCAL1 could utilize its high affinity DNA binding to localize to telomeres since SMARCAL1 is capable of binding a variety of DNA structures with high affinity in the absence of RPA (Betous et al., 2012; Muthuswami et al., 2000; Yusufzai and Kadonaga, 2008). I have only been able to detect SMARCAL1 localization at telomeres when it is highly overexpressed (Figure 3.8). Thus, its localization to telomeres may only occur during the short window when telomere replication is happening.

Alternatively, it is possible that SMARCAL1 exerts its telomere maintenance function indirectly through some kind of signaling mechanism. I think this is unlikely given that its DNA-dependent ATPase activity is needed for its telomere maintenance function. Finally, I cannot rule out the possibility that the two-fold overexpression of the RPA binding mutant of SMARCAL1 is able to overcome the requirement of RPA to localize to sites of replication stress at telomeres. However, this level of expression is unable to rescue other phenotypes associated with SMARCAL1 silencing including increased intensity of the DNA damage marker γH2AX and defects in cell cycle progression (Bansbach et al., 2009; Ciccia et al., 2009).

Beyond localization, RPA also regulates SMARCAL1 enzymatic activity (Betous et al., 2013a). RPA stimulates SMARCAL1 fork reversal function on stalled replication fork substrates while inhibiting fork reversal on normal replication intermediates (Betous et al., 2013a; Bhat et al., 2015). Conversely, fork restoration by SMARCAL1 is inhibited by RPA on stalled replication forks but stimulated on substrates that would yield a normal replication fork structure. Since RPA appears to be dispensable for the telomere function of SMARCAL1, how this level of regulation is achieved at telomeres remains undetermined. Like localization, another protein could regulate SMARCAL1's function at telomeres; however, this factor has not yet been identified. Since identification of a telomere-specific factor has proven difficult, performing the *in vitro* fork remodeling assays with candidate proteins or protein complexes (such as components of the shelterin complex or the CST complex) might be a useful way to investigate this aspect of SMARCAL1 regulation.

Conclusions

My data provide the first indication of an endogenous replication stress that SMARCAL1 acts to resolve. It also separates the function of SMARCAL1 from related fork repair enzymes. SMARCAL1 deficiency causes the human disease SIOD. Symptoms associated with SIOD include renal failure, growth defects, immune deficiencies, and a slight predisposition to cancer (Baradaran-Heravi et al., 2012b; Boerkoel et al., 2002; Carroll et al., 2013). This disease lacks obvious telomere defect phenotypes including a strong pre-disposition to cancer or premature aging (Blasco, 2005). Onset of these telomere-associated phenotypes typically occurs often later in life; however, SIOD patients usually do not survive beyond two decades of life (Lou et al.,

2002). Thus, the telomere phenotypes may not manifest within the timeframe that SIOD patients live.

Interestingly, the SMARCAL1 mutant mouse does not display any disease phenotypes associated with SIOD in humans (Puccetti et al., 2017). One key difference between humans and mice at a cellular level is that murine telomeres are nearly five times longer than humans (Calado and Dumitriu, 2013). Further, telomerase is constantly active in mice indicating an active mechanism to maintain longer telomeres. It is tempting to speculate that the telomere function of SMARCAL1 might contribute to some disease phenotypes of SIOD. Studies with a telomerase knockout SMARCAL1-deficient mouse would be informative for elaborating on this topic. Future work characterizing the etiology of SIOD will be important to investigate the contribution of the telomere replication functions of SMARCAL1 to replication fidelity and genome stability.

CHAPTER IV

ADAPTING NEXT GENERATION SEQUENCING TO PROBE SMARCAL1 FUNCTION

DURING DNA REPLICATION AND REPAIR

Introduction

In addition to telomeres, the function of SMARCAL1 is required for replication through

other undetermined regions of the genome (Bansbach et al., 2010; Poole et al., 2015). There are

several other candidate repetitive regions such as satellite or ribosomal DNA repeats that might

require SMARCAL1 function. Alternatively, SMARCAL1 could be required to stabilize

replication forks that have encountered a specific fork stalling lesion on the DNA template or at

sites where then transcription machinery collides with the replisome. Several methods have been

developed to investigate each of these sources of replication stress individually; however, using a

candidate approach to examine all potential sources of replication stress is expensive and

inefficient. Adapting an assay to examine several regions in the genome within one experiment is

necessary to further characterize where SMARCAL1 is needed during DNA replication and begin

work investigating the requirement for other SNF2 fork remodeling enzymes.

Due to decreased costs and the development of several methodologies, next-generation

DNA sequencing (NGS) has become a promising method to analyze multiple regions in the

genome simultaneously. Several studies have utilized NGS-based methods to investigate diverse

aspects of DNA repair including identifying where DNA repair proteins function. In this chapter,

I will highlight some examples of NGS-based methods that have successfully been used to

characterize DNA repair proteins, present my work using these methods to characterize

SMARCAL1 during DNA replication, and discuss future directions to investigate the functions of SNF2 DNA translocases.

Characterizing the location of DNA repair proteins using ChIP-seq

Chromatin immunoprecipitation with DNA sequencing (ChIP-seq) has proven an invaluable tool to characterize proteins involved with DNA repair. Identifying where proteins bind on a global scale permits identifying patterns in localization that has been used to draw conclusions on many aspects of DNA repair including influences on pathway choice (Aymard et al., 2014) and locations of novel, difficult to replicate regions of the genome (Barlow et al., 2013).

Identification of sites of DNA damage has been done primarily by immunoprecipitation of the ssDNA binding protein RPA or the phosphorylated histone variant γH2AX, a histone modification that occurs in the early stages of DNA damage recognition. Much of this work has revolved around cell culture models with site-induced DNA breaks. Several characteristics of DNA damage sensing and signal propagation were determined using this system. For example, γH2AX propagates bi-directionally for kilobases in a manner dependent on the transcription state near the site of the DNA break (Iacovoni et al., 2010). Similar studies have also been performed using RPA as the marker of damage (Yamane et al., 2011; Yamane et al., 2013).

These methods have also been successfully applied in the absence of induced site-specific breaks. ChIP-seq using antibodies against RPA was used to identify previously uncharacterized sites of endogenous replication stress in mouse models (Barlow et al., 2013). In this study, novel, difficult to replicate sequences – primarily in repetitive and highly methylated regions – were identified in the early stages of DNA replication in mice (Barlow et al., 2013). This method was also used to determine that the CST complex, a RPA-like heterotrimeric complex that binds

ssDNA in a sequence independent manner, facilitates replication fork restart in response to fork stalling at G-rich repetitive regions of the genome (Chastain et al., 2016). These applications of ChIP-seq would closely mirror the experimental setup for investigating sources of endogenous replication stress sources that require the activity of SNF2 fork remodelers.

DNA break sequencing

ChIP-seq is an effective way to characterize where DNA repair proteins localize. However, using this method to identify where replication stress-induced DNA breaks occur is less effective as it relies on predictions based on indirect measurements. Additionally, proteins typically immunoprecipitated for these assays, specifically γH2AX, can be located kilobases away making it difficult to obtain high resolution information on the location of DNA breaks (Iacovoni et al., 2010). Measuring the locations of the breaks through more direct methods would be informative to answer questions associated with SMARCAL1 deficiency and address other aspects of DNA repair.

Within the last decade, several methods that directly sequence DNA breaks have been developed. These methods typically fall into two categories. Methods such as DNA end sequencing (END-seq) and BLESS (break labeling, enrichment on streptavidin, and next generation sequencing) require direct purification of the DNA break prior to sequencing (Figure 4.1). The other category of DNA break sequencing methods requires creating a site-specific DNA break that serves as the "bait" that can be fused to "prey" DNA breaks thus creating genomic rearrangements detected through DNA sequencing (Figure 4.1). This concept is used by methods such as high throughput, genome-wide translocation sequencing (HTGTS).
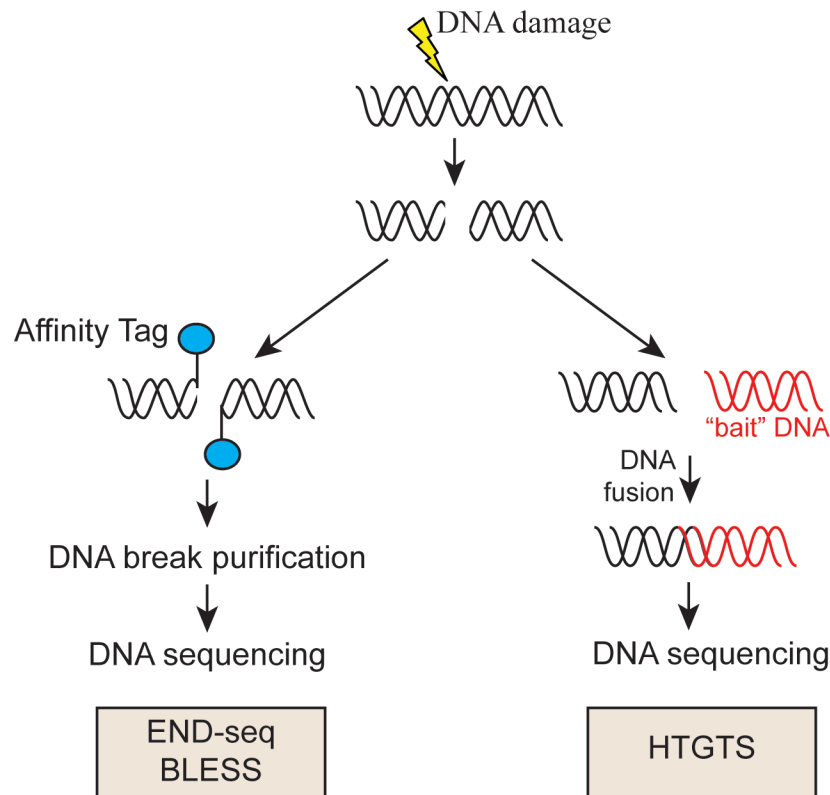
Figure 4.1. Overview of methods to analyze the DNA breakome. The location of DNA breaks can be determined using methods that directly purify the DNA break. Alternatively, the location of DNA breaks can be determined by generating a "bait" DNA break that will fuse to breaks that form after DNA damage resuling in a DNA fusion that is detected by sequencing. Methods that employ these approaches are indicated in boxes. END-seq, DNA end sequencing. BLESS, break labeling, enrichment on streptavidin, and next generation sequencing. HTGTS, high throughput, genome-wide translocation sequencing.

71

Both categories of breakome analyses, though very different, have been used to study several aspects of DNA repair. HTGTS was effectively used to show DNA breaks occur at a high frequency in highly transcribed regions and display a strong preference to be repaired through intrachromosomal fusions driven by non-homologous end joining (NHEJ) (Chiarle et al., 2011). END-seq, the newest of these methods, was used to map resection events that occurred at site-directed dsDNA breaks in NHEJ-deficient murine cells (Canela et al., 2016). Novel aphidicolin-sensitive regions in the genome were identified in HeLa cells using BLESS to selectively purify breaks (Crosetto et al., 2013). Applying these methods to characterize breaks generated in SMARCAL1-deficient cells would be informative in discerning where this protein is needed during DNA replication.

Genome sequencing

Lowering sequencing costs and increasing efforts to create personalized cancer treatments has resulted in large increases of whole genome sequence data sets. This increase in data has permitted identification of patterns associated with mutational processes that drive human disease (Alexandrov et al., 2013; Nik-Zainal et al., 2012). Using extensive bioinformatics and statistical analyses, DNA mutation patterns associated with defects in several DNA repair pathways have been identified (reviewed in (Helleday et al., 2014)). For example, mutations in *BRCA1* or *BRCA2,* crucial components of the homology directed repair (HDR) pathway that are frequently mutated in breast and ovarian cancers, result in indels surrounded by regions of microhomology (Alexandrov et al., 2013; Helleday et al., 2014). This mutation signature strongly implies DNA repair is shunted to alternative pathways when HDR is impaired. Signatures associated with

defects in other DNA repair pathways such as mismatch repair, base-excision repair, and nucleotide excision repair have also been reported (Helleday et al., 2014).

These large-scale sequencing data sets have also permitted the identification mutation patterns associated with failure to resolve endogenous and exogenous sources of DNA damage (Alexandrov et al., 2013). Failure to resolve pyrimidine dimers generated by UV damage results in an increased frequency of cytidine to thymidine (C>T) transitions (Alexandrov et al., 2013). This type of genome sequencing can be used to identify patterns down to single nucleotide changes that occur randomly throughout the genome. Though computationally expensive, whole genome sequencing has great advantages in that it does not depend on the efficacy of antibodies or *in situ* DNA labeling methods. Genome sequencing can complement and sometimes go beyond the capabilities of other NGS methods to answer important questions in DNA repair.

Results

I sought to investigate where SMARCAL1 functions during replication using a multi-faceted sequencing approach. When the replication fork stalls, RPA binds to ssDNA generated at these sites of stress (Figure 4.2). SMARCAL1 is then recruited to stabilize the stalled fork and promote restart after the source of replication stress is addressed. In conditions when SMARCAL1 is absent, RPA persists on the ssDNA and the stalled replication fork eventually collapses into a dsDNA break due to the action of structure-specific nucleases such as MUS81 (Betous et al., 2012; Dungrawala et al., 2017). I adopted a plan to use several NGS methods, including RPA ChIP-seq and breakome sequencing, to exploit these observations and determine where SMARCAL1 is needed during DNA replication (Figure 4.2).
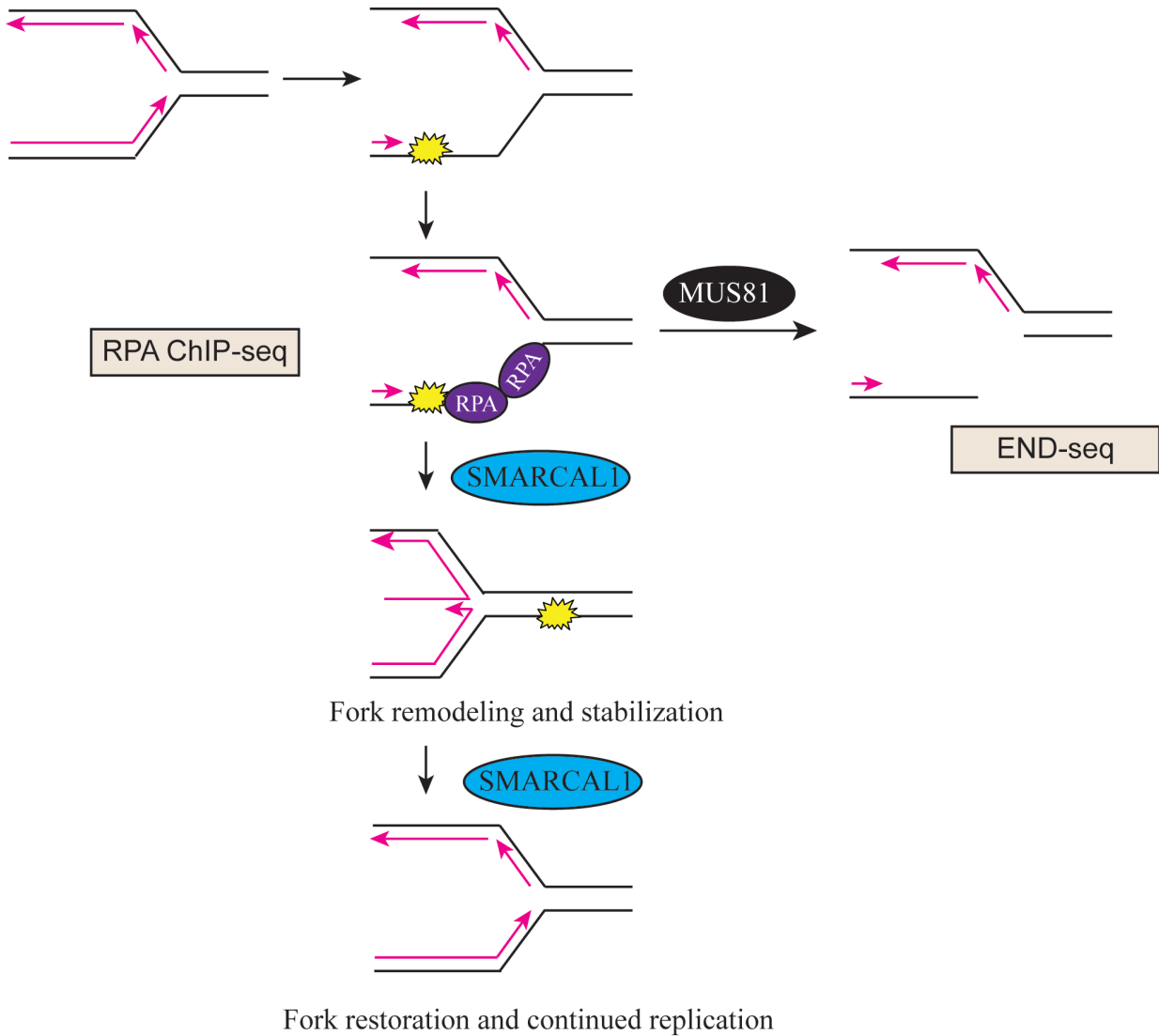
Figure 4.2. Model for sequencing approaches to study SMARCAL1. When replication forks stall, RPA coats the single-stranded DNA and recruits SMARCAL1 to perform fork stabilization through fork remodeling. In SMARCAL1-deficient cells, stalled replication forks are cleaved by nuclease such as MUS81 and result in DNA breaks. Methods employed in this study are shown in boxes. A lesion in the leading strand is indicated by the yellow symbol.

74

END-seq to characterize DNA breaks in SMARCAL1 knockdown cells

SMARCAL1 knockdown in human cells results in DNA breaks in the absence of any added genotoxic agent (Bansbach et al., 2009; Betous et al., 2012; Dungrawala et al., 2017; Poole et al., 2015). Identification of where these breaks occur would be informative to determine where SMARCAL1 is needed during DNA replication. I opted to use END-seq – which involves the direct labelling, and subsequent purification, of DNA ends from cells immobilized in agarose plugs (Canela et al., 2016). This method was determined to be significantly more sensitive than other DNA break purification methods with a lower limit of detection of 1 break in 10,000 cells (Canela et al., 2016).

I transiently knocked down SMARCAL1 in U2OS cells, an aneuploid osteosarcoma cell line with high levels of DNA breaks induced by SMARCAL1 deficiency (Dungrawala et al., 2017). To increase the intensity of the DNA break signal, I also treated cells with 2mM HU for 4 hours prior to embedding the cells in agarose plugs (Figure 4.3A-B). The END-seq library preparation and DNA sequencing was performed in collaboration with the Nussenzweig lab at the National Institute of Health.

Using END-seq, I obtained a similar number of DNA ends from U2OS cells treated with SMARCAL1 siRNA and a non-targeting control (Figure 4.3C). This was true with and without 2mM HU. When I filtered out DNA breaks found in the non-targeting control, the number of DNA ends specific to SMARCAL1 knockdown was reduced to 49. This number increased to 163 in the presence of HU. The amount of overlap between DNA ends from SMARCAL1 knockdown cells treated with HU or a vehicle control was minimal with only 6 peaks that overlapped between the two conditions. Thus, HU treatment did not amplify the frequency of DNA breaks at a specific
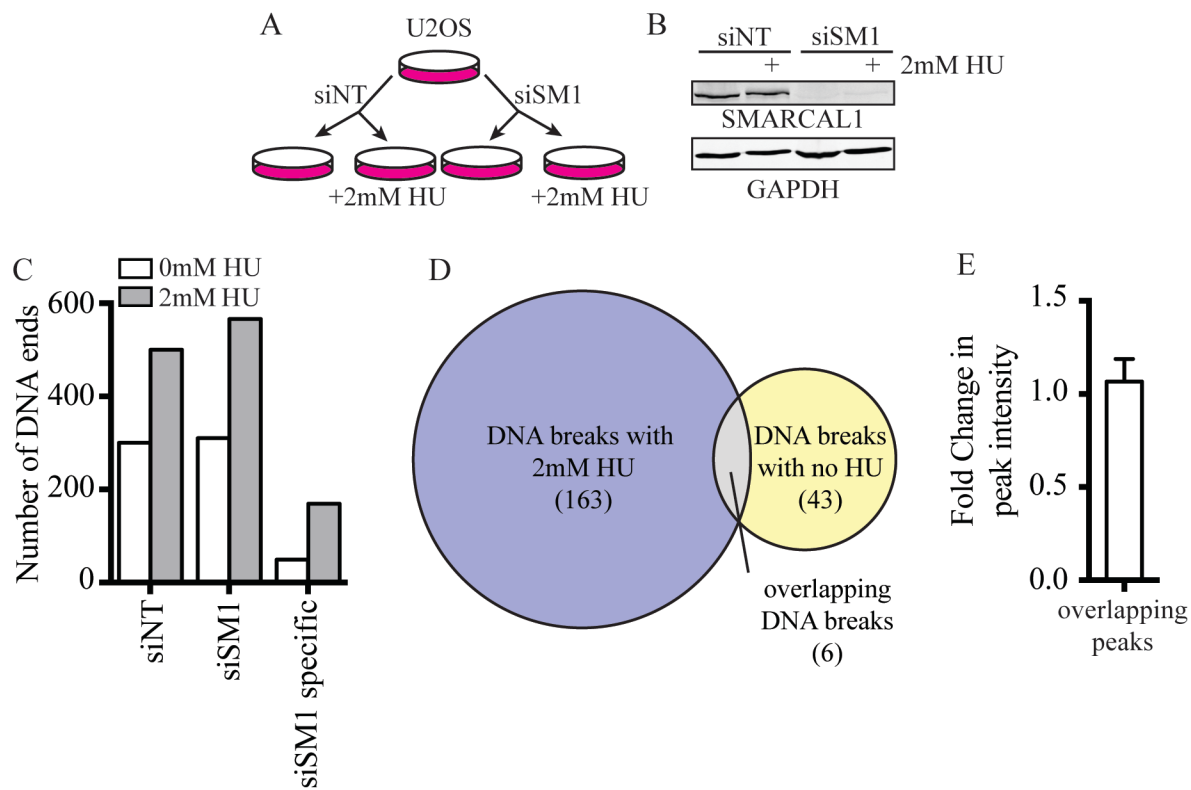
Figure 4.3. END-seq anlaysis of SMARCAL1-deficient cells. (A) Schematic of END-seq experimental sample setup. Cells treated with 2mM HU were treated for 4 hours prior to harvesting. (B) Western blot analysis of SMARCAL1 knockdown for END-seq samples. (C) Total number of DNA ends detected using END-seq and peak calling by HOMER. (D) Comparison of DNA breaks specific to SMARCAL1 knockdown with and without 2mM HU treatment. (E) Fold change of peak intensity for peaks found in both siNT and siSM1 conditions. siNT, siNON-TARGETING; siSM1, siSMARCAL1.

location in SMARCAL1 knockdown cells but instead generated new ones. To exclude the possibility that SMARCAL1 deficiency results in amplification of DNA breaks already present in wild-type cells, I examined the change in the signal intensity of DNA breaks present in control and SMARCAL1 knockdown cells. The intensity of the peaks associated with DNA breaks found in both conditions did not change indicating these breaks are not exacerbated by loss of SMARCAL1 (Figure 4.3E). Examples of DNA break peaks called in this experiment are shown in Figure 4.4.

Unfortunately, most of the peaks in both SMARCAL1-proficient and deficient untreated cells were composed of very few reads – as indicated by the read pileup – suggesting less confidence in these peaks (Figure 4.5A). The same trend was also observed for HU treated cells. Peaks containing less than 50 reads were filtered to exclude a large percentage of low-confidence peaks. Further, the caveats of END-seq include labeling telomere ends – which resemble DNA breaks (Canela et al., 2016). Overall, the number of peaks at telomeres was less than 10% of total peaks (Figure 4.5B), which corresponds to the reported prevalence of telomeres in END-seq data (Canela et al., 2016). However, the telomere peaks represented many of the higher confidence peaks with a higher read pileup (Figure 4.5C). Thus, the highest confidence peaks in my data set appear to predominately be in the telomere repeats.

I next investigated where the other DNA breaks were occurring in SMARCAL1-deficient cells. For example, it is plausible SMARCAL1 could be required for replication through other repetitive regions in the genome given its function in replication through telomeres. Failure to resolve the replication stress associated with repetitive sequences in the genome would result in an increased frequency of DNA breaks in these regions detected using END-seq. When I compared the percent of DNA breaks present in various repetitive elements to the percent of the genome comprised of these elements, I observed a striking increase in DNA breaks at satellite repeats
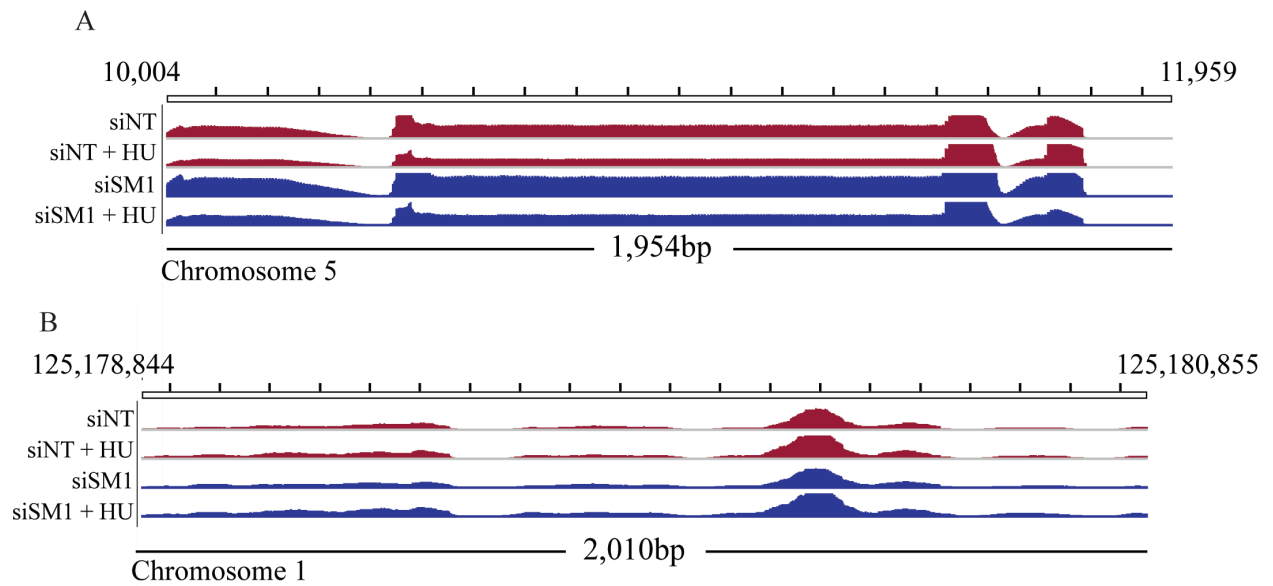
Figure 4.4. Example peaks from END-seq data. Representative peaks on chromosome 5 (A) and chromosome 1 (B) for all END-seq conditions. Locations of peaks are listed above the graphs. Images were created from screenshots from the Integrative Genomics Viewer (IGV). Peak heights were set at 7000 in (A) and 3000 in (B). siNT, siNON-TARGETING; siSM1, siSMARCAL1; HU, hydroxyurea.
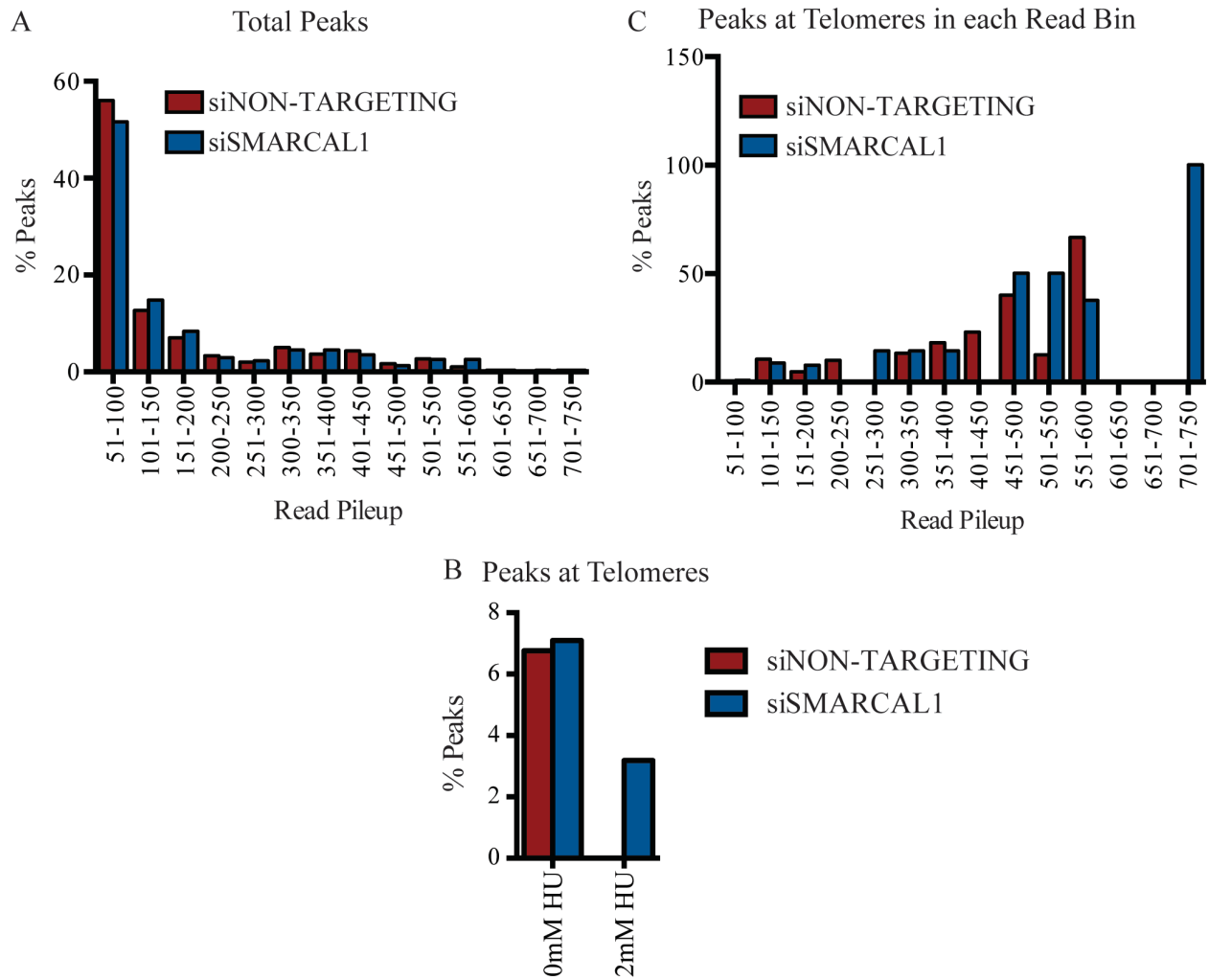
Figure 4.5. Overview of peak distribution in untreated samples. (A) Histogram showing the distribution of peaks composed of the indicated read pileup (number of reads that contributed to the peak calling). (B) Percent of total peaks present at telomeres. (C) Percent of peaks called with the specified read pileup that were found within telomeres.

(Figure 4.6) (Treangen and Salzberg, 2011). This enrichment is also true in the presence of exogenous replication stress. This was not a phenomenon associated with all repetitive elements since the long and short interspersed nuclear elements (LINEs and SINEs) were not enriched compared to the composition of the genome (Figure 4.6). However, this enrichment was observed in wild-type and SMARCAL1 knockdown conditions suggesting there may not be a bias for SMARCAL1-deficiency.

Given the high prevalence of breaks at microsatellites by the END-seq analysis, I sought to further investigate this finding by assaying for microsatellite stability using a secondary method based on polymerase chain reaction (PCR) (Matsuzaki et al., 2015). In this assay, microsatellite instability (MSI) will manifest as differences in the banding pattern compared to the wild-type control. I investigated MSI in two human cell models - SMARCAL1 knockout U2OS clones generated by CRISPR (Appendix A) and long-term siRNA knockdown every 3 days for a total of 7 transfections in HeLa long-telomere cells. In both cell lines, SMARCAL1 deficiency did not result in MSI for a variety of different repeats (Figure 4.7A). There was also no observable MSI associated with the SMARCAL1$^{\Delta/\Delta}$ MEFs when several different microsatellites were tested (Figure 4.7B).

I also investigated the prevalence of DNA breaks at other regions in the genome that could be difficult to replicate. I examined the percent of DNA ends near transcription start sites as an initial investigation into the contribution of SMARCAL1 in resolving conflicts between the replication and transcription machinery. Most of the DNA ends were located more than 200 kilobases from a transcription start site (Figure 4.8A). This enrichment decreased when cells were treated with 2mM HU, again suggesting HU did not exacerbate DNA breaks associated with SMARCAL1 but, rather, generated several new ones. A similar trend was observed for peaks that
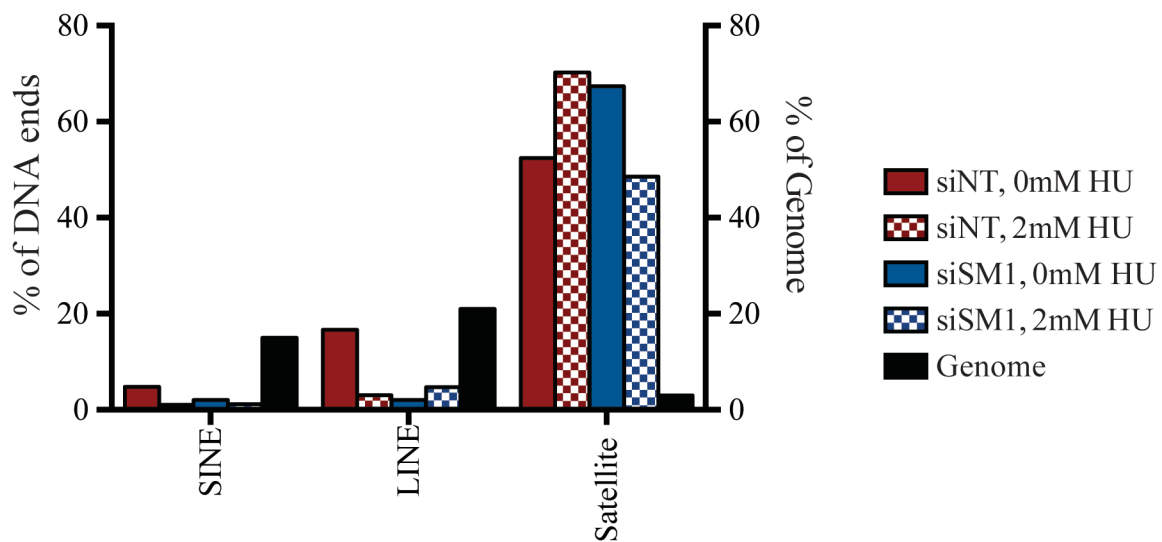
Figure 4.6. Genomic locations of DNA breaks in SMARCAL1-deficient cells. Percent of peaks present in repetitive elements (red and blue solid and checkers) and percent of the genome composed of each repetitive element (black) Genome composition was obtained from by (Treangen and Salzberg, 2011).
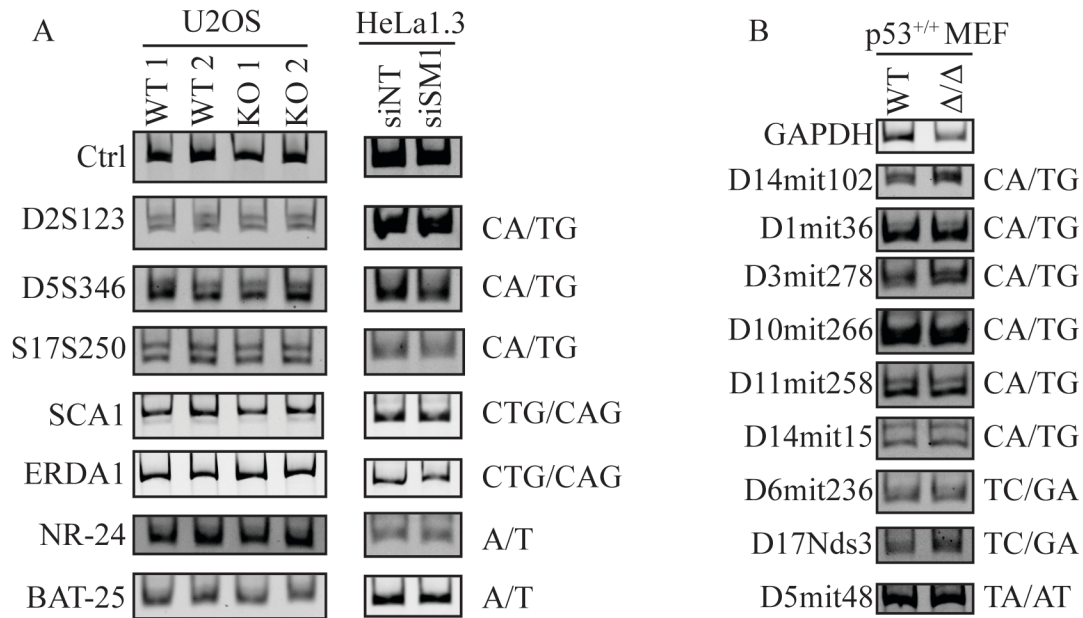
Figure 4.7. SMARCAL1-deficient cells do not display microsatellite instability. SMARCAL1 knockout U2OS cells (A, left), SMARCAL1 transient knockdown HeLa1.3 cells (A, right), and SMARCAL1 mutant MEFs do not display changes in microsatellite repeat number using PCR. HeLa LT cells were transfected with siRNA against SMARCAL1 every 3 days for a total of 7 transfections. Names of specific microsatellite regions are identified to the left of each sample; repeat type is labeled to the right of each sample. DNA was detected with ethidium bromide staining.
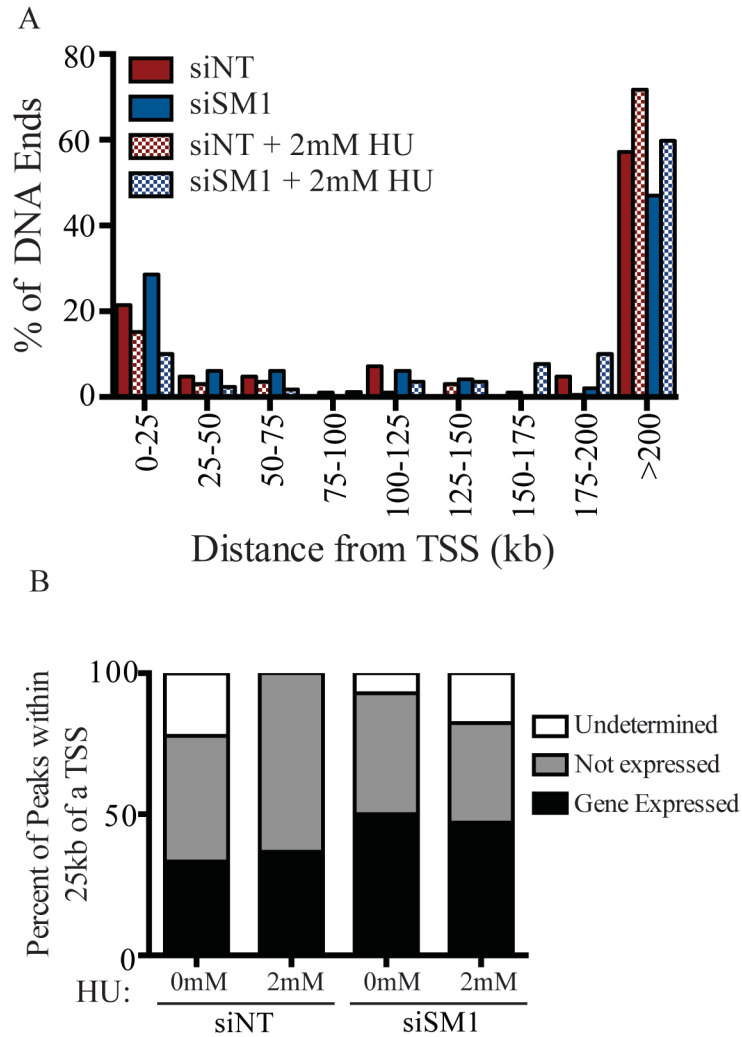
Figure 4.8. Prevalence of DNA breaks at TSS in SMARCAL1-deficient cells. (A) Histogram of the distance between DNA ends and the closest transcription start site (TSS). (B) Gene expression profile of peaks within 25kb of a TSS. siNT, siNON-TARGETING; siSM1, siSMARCAL1; HU, hydroxyurea.

Interesting to note, nearly a quarter of the DNA ends observed in the END-seq analysis were found within 25 kilobases of a transcription start site (Figure 4.8A). When I examined the expression profiles of the genes located within 25 kilobases of a DNA break, many of the genes were not expressed in U2OS cells based on RNA sequencing data from our laboratory (Figure 4.8B).

RPA ChIP-seq on a positive control cell line requires further optimization

In my efforts to identify where the functions of SMARCAL1 are required during DNA replication, I also decided to utilize ChIP-seq. In this approach, I opted to use antibodies to immunoprecipitate RPA since SMARCAL1 deficiency should result in prolonged occupancy of RPA on DNA and, thus, would be a good marker of where SMARCAL1 is needed during replication (Figure 4.2). Further, the signal associated with RPA ChIP yields narrower peaks compared to other markers of replication stress, like γH2AX (Barlow et al., 2013). Antibodies against RPA have been effectively used for RPA ChIP-seq in multiple studies to identify sites of replication stress; therefore, verified reagents were already available for use (Barlow et al., 2013; Yamane et al., 2011; Yamane et al., 2013). Finally, immunoprecipitating RPA rather than other markers of DNA damage, such as γH2AX, would permit identification of sites of replication stress that do not necessarily collapse into dsDNA breaks. Although SMARCAL1 knockdown results in a significant induction of DNA breaks, knockdown of ZRANB3 or HLTF does not result in DNA breaks in the absence of exogenous damage (Achar et al., 2015; Betous et al., 2012; Dungrawala et al., 2017). Thus, adapting a method that would be applicable to all members of this family would also be beneficial in characterizing differences between these enzymes.

I sought to confirm my methodology using a positive control cell line prior to proceeding with my experimental samples. I used a U2OS cell line with an HA-tagged AsiSI enzyme fused to

the estrogen receptor (Iacovoni et al., 2010). This enzyme typically resides in the cytoplasm where it is unable to cleave DNA (Figure 4.9). Upon addition of the small molecule 4-hydroxytamoxofin (4OHT), the fusion enzyme relocates to the nucleus and cuts the genome in a sequence-specific manner. Although there are over 1000 potential AsiSI cut sites present in genome, only approximately 200 are cleaved based on the chromatin landscape associated with the cut sites (Massip et al., 2010). I confirmed induction of the enzyme after treatment with 4OHT by immunofluorescence. After 4 hours of treatment with 4OHT, the AsiSI enzyme completely re-localized to the nucleus. Further, the enzyme cleaved the DNA as indicated by the induction of γH2AX foci. This induction was absent in the vehicle control cells treated with ethanol (EtOH) (Figure 4.9).

I effectively optimize RPA immunoprecipitation from the induced U2OS cells; however, significant amounts of RPA still remained in the fractions that were not bound by the antibody (Figure 4.10). When I compared the location of the immunoprecipitated RPA to sites of AsiSI cut sites in the genome using quantitative PCR, I saw an association of RPA with known AsiSI cut sites in cells where the enzyme had been induced with 4OHT (Figure 4.10).

When I extended my analysis to include DNA sequencing rather than qPCR, I saw a significantly higher number of ChIP-seq peaks compared to cells treated with the EtOH control (Figure 4.11A). However, many of these peaks were composed of very few reads indicating lower confidence and higher probability of a false positive signal. When I filtered the list with a cutoff to exclude these low confidence signals, the number of peaks significantly decreased such that there was a difference of only 200 between the treatment and control samples (Figure 4.11B). Further filtering to exclude peaks present in control set reduced the number of 4OHT treatment specific peaks to 324. However, only around 20% of these peaks were within 20kb of any AsiSI
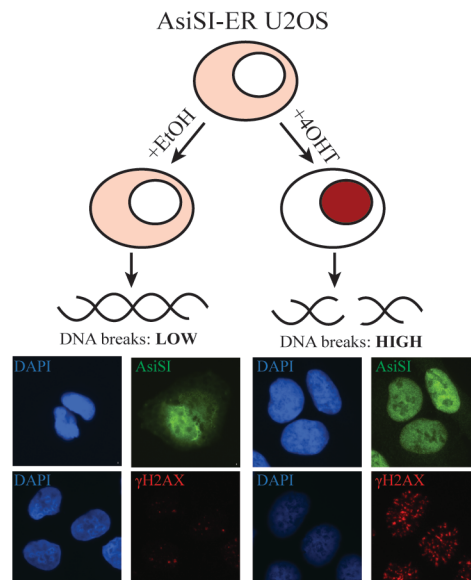
Figure 4.9. Schematic of positive control cell line for RPA ChIP-seq optimization. Top, overview of U2OS cell system stably expressing HA-AsiSI. Without induction, the restriction enzyme AsiSI is retained in the cytoplasm and cannot cleave the DNA. Upon induction with 4-hydroxytamoxofin, the enzyme relocates to the nucleus and cuts the DNA resulting γH2AX induction. Bottom, representative immunofluorescence images of HA-AsiSI (green) and γH2AX (red) for each treatment condition.
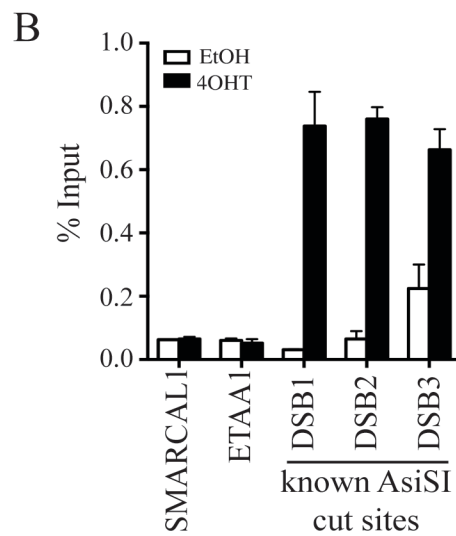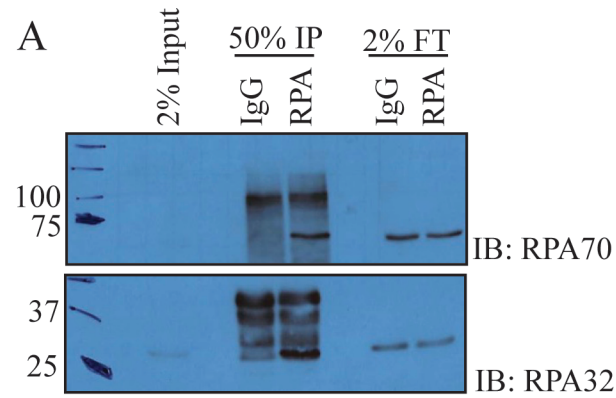
Figure 4.10. RPA ChIP optimization. (A) Immunoblot following RPA immunoprecipitation from unin-duced AsiSI U2OS cells. Numbers on left indicate the DNA ladder. (B) Quantitative PCR analysis from RPA ChIP of AsiSI U2OS with and without induction. Primer sets SMARCAL1 and ETAA1 are controls where no signal should occur. DSB1-3 indicate primer sets at known AsiSI cut sites in the genome. EtOH, ethanol; 4OHT, 4-hydroxytamoxofin.
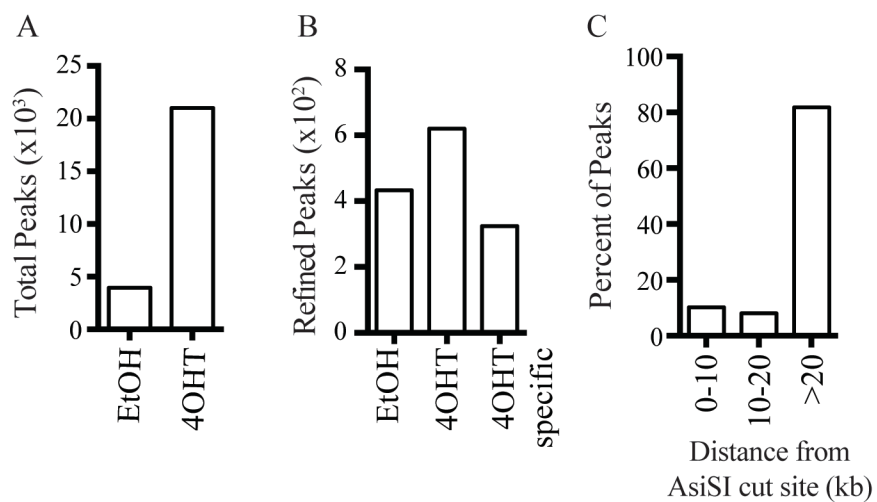
Figure 4.11. RPA ChIP-seq results. (A) Bar graph indicating the total number of RPA ChIP-seq peaks called using MACS2. (B) Bar graph indicating the refined number of peaks. (C) Histogram of the number of peaks near known AsiSI cut sites. EtOH, ethanol; 4OHT, 4-hydroxytamoxofin; 4OHT specific, peaks found in the 4OHT treated sample only.

cut site (Figure 4.11C). While there was an increase in peaks associated with 4OHT treatment, these were not located near an induced DNA break. This result was consistent between two replicates where one library was prepared by the sequencing core at Hudson Alpha and the other was prepared by me at Vanderbilt University.

Discussion

Next generation sequencing is an efficient and relatively inexpensive method that is widely used to characterize proteins involved in multiple aspects of DNA replication and repair. Using various NGS methods, I sought to determine the replication stress contexts requiring the function of SMARCAL1. These studies could then be applied to its closely related family members HLTF and ZRANB3 to elucidate the requirement for 3 biochemically similar enzymes.

RPA ChIP-seq needs further optimization

Using ChIP-seq, I was unable to detect enrichment of RPA at sites of induced AsiSI breaks in my positive control samples (Figure 4.11). This result is concerning as RPA ChIP-seq has been successfully used with AsiSI-induced breaks (Canela et al., 2016; Lopez-Saavedra et al., 2016). Further optimization of the immunoprecipitation is likely needed for the success of this protocol. I immunoprecipitated RPA from the U2OS cells; however, much of the RPA remained in the fraction not bound by the antibody (Figure 4.10). This work was done using a single RPA antibody, so testing alternative antibodies might improve the protein yield. Alternatively, the signal of RPA bound to sites of AsiSI cleavage has been enhanced by co-deletion of 53BP1 and LIGIV, two components of the NHEJ repair pathway (Canela et al., 2016). Preventing NHEJ-mediated repair would enhance RPA occupancy at the sites of DNA breaks and would likely enhance the signal in

this positive control experiment. However, it is not known if including these extra gene knockouts would affect RPA binding in SMARCAL1-deficient settings, further complicating the analyses of my future RPA ChIP-seq data sets.

As an alternative to RPA, ChIP-seq with the DDR marker γH2AX might bypass some obstacles associated with immunoprecipitating RPA. In addition to coating ssDNA at stalled replication forks, RPA also functions in normal DNA replication during elongation and Okazaki fragment maturation on the lagging strand (Oakley and Patrick, 2010). Therefore, RPA ChIP-seq likely includes DNA bound to these regions and could result in higher background. Conversely, chromatin bound γH2AX levels are low in conditions without replication stress and, thus, would provide a lower background. Additionally, I was able to detect a significant increase of γH2AX bound to DNA after induction with 4OHT suggesting this would be a feasible alternative (Figure 4.9). However, the best alternative might be to perform SMARCAL1 ChIP-seq to directly assess where this protein localizes. SMARCAL1-specific antibodies have not been tested for ChIP applications; however, this may be a useful tool to investigate where SMARCAL1 functions during replication. Alternatively, using CRISPR to insert an affinity tag at the SMARCAL1 locus could also be useful for purifying SMARCAL1 with commercially available and experimentally validated antibodies. Further, this method would permit purifying endogenous levels of protein that would prevent introducing DNA damage associated with transient overexpression of SMARCAL1.

END-seq

An initial experiment using END-seq indicated a large percentage of DNA breaks were located in repetitive regions. This result is curious as repetitive elements are typically considered

low confidence as they difficult to align to the genome (Treangen and Salzberg, 2011). END-seq suggested an increased frequency of DNA breaks at microsatellites, though this was not specific to SMARCAL1-deficiency (Figure 4.6). I was unable to see any MSI at a variety of microsatellites tested using PCR in both human and murine cells (Figure 4.7). Although it seems unlikely that SMARCAL1 functions specifically at microsatellites, I only tested a subset of microsatellites present in both species by PCR so it is possible other untested regions could be effected. Alternatively, the DNA breaks caused by SMARCAL1 knockdown may not result in microsatellite expansions or contractions that are detectable by this PCR-based method. The prevalence of microsatellites in the END-seq data suggests U2OS cells may be constantly experiencing stress indicating this cell line may not be the most appropriate for these analyses. Though the data suggests SMARCAL1 may not be required for microsatellite stability, further work is needed to completely rule out a function of SMARCAL1 at microsatellites.

In addition to the microsatellites, I saw a slight enrichment of DNA breaks that were present within 25 kilobases of a transcription start site, suggesting SMARCAL1 could be required to resolve conflicts between the replication and transcription machinery (Figure 4.8A). SMARCAL1 is capable of remodeling RNA-DNA structures lending support to this idea (Kassavetis and Kadonaga, 2014). However, only half (at most) of the genes located within 25 kilobases of a DNA break were actually expressed in U2OS cells (Figure 4.8B). Alternatively, SMARCAL1 could be functioning to resolve replication-transcription conflicts further within the gene body. Analyses to look at the prevalence of DNA breaks at specific genes that are highly expressed might be informative to further probe a function of SMARCAL1 in resolving transcription conflicts.

Interestingly, the top hits from END-seq (as measured by read pileup) in SMARCAL1 WT and knockdown cells were located at telomeres (Figure 4.5C). SMARCAL1 depletion results in

DNA breaks at telomeres, so this was not an unexpected result (Cox et al., 2016; Poole et al., 2015). However, U2OS cells utilize the ALT pathway and, as a result, have telomeres that are constantly breaking and elongating. Thus, the enrichment observed in my END-seq samples could be due to the telomere maintenance pathway rather than DNA breaks associated with loss of SMARCAL1. Changing the cell type in future studies would address this issue and prevent flooding the system with telomere ends over DNA breaks that occur in other regions of the genome.

Conclusions

Although this work is in its initial stages and likely requires more optimization, it's unclear if these methods will be appropriate for determining where SMARCAL1 is needed for replication. Most studies utilizing these methods have involved induction of sequence specific DNA breaks. This method results in many DNA sequencing reads stacked in the same location flanking the break, resulting in a strong signal and easy identification of breaks (Figure 4.12). However, breaks in SMARCAL1-deficient cells would not likely occur at a specific sequence since SMARCAL1 binds DNA in a sequence independent manner. Instead, breaks generated in SMARCAL1-deficient cells would likely be localized more generally to a specific DNA region. If this is the case, DNA reads gathered from ChIP-seq or breakome analyses would be spread across a larger region resulting in a less intense signal and make it more difficult to identify locations of DNA breaks in SMARCAL1-deficient settings above the background noise (Figure 4.12).

Instead, the best approach might be to perform some form of genome sequencing. These methods are more likely to yield promising results for sequence-independent DNA damage, a situation that more closely resembles the hypothesis for regions requiring SMARCAL1 function.
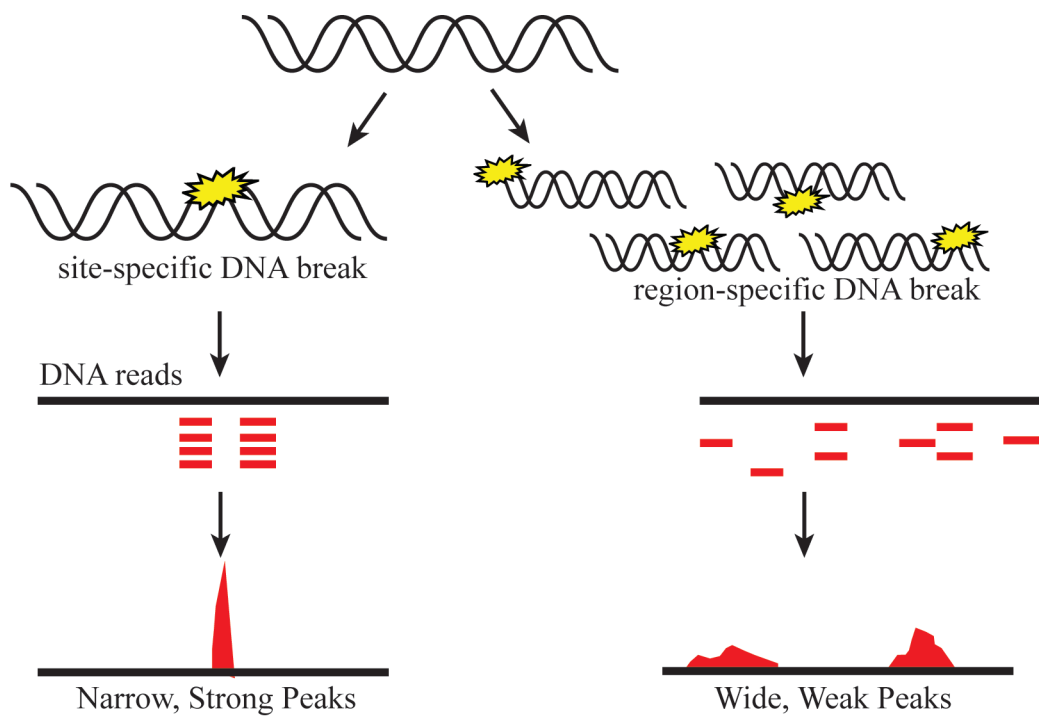
Figure 4.12. Models for DNA sequencing data analyses. Examples of site-specific (A) and region-specific (B) DNA break patterns and resulting DNA sequencing peaks.

However, completion of these studies will require significant computational power and collaborations with a bioinformatician to discern the significance of mutations identified in SMARCAL1-deficient cells. Due to the financial and computational costs associated with whole genome sequencing, it is important to carefully devise the proper model system for these analyses. Sequencing the genome of SIOD patients would be an ideal situation and would have obvious application to understanding the etiology of the disease. However, due to the rarity of the disease and considering the emotional distress experienced by patients and their families, this does not seem a feasible option. Alternatively, sequencing tumors that develop in SMARCAL1 mutant mice after ionizing radiation would be a strong alternative to the human disease model as it still involves an *in vivo* approach (Puccetti et al., 2017). A strong cell culture alternative would be generating SMARCAL1 knockout cell lines using CRISPR. This approach would permit generation of several biological replicates for sequencing and could easily be extended to include knockouts of ZRANB3 and HLTF while preserving a similar biological background.

CHAPTER V[†]

DISCUSSION AND FUTURE DIRECTIONS

Summary of dissertation work

DNA replication is constantly challenged by several sources of replication stress that threaten the integrity of the genome and promote human disease (Zeman and Cimprich, 2014). The DDR functions to address these sources of replication stress through multiple pathways including proteins like the SNF2 family members SMARCAL1, ZRANB3, and HLTF that facilitate stabilization of the replication fork through fork remodeling. Despite our knowledge on the enzymatic functions of these enzymes, several questions concerning where these proteins function in cells still remain.

I identified telomeres as the first endogenous source of replication stress that requires the function of SMARCAL1 (Chapter III and (Poole et al., 2015)). Importantly, this function is not shared by ZRANB3 or HLTF (Figure 3.4) suggesting these enzymes may be required in specific replication stress contexts or, perhaps, at specific regions of the genome. Using next generation sequencing approaches, I tried to identify other regions of the genome that require the function of SMARCAL1 with the hope of extending these studies to examine the functions of ZRANB3 and HLTF in the future (Chapter IV). Though unsuccessful, I was able to generate several tools for future studies in the lab including pipelines for DNA and RNA sequencing analyses and several SMARCAL1 knockout cells lines (Chapter IV and Appendix A).

[†] Parts of this chapter were adapted from Poole, L.A., and Cortez, D. (2017). Functions of SMARCAL1, ZRANB3, and HLTF in maintaining genome stability. Crit Rev Biochem Mol Biol, 1-19.

Though my work was driven by the hypothesis that SMARCAL1, ZRANB3, and HLTF are required for replication through specific regions of the genome, there are several alternative hypotheses to explain the requirement for each of these enzymes. In the following sections, I will discuss some of these alternatives and present future directions to tests these models.

Model 1: SMARCAL1, ZRANB3, and HLTF are needed in response to specific DNA structures

The requirement for each of these fork remodeling enzymes could be to address a specific DNA structure that forms when a replication fork has stalled. SMARCAL1, ZRANB3, and HLTF can catalyze fork reversal in the absence of any additional proteins. The efficiency of these reactions begins to change once DNA gaps are introduced to the substrate and supplemental DNA binding proteins are added to the reactions (Figure 5.1, Table 5.1). Fork reversal by SMARCAL1 on a leading strand gapped DNA substrate is stimulated by the presence of RPA whereas fork reversal by ZRANB3 is inhibited and HLTF function may be unaffected (Figure 5.1) (Achar et al., 2011; Betous et al., 2013a). The SRD HIRAN in HLTF specifically recognizes a 3'-OH while SMARCAL1 does not display a bias for this functional group (Kile et al., 2015). These differences in substrate preferences, conferred by the SRDs, would diversify the pool of potential substrates and could contribute to the requirement for each of these enzymes in cells.

In vitro biochemical approaches would be the most appropriate method to further investigate this model. Although I have detailed several differences that differentiate these proteins, these studies are not complete. Exhaustive studies to examine the DNA binding preferences of SMARCAL1 have been performed; however, HLTF and ZRANB3 preferences for substrates such as ssDNA gaps or different ssDNA overhang structures have not been determined. Differences in fork remodeling efficiencies with supplemental proteins, such as RPA or PCNA, have not been
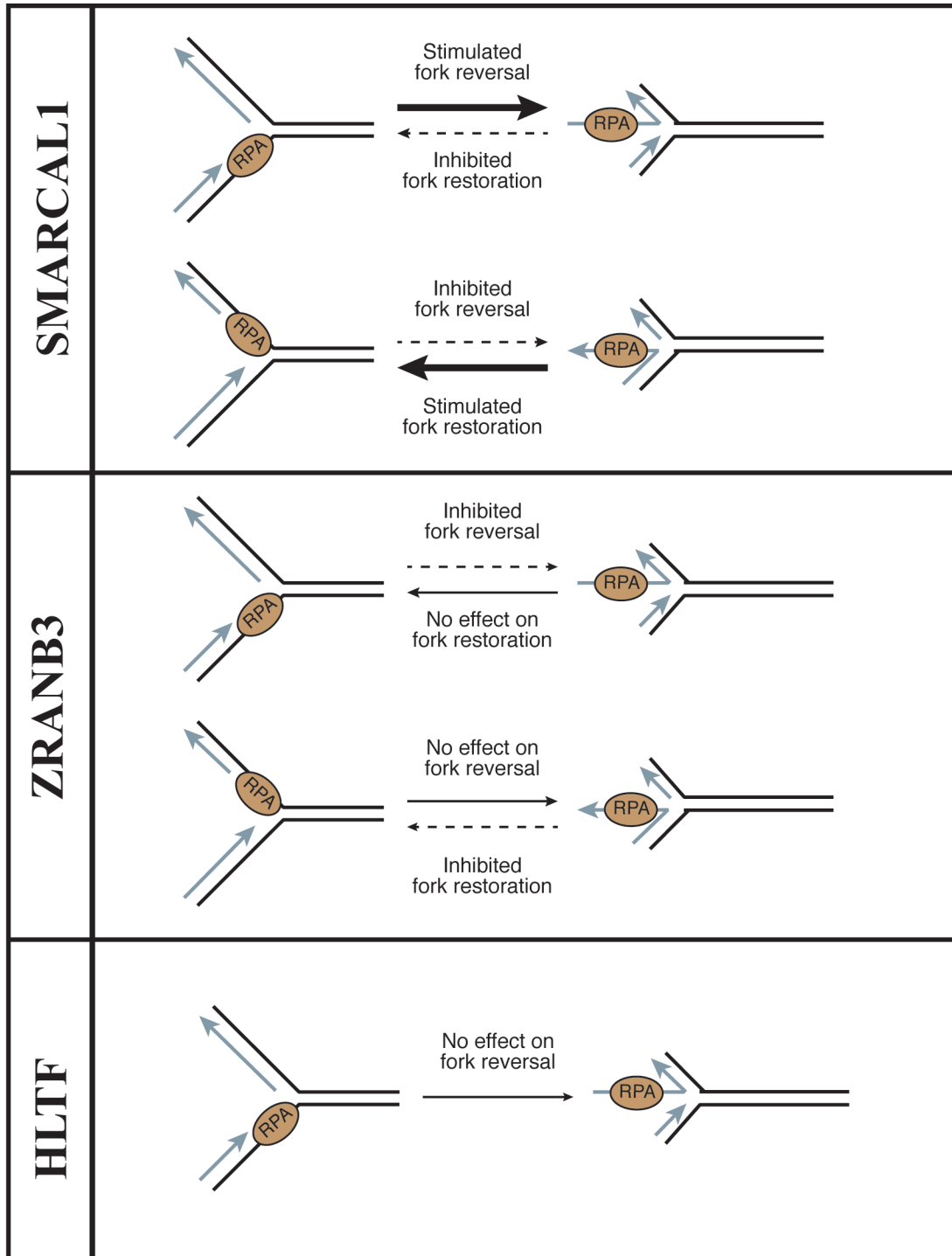
Figure 5.1. Effect of RPA on fork remodeling by SNF2 enzymes. Schematic of the effects of RPA on fork reversal and fork restoration for SMARCAL1, ZRANB3, and HLTF. Arrows indicate the direction of DNA polymerization.
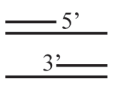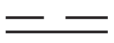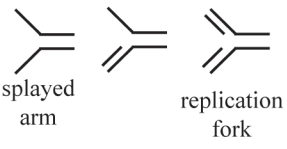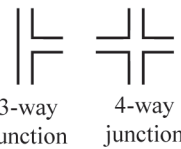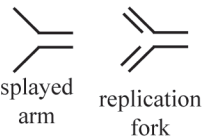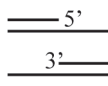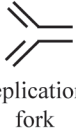
Table 5.1. Table of DNA structures bound by each of the SNF2 enzymes. Summary of prefered DNA binding structures tested by SMARCAL1, ZRANB3, and HLTF.

determined with all these enzymes. Extending these studies to include the other DNA binding proteins implicated in fork reversal, such as RAD51 or RADX, could identify major differences between the SNF2 fork remodelers and lend support to this model (Dungrawala et al., 2017; Zellweger et al., 2015).

Model 2: Multi-faceted regulation of recruitment and function

The redundancies in fork remodeling functions of SMARCAL1, ZRANB3, and HLTF suggest they could be performing the same reactions in cells. Another model to explain the requirement for all of these biochemically similar enzymes could be different mechanisms of regulation – both in recruitment to sites of damage and regulation of enzymatic function.

SMARCAL1 contains an RPA binding domain at its N-terminus that is required for its localization to sites of stress (Figure 1.4) (Bansbach et al., 2009; Ciccia et al., 2009; Postow et al., 2009; Yuan et al., 2009; Yusufzai et al., 2009). ZRANB3 contains multiple motifs and domains that permit binding to polyubiquitinated PCNA in response to replication stress (Figure 1.4) (Ciccia et al., 2012; Weston et al., 2012; Yuan et al., 2012). These differences in protein interactions could also translate into spatial regulation of these enzymes. This could, perhaps, also be extended to include some degree of temporal regulation depending on the speed at which PCNA modifications occur. How HLTF is recruited to sites of replication stress has not been determined. One study looking to identify HLTF interacting partners in unperturbed cells identified RPA70 and the dsDNA break repair regulatory protein PTIP (Pax transactivation domain-interacting protein), though the functional importance of these interactions has not been determined. At least a subset of HLTF travels with the replication fork in unperturbed conditions suggesting it may not require a supplemental protein interaction to localize to sites of replication stress (Kile et al., 2015).

However, research to understand recruitment of HLTF to sites of replication stress is largely unexplored.

Conversely, all of these enzymes could be recruited whenever the replication fork encounters stress. This model is supported by the observation that, like HLTF, at least a subset of SMARCAL1 also travels with the replication fork in unperturbed conditions (Betous et al., 2012; Dungrawala et al., 2015). With this model, the function of these proteins would need to be regulated to prevent these enzymes from potentially competing with each other and delaying rapid fork remodeling in response to replication stress. Phosphorylation was already determined to be a critical factor regulating SMARCAL1 function at replication forks (Carroll et al., 2014; Couch et al., 2013). Phosphorylation of SMARCAL1 in the linker between the bi-lobed ATPase domains limits its function after SMARCAL1 has localized to sites of replication stress (Couch et al., 2013). Initial microscopy studies indicate ZRANB3 retention at sites of replication stress is also dependent on phosphorylation by DDR kinases (Ciccia et al., 2012). With this model, these proteins could all be recruited to the same site of replication stress but require either activation of their enzymatic function or modifications that limit their activity to facilitate repair by the appropriate enzyme. To date, no studies on the regulation of HLTF by post-translational modifications (PTMs) has been reported. Mapping and characterizing the sites of PTMs in each of these enzymes would facilitate investigating this model.

Model 3: Differences in enzymatic activities drive the requirement for each enzyme

In addition to the core ATPase domains, HLTF and ZRANB3 possess additional enzymatic domains that confer additional functions (Table 1.1). HLTF contains an RING E3 ubiquitin ligase domain that facilitated poly-ubiquitination of PCNA in response to replication stress (Motegi et

al., 2008; Unk et al., 2008). The function of this domain has strong implications on DNA replication as mutations in the RING domain result in increased frequencies of fork stalling and fork collapse after treatment with DNA damaging agents (Blastyak et al., 2010). However, it is unclear if the functions of the RING and ATPase domains are somehow linked. More specifically, it is unclear if HLTF is recruited to a stalled fork to polyubiquitinate PCNA then continues to remodel the same stalled replication fork. The ability to perform both activities would be beneficial and forego the necessity to recruit additional proteins that perform the same reaction. Though a provocative thought, this remains untested.

Additionally, ZRANB3 contains an endonuclease domain that permits cleaving the DNA two nucleotides upstream of the replication fork junction (Badu-Nkansah et al., 2016b; Sebesta et al., 2017; Weston et al., 2012). It's tempting to speculate that ZRANB3 would preferentially be recruited to replication forks that required DNA cleavage during fork remodeling to stabilize the stalled fork. This nuclease activity is coupled with ATPase activity indicating ZRANB3 could remodel and cleave a single fork, lending support to this model.

Model 4: Cooperating functions of fork remodeling enzymes

My work has focused on the hypothesis that SMARCAL1, ZRANB3, and HLTF function in specific and distinct replication stress contexts. Alternatively, these enzymes could be operating sequentially or cooperatively to facilitate stabilization or repair of the same stalled replication fork in cells. PCNA ubiquitylation by HLTF could be part of the signal that directs ZRANB3 to act on a stalled fork (Figure 5.2). Also, the DNA substrate changes during the process of fork reversal. A
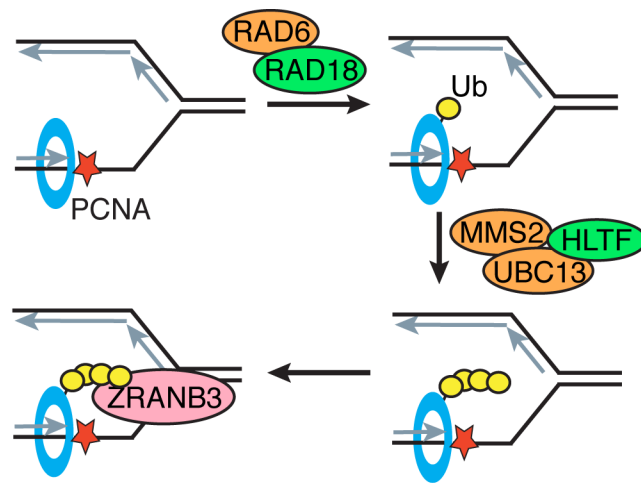
Figure 5.2. Model for cooperativity between HLTF and ZRANB3. PCNA is monoubiquitinated by the RAD6/RAD18 complex after the replication fork stalls at a lesion on the leading strand. Polyubiquitin chains are added by MMS2/UBC13 with the ubiquitin ligase HLTF. ZRANB3 then binds polyubiquitinated PCNA (using its NZF motif, PIP box, and APIM) and remodels or cleaves the stalled replication fork. Ub, ubiquitin.

leading strand gapped fork with RPA bound could be a good substrate for SMARCAL1, but as soon as the fork has reversed far enough to remove RPA and bring the 3' end of the leading strand close to the fork junction, it would be an excellent substrate for HLTF recognition (Figure 5.3). Regulatory signals could mediate a handoff mechanism to coordinate the fork remodeling functions of these enzymes. Further, the function of nucleases both before and after fork reversal could change the structure of the replication fork making it a better or worse substrate for one of the enzymes.

In this model, the differences in phenotypic outcomes caused by inactivation of each enzyme could be due to alternative functions outside fork remodeling for each of these enzymes (Table 1.1). SMARCAL1 was reported to anneal unwound DNA at breaks to facilitate repair of dsDNA breaks through non-homologous end-joining (NHEJ) (Keka et al., 2015). SMARCAL1 also has been reported to have a regulatory role in the transcription of a subset of genes (Baradaran-Heravi et al., 2012a; Haokip et al., 2016; Patne et al., 2017; Sharma et al., 2015). ZRANB3 and HLTF have reported functions in regulating cellular levels of recombination intermediates such as the displacement loop (D-loop) utilized in the HR repair pathway (Burkovics et al., 2014; Ciccia et al., 2012). Alternatively, failures at any one step in a fork repair process would cause shunting of the stalled fork into alternative pathways that may differ depending on which enzyme was inactivated, also leading to different phenotypes.

Other proteins are important for fork reversal

Beyond the SNF2 family, there are several other enzymes capable of performing fork remodeling *in vitro* (Table 5.2). The Fanconi anemia pathway helicase FANCM and the helicase FBH1 can also catalyze fork reversal while RAD54 can catalyze fork reversal and fork restoration
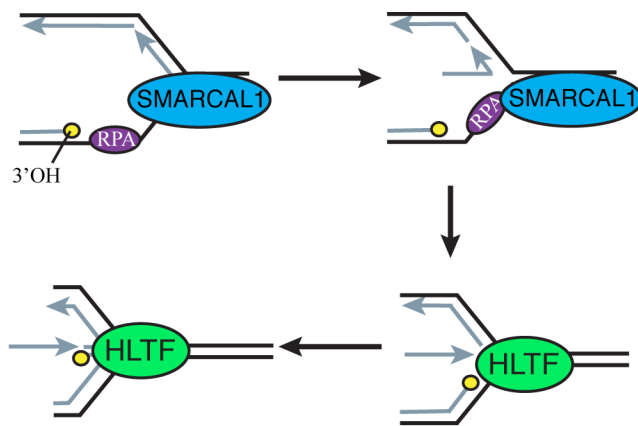
Figure 5.3. Model for substrate handoff between SMARCAL1 and HLTF. SMARCAL1 is recruited to sites of replication stress by RPA and reverses the replication fork until RPA is dissociated and the 3'-hydroxyl (3'-OH) group is close to the fork junction. HLTF can then bind the 3'OH group and continue reversing the replication fork.

Protein Name	Fork Reversal	Fork Restoration	References
SMARCAL1	Yes	Yes	Betous et al., 2012; Betous et al., 2013; Bhat et al.,2015; Kolinjivadi et al., 2017
ZRANB3	Yes	Yes	Betous et al., 2012; Betous et al., 2013; Ciccia et al., 2012
HLTF	Yes	ND	Blastyak et al., 2010; Achar et al., 2011; Achar et al., 2015; Kile et al., 2015
RECQ1	No	Yes	Popuri et al., 2008; Berti et al., 2013
RECQ5	Yes	ND	Kanagaraj et al., 2006
BLM	Yes	Yes	Machwe et al., 2006; Ralf et al., 2006; Machwe et al., 2011a; Machwe et al., 2011b
WRN	Yes	Yes	Kanagaraj et al., 2006; Machwe et al., 2006; Machwe et al., 2007; Machwe et al., 2011a; Berti et al., 2013
FANCM	Yes	ND	Gari et al., 2008a; Gari et al., 2008b; Xue et al., 2008
RAD54	Yes	Yes	Bugreev et al., 2011
FBH1	Yes	ND	Fugger et al., 2015

Table 5.2. List of fork remodeling proteins. Human proteins capable of performing fork reversal and fork restoration are listed with appropriate references. ND indicates activities that have not yet been determined.

in vitro (Bugreev et al., 2011; Fugger et al., 2015; Gari et al., 2008a; Gari et al., 2008b; Xue et al., 2008). Additionally, several members of the RECQ-like family of DNA helicases such as BLM, WRN, RECQ1, and RECQ5 can catalyze fork reversal or fork restoration using oligonucleotide or plasmid based assays (Berti et al., 2013; Kanagaraj et al., 2006; Machwe et al., 2011a; Machwe et al., 2011b; Machwe et al., 2006; Machwe et al., 2007; Popuri et al., 2008; Ralf et al., 2006). Much of the work characterizing the fork remodeling functions and preferences of these additional enzymes is in its preliminary stages. Experiments to exhaustively characterize their DNA binding preferences, mechanisms of regulation, and protein structure are needed for full comparisons to the SNF2 enzymes. These studies are important for describing the requirement of these proteins during DNA repair.

Among all the proteins capable of catalyzing fork reversal, only SMARCAL1 and ZRANB3 were shown to regress forks *in vivo* using electron microscopy (Kolinjivadi et al., 2017; Vujanovic et al., 2017). This observation was important as it validated what only *in vitro* and indirect cellular assays had shown previously (Badu-Nkansah et al., 2016a; Betous et al., 2013a; Betous et al., 2012; Bhat et al., 2015; Ciccia et al., 2012; Couch et al., 2013). While there are many proteins capable of performing these functions on naked or minimally decorated DNA *in vitro*, whether these enzymes can perform this function on DNA extensively coated with proteins or under torsional stress remains unexplored.

Despite the prevalence of fork remodeling proteins and the extensive work characterizing these proteins, several details about the mechanism of fork reversal remain undetermined. The ssDNA binding protein RAD51 was shown to be essential for fork reversal in response to various exogenous sources of replication stress, though it is not clear why RAD51 is needed (Zellweger et al., 2015). RAD51 is thought to replace RPA on the template DNA at sites of replication stress

and also stabilize the nascent ssDNA on the chicken foot structure, though this remains only a model (Neelsen and Lopes, 2015; Zellweger et al., 2015). Interestingly, co-depletion of SMARCAL1 and RAD51 additively reduces the number of reversed replication forks after exogenous stress (Kolinjivadi et al., 2017). This result raises the possibility that RAD51 is in an alternative pathway to SMARCAL1 in promoting fork reversal, though more work is needed to establish this model.

Concluding remarks

My work with SMARCAL1 resulted in the first identified source of replication stress that requires the function of this protein in unperturbed DNA replication. The work in this thesis required pioneering several new methods in the lab that will be beneficial in continuing work on the function of the SNF2 family of DNA translocases and their fork remodeling activities. Continuing this work will provide important details on the mechanisms of fork reversal in cells and endogenous sources of stress that trigger this response. Further, continuing my work to characterize when these enzymes are needed during replication will be invaluable for understanding the phenotypes associated with human diseases caused by defects in fork remodeling enzymes.

APPENDIX A

GENERATION OF *SMARCAL1*$^{-/-}$ CELLS USING CRISPR/CAS9

This appendix contains information associated with SMARCAL1-null cell lines that I generated using CRISPR/Cas9 genome editing. Using primers designed by a Thomas Bass, a student in the lab, I targeted exon 7 of the SMARCAL1 gene – which encodes for part of the N-terminal ATPase domain – and its upstream intron using a combination of two guide RNAs (gRNAs) (Figure AA.1A-B). I was able to generate two SMARCAL1 knockout U2OS clones (U2OS KO22 and KO 25) and one heterozygous knockout clone (U2OS KO33). Western blot analysis of SMARCAL1 protein in these clones indicates full length protein was not expressed in KO22 and KO25 while KO33 showed a faint band indicating SMARCAL1 protein was expressed but at a lower level (Figure AA.1C). PCR analysis of the SMARCAL1 locus near exon 7 and its upstream intron indicate editing occurred in all clones while KO33 showed multiple PCR products indicative of edited and unedited alleles (Figure AA.1D). The exact modification of each allele in these clones was not examined; however, the PCR products were sequenced and showed strong evidence of editing near the genomic locations targeted by the gRNAs (Figure AA.2-4). Cloning these PCR fragments and sequencing individual alleles is needed to determine the exact nature of the editing that occurred in these clones.
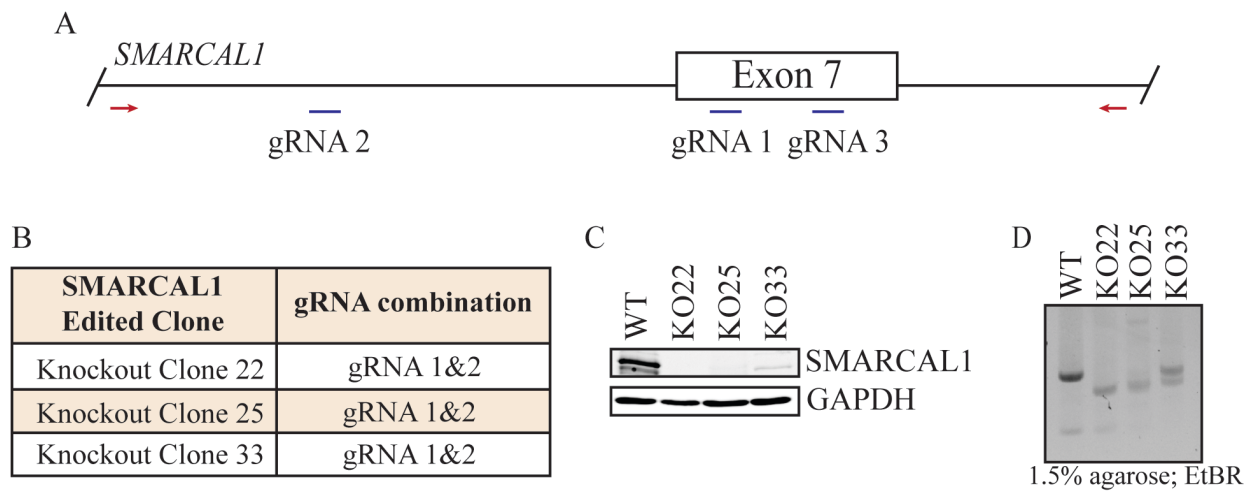
Figure AA.1. Overview of *SMARCAL1* **editing by CRISPR/Cas9.** (A) Schematic of SMARCAL1 locus with the locations targed by guide RNAs and PCR primers for screening (red arrows). (B) Overview of clones generated by CRISPR/Cas9 editing. (C) Western blot analysis of SMARCAL1 protein levels in edited clones. (D) PCR analysis of *SMARCAL1* locus in each edited clone.

A

```
SM1 genome    351   CAACCAACAGCTGCTGTGCTGGGTGGTCATTGTAGCTTTGTTCCCTCCTGTCATCCACAG   410
                    ||||||||||||||||||||||||
KO22          324   CAACCAACAGCTGCTGTGCTGGG-------------------------------------   346


SM1 genome    411   GCCTTCCTTCGGTGGCTGCCATCTCTGAGCCCAGATTGCATCAACGTCGTGGTGACTGGG   470

KO22          347   ------------------------------------------------------------   346


SM1 genome    471   AAGGACCGCCTGACAGCTGGCCTGATCAACATTGTCAGCTTTGACCTTCTTAGCAAGTTG   530
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
KO22          347   AAGGACCGCCTGACAGCTGGCCTGATCAACATTGTCAGCTTTGACCTTCTTAGCAAGTTG   406
```

B

Figure AA.2. CRISPR editing of SMARCAL1 locus in U2OS KO clone 22. (A) Serial Cloner local alignment between PCR product of KO22 and the genomic sequence of the SMARCAL1 locus. (B) FinchTV nucleotide trace of the indicate region of the KO22 PCR sequence. SM1, SMARCAL1.

A
```
SM1 genome  408   CAGGCCTTCCTTCGGTGGCTGCCATCTCTGAGCCCAGATTGCATCAACGTCGTGGTGACT   467
                  ||||||||||||||||||||||||||||||||||||||
KO25        378   CAGGCCTTCCTTCGGTGGCTGCCATCTCTGAGCCCA------------------------   413


SM1 genome  468   GGGAAGGACCGCCTGACAGCTGGCCTGATCAACATTGTCAGCTTTGACCTTCTTAGCAAG   527

KO25        414   ------------------------------------------------------------   413


SM1 genome  528   TTGGAAAAACAGCTAAAAAC-CCCTTTTAAAGTTGTCATCATTGTAAGAAACTTGGCAAA   586
                             : ||::| :| ||| || || :    | : || || | : ||
KO25        414   -------------N-AANNCNNCTTTTAAATGTGTNATCATTGNAAAAAAACTTGNCAAAG   459


SM1 genome  587   GTCTTTAAGTACTTTATCTCTCTGGAAACTTTCTTTGTAAAAGCTCTGAGAACTTTCATG   646
                  :|| :  ::|:    ::: | ||::|: || ::||| : ::: | :: |
KO25        460   TCNTTANCAANNTNATATNNNTGGAAANNTNCTTTGNNAAAATNANNNATAAGTNNACTGT   519


SM1 genome  647   TCTGTAATCTCCTTGAGCCCCATTTAGTTTCTAGAAGAATAGAAATCACTGTCTTTATTT   706
                  ::   |:|    :: : || :  |  :: |:::: :  |    :    ||   |:|
KO25        520   CTGTGNNCTCCNTGAGANNANTTTANTGTCTNNAANNNNANAGATCACNGTCTTTATTNT   579
```
B

Figure AA.3. CRISPR editing of SMARCAL1 locus in U2OS KO clone 25. (A) Serial Cloner local alignment between PCR product of KO25 and the genomic sequence of the SMARCAL1 locus. Blue sequence indicates gRNA #2; green sequence indicates gRNA #3. (B) FinchTV nucleotide trace of the indicated region of the KO25 PCR sequence. SM1, SMARCAL1.

A

```
SM1 genome   408  CAGGCCTTCCTTCGGTGGCTGCCATCTCTGAGCCCAGATTGCATCAACGTCGTGGTGACT   467
                  |||||||||||||||||||||||||||||||||||||||||||
KO25         378  CAGGCCTTCCTTCGGTGGCTGCCATCTCTGAGCCCA------------------------   413


SM1 genome   468  GGGAAGGACCGCCTGACAGCTGGCCTGATCAACATTGTCAGCTTTGACCTTCTTAGCAAG   527

KO25         414  ------------------------------------------------------------   413


SM1 genome   528  TTGGAAAAACAGCTAAAAAC-CCCTTTTAAAGTTGTCATCATTGTAAGAAACTTGGCAAA   586
                          : ||::| :| ||| || ||  :        | : || ||  | :  ||
KO25         414  -------------N-AANNCNNCTTTTAAATGTGTNATCATTGNAAAAAACTTGNCAAAG   459


SM1 genome   587  GTCTTTAAGTACTTTATCTCTCTGGAAACTTTCTTTGTAAAAGCTCTGAGAACTTTCATG   646
                      :|| :   ::|:     ::: | ||::|:  || ::||| : ::: |   :: |
KO25         460  TCNTTANCAANNTNATATNNNTGGAAANNTNCTTTGNNAAATNANNNATAAGTNNACTGT   519


SM1 genome   647  TCTGTAATCTCCTTGAGCCCCATTTAGTTTCTAGAAGAATAGAAATCACTGTCTTTATTT   706
                     ::    |:|     :: : || :  |   :: |::::  :  |    :   || |:|
KO25         520  CTGTGNNCTCCNTGAGANNANTTTANTGTCTNNAANNNNANAGATCACNGTCTTTATTNT   579
```
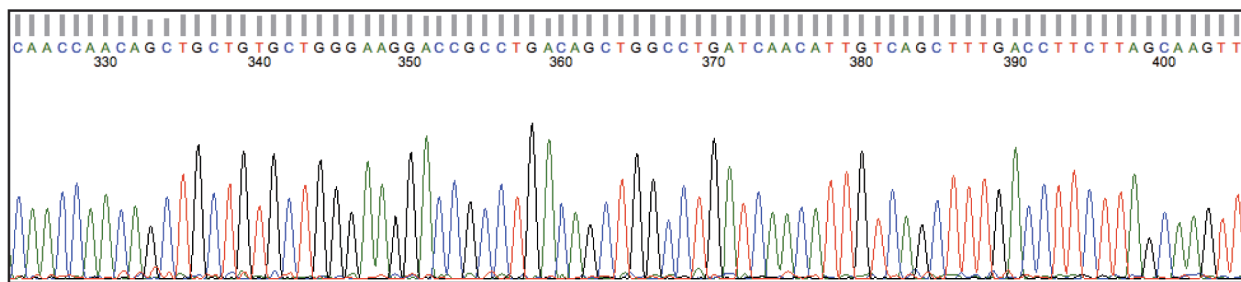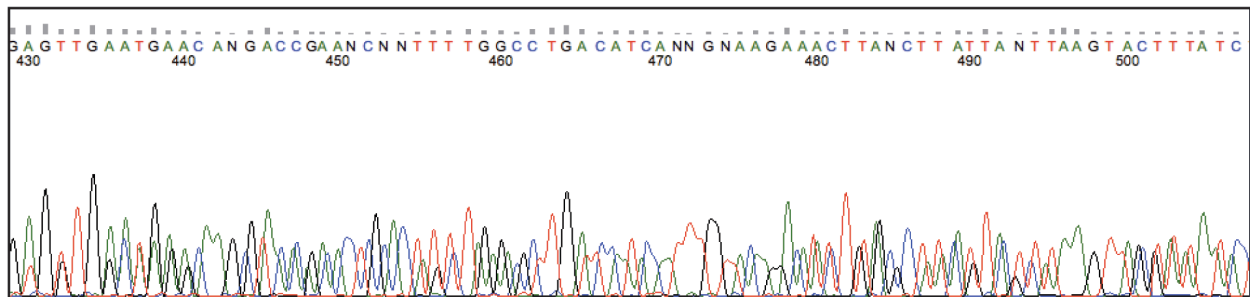
B

Figure AA.4. CRISPR editing of SMARCAL1 locus in U2OS KO clone 33. (A) Serial Cloner local alignment between PCR product of KO33 and the genomic sequence of the SMARCAL1 locus. (B) FinchTV nucleotide trace of the indicated region of the KO33 PCR sequence. SM1, SMARCAL1.

112

REFERENCES

Achar, Y.J., Balogh, D., and Haracska, L. (2011). Coordinated protein and DNA remodeling by human HLTF on stalled replication fork. Proc Natl Acad Sci U S A *108*, 14073-14078.

Achar, Y.J., Balogh, D., Neculai, D., Juhasz, S., Morocz, M., Gali, H., Dhe-Paganon, S., Venclovas, C., and Haracska, L. (2015). Human HLTF mediates postreplication repair by its HIRAN domain-dependent replication fork remodelling. Nucleic Acids Res *43*, 10277-10291.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L.*, et al.* (2013). Signatures of mutational processes in human cancer. Nature *500*, 415-421.

Aymard, F., Bugler, B., Schmidt, C.K., Guillou, E., Caron, P., Briois, S., Iacovoni, J.S., Daburon, V., Miller, K.M., Jackson, S.P.*, et al.* (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nat Struct Mol Biol *21*, 366-374.

Badu-Nkansah, A., Mason, A.C., Eichman, B.F., and Cortez, D. (2016a). Identification of a Substrate Recognition Domain in the Replication Stress Response Protein Zinc Finger Ran-binding Domain Containing Protein 3 (ZRANB3). J Biol Chem.

Badu-Nkansah, A., Mason, A.C., Eichman, B.F., and Cortez, D. (2016b). Identification of a Substrate Recognition Domain in the Replication Stress Response Protein Zinc Finger Ran-binding Domain-containing Protein 3 (ZRANB3). J Biol Chem *291*, 8251-8257.

Bansbach, C.E., Betous, R., Lovejoy, C.A., Glick, G.G., and Cortez, D. (2009). The annealing helicase SMARCAL1 maintains genome integrity at stalled replication forks. Genes Dev *23*, 2405-2414.

Bansbach, C.E., Boerkoel, C.F., and Cortez, D. (2010). SMARCAL1 and replication stress: an explanation for SIOD? Nucleus *1*, 245-248.

Baradaran-Heravi, A., Cho, K.S., Tolhuis, B., Sanyal, M., Morozova, O., Morimoto, M., Elizondo, L.I., Bridgewater, D., Lubieniecka, J., Beirnes, K.*, et al.* (2012a). Penetrance of biallelic SMARCAL1 mutations is associated with environmental and genetic disturbances of gene expression. Hum Mol Genet *21*, 2572-2587.

Baradaran-Heravi, A., Raams, A., Lubieniecka, J., Cho, K.S., DeHaai, K.A., Basiratnia, M., Mari, P.O., Xue, Y., Rauth, M., Olney, A.H.*, et al.* (2012b). SMARCAL1 deficiency predisposes to non-Hodgkin lymphoma and hypersensitivity to genotoxic agents in vivo. Am J Med Genet A *158A*, 2204-2213.

Barefield, C., and Karlseder, J. (2012). The BLM helicase contributes to telomere maintenance through processing of late-replicating intermediate structures. Nucleic Acids Res *40*, 7358-7367.

Barlow, J.H., Faryabi, R.B., Callen, E., Wong, N., Malhowski, A., Chen, H.T., Gutierrez-Cruz, G., Sun, H.W., McKinnon, P., Wright, G.*, et al.* (2013). Identification of early replicating fragile sites that contribute to genome instability. Cell *152*, 620-632.

Berti, M., Ray Chaudhuri, A., Thangavel, S., Gomathinayagam, S., Kenig, S., Vujanovic, M., Odreman, F., Glatter, T., Graziano, S., Mendoza-Maldonado, R.*, et al.* (2013). Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition. Nat Struct Mol Biol *20*, 347-354.

Betous, R., Couch, F.B., Mason, A.C., Eichman, B.F., Manosas, M., and Cortez, D. (2013a). Substrate-selective repair and restart of replication forks by DNA translocases. Cell Rep *3*, 1958-1969.

Betous, R., Glick, G.G., Zhao, R., and Cortez, D. (2013b). Identification and characterization of SMARCAL1 protein complexes. PLoS One *8*, e63149.

Betous, R., Mason, A.C., Rambo, R.P., Bansbach, C.E., Badu-Nkansah, A., Sirbu, B.M., Eichman, B.F., and Cortez, D. (2012). SMARCAL1 catalyzes fork regression and Holliday junction migration to maintain genome stability during DNA replication. Genes Dev *26*, 151-162.

Bhat, K.P., Betous, R., and Cortez, D. (2015). High-affinity DNA-binding domains of replication protein A (RPA) direct SMARCAL1-dependent replication fork remodeling. J Biol Chem *290*, 4110-4117.

Blasco, M.A. (2005). Telomeres and human disease: ageing, cancer and beyond. Nat Rev Genet *6*, 611-622.

Blastyak, A., Hajdu, I., Unk, I., and Haracska, L. (2010). Role of double-stranded DNA translocase activity of human HLTF in replication of damaged DNA. Mol Cell Biol *30*, 684-693.

Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. Nat Rev Genet *13*, 770-780.

Boerkoel, C.F., Takashima, H., John, J., Yan, J., Stankiewicz, P., Rosenbarker, L., Andre, J.L., Bogdanovic, R., Burguet, A., Cockfield, S.*, et al.* (2002). Mutant chromatin remodeling protein SMARCAL1 causes Schimke immuno-osseous dysplasia. Nat Genet *30*, 215-220.

Bugreev, D.V., Rossi, M.J., and Mazin, A.V. (2011). Cooperation of RAD51 and RAD54 in regression of a model replication fork. Nucleic Acids Res *39*, 2153-2164.

Burkovics, P., Sebesta, M., Balogh, D., Haracska, L., and Krejci, L. (2014). Strand invasion by HLTF as a mechanism for template switch in fork rescue. Nucleic Acids Res *42*, 1711-1720.

Byun, T.S., Pacek, M., Yee, M.C., Walter, J.C., and Cimprich, K.A. (2005). Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. Genes Dev *19*, 1040-1052.

Calado, R.T., and Dumitriu, B. (2013). Telomere dynamics in mice and humans. Semin Hematol *50*, 165-174.

Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B.P., and Nussenzweig, A. (2016). DNA Breaks and End Resection Measured Genome-wide by End Sequencing. Mol Cell *63*, 898-911.

Carroll, C., Badu-Nkansah, A., Hunley, T., Baradaran-Heravi, A., Cortez, D., and Frangoul, H. (2013). Schimke Immunoosseous Dysplasia associated with undifferentiated carcinoma and a novel SMARCAL1 mutation in a child. Pediatr Blood Cancer *60*, E88-90.

Carroll, C., Bansbach, C.E., Zhao, R., Jung, S.Y., Qin, J., and Cortez, D. (2014). Phosphorylation of a C-terminal auto-inhibitory domain increases SMARCAL1 activity. Nucleic Acids Res *42*, 918-925.

Cesare, A.J., and Reddel, R.R. (2010). Alternative lengthening of telomeres: models, mechanisms and implications. Nat Rev Genet *11*, 319-330.

Chastain, M., Zhou, Q., Shiva, O., Whitmore, L., Jia, P., Dai, X., Huang, C., Fadri-Moskwik, M., Ye, P., and Chai, W. (2016). Human CST Facilitates Genome-wide RAD51 Recruitment to GC-Rich Repetitive Sequences in Response to Replication Stress. Cell Rep *16*, 1300-1314.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J.*, et al.* (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107-119.

Chung, I., Osterwald, S., Deeg, K.I., and Rippe, K. (2012). PML body meets telomere: the beginning of an ALTernate ending? Nucleus *3*, 263-275.

Ciccia, A., Bredemeyer, A.L., Sowa, M.E., Terret, M.E., Jallepalli, P.V., Harper, J.W., and Elledge, S.J. (2009). The SIOD disorder protein SMARCAL1 is an RPA-interacting protein involved in replication fork restart. Genes Dev *23*, 2415-2425.

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

Ciccia, A., Nimonkar, A.V., Hu, Y., Hajdu, I., Achar, Y.J., Izhar, L., Petit, S.A., Adamson, B., Yoon, J.C., Kowalczykowski, S.C.*, et al.* (2012). Polyubiquitinated PCNA recruits the ZRANB3 translocase to maintain genomic integrity after replication stress. Mol Cell *47*, 396-409.

Clewing, J.M., Fryssira, H., Goodman, D., Smithson, S.F., Sloan, E.A., Lou, S., Huang, Y., Choi, K., Lucke, T., Alpay, H.*, et al.* (2007). Schimke immunoosseous dysplasia: suggestions of genetic diversity. Hum Mutat *28*, 273-283.

Coleman, M.A., Eisen, J.A., and Mohrenweiser, H.W. (2000). Cloning and characterization of HARP/SMARCAL1: a prokaryotic HepA-related SNF2 helicase protein from human and mouse. Genomics *65*, 274-282.

Cortez, D. (2015). Preventing replication fork collapse to maintain genome integrity. DNA Repair (Amst) *32*, 149-157.

Couch, F.B., Bansbach, C.E., Driscoll, R., Luzwick, J.W., Glick, G.G., Betous, R., Carroll, C.M., Jung, S.Y., Qin, J., Cimprich, K.A.*, et al.* (2013). ATR phosphorylates SMARCAL1 to prevent replication fork collapse. Genes Dev *27*, 1610-1623.

Cox, K.E., Marechal, A., and Flynn, R.L. (2016). SMARCAL1 Resolves Replication Stress at ALT Telomeres. Cell Rep *14*, 1032-1040.

Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K.*, et al.* (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat Methods *10*, 361-365.

de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev *19*, 2100-2110.

Drosopoulos, W.C., Kosiyatrakul, S.T., and Schildkraut, C.L. (2015). BLM helicase facilitates telomere replication during leading strand synthesis of telomeres. J Cell Biol *210*, 191-208.

Dungrawala, H., Bhat, K.P., Le Meur, R., Chazin, W.J., Ding, X., Sharan, S.K., Wessel, S.R., Sathe, A.A., Zhao, R., and Cortez, D. (2017). RADX Promotes Genome Stability and Modulates Chemosensitivity by Regulating RAD51 at Replication Forks. Mol Cell *67*, 374-386 e375.

Dungrawala, H., Rose, K.L., Bhat, K.P., Mohni, K.N., Glick, G.G., Couch, F.B., and Cortez, D. (2015). The Replication Checkpoint Prevents Two Types of Fork Collapse without Regulating Replisome Stability. Mol Cell *59*, 998-1010.

Flaus, A., Martin, D.M., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res *34*, 2887-2905.

Flaus, A., and Owen-Hughes, T. (2011). Mechanisms for ATP-dependent chromatin remodelling: the means to the end. FEBS J *278*, 3579-3595.

Fugger, K., Mistrik, M., Neelsen, K.J., Yao, Q., Zellweger, R., Kousholt, A.N., Haahr, P., Chu, W.K., Bartek, J., Lopes, M*., et al.* (2015). FBH1 Catalyzes Regression of Stalled Replication Forks. Cell Rep.

Gari, K., Decaillet, C., Delannoy, M., Wu, L., and Constantinou, A. (2008a). Remodeling of DNA replication structures by the branch point translocase FANCM. Proc Natl Acad Sci U S A *105*, 16107-16112.

Gari, K., Decaillet, C., Stasiak, A.Z., Stasiak, A., and Constantinou, A. (2008b). The Fanconi anemia protein FANCM can promote branch migration of Holliday junctions and replication forks. Mol Cell *29*, 141-148.

Ghosal, G., Yuan, J., and Chen, J. (2011). The HARP domain dictates the annealing helicase activity of HARP/SMARCAL1. EMBO Rep *12*, 574-580.

Gilson, E., and Geli, V. (2007). How telomeres are replicated. Nat Rev Mol Cell Biol *8*, 825-838.

Gong, Y., and de Lange, T. (2010). A Shld1-controlled POT1a provides support for repression of ATR signaling at telomeres through RPA exclusion. Mol Cell *40*, 377-387.

Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H., and de Lange, T. (1999). Mammalian Telomeres End in a Large Duplex Loop. Cell *97*, 503-514.

Haokip, D.T., Goel, I., Arya, V., Sharma, T., Kumari, R., Priya, R., Singh, M., and Muthuswami, R. (2016). Transcriptional Regulation of Atp-Dependent Chromatin Remodeling Factors: Smarcal1 and Brg1 Mutually Co-Regulate Each Other. Sci Rep *6*, 20532.

Helleday, T., Eshtad, S., and Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet *15*, 585-598.

Henson, J.D., Cao, Y., Huschtscha, L.I., Chang, A.C., Au, A.Y., Pickett, H.A., and Reddel, R.R. (2009). DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. Nat Biotechnol *27*, 1181-1185.

Hishiki, A., Hara, K., Ikegaya, Y., Yokoyama, H., Shimizu, T., Sato, M., and Hashimoto, H. (2015). Structure of a Novel DNA-binding Domain of Helicase-like Transcription Factor (HLTF) and Its Functional Implication in DNA Damage Tolerance. J Biol Chem *290*, 13215-13223.

Hoege, C., Pfander, B., Moldovan, G.-L., Pyrowolakis, G., and Jensch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature *419*, 135-141.

Iacovoni, J.S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D., and Legube, G. (2010). High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. EMBO J *29*, 1446-1457.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Kanagaraj, R., Saydam, N., Garcia, P.L., Zheng, L., and Janscak, P. (2006). Human RECQ5beta helicase promotes strand exchange on synthetic DNA structures resembling a stalled replication fork. Nucleic Acids Res *34*, 5217-5231.

Kassavetis, G.A., and Kadonaga, J.T. (2014). The annealing helicase and branch migration activities of Drosophila HARP. PLoS One *9*, e98173.

Keka, I.S., Mohiuddin, Maede, Y., Rahman, M.M., Sakuma, T., Honma, M., Yamamoto, T., Takeda, S., and Sasanuma, H. (2015). Smarcal1 promotes double-strand-break repair by nonhomologous end-joining. Nucleic Acids Res *43*, 6359-6372.

Kile, A.C., Chavez, D.A., Bacal, J., Eldirany, S., Korzhnev, D.M., Bezsonova, I., Eichman, B.F., and Cimprich, K.A. (2015). HLTF's Ancient HIRAN Domain Binds 3' DNA Ends to Drive Replication Fork Reversal. Mol Cell *58*, 1090-1100.

Kolinjivadi, A.M., Sannino, V., De Antoni, A., Zadorozhny, K., Kilkenny, M., Techer, H., Baldi, G., Shen, R., Ciccia, A., Pellegrini, L*., et al.* (2017). Smarcal1-Mediated Fork Reversal Triggers Mre11-Dependent Degradation of Nascent DNA in the Absence of Brca2 and Stable Rad51 Nucleofilaments. Mol Cell *67*, 867-881 e867.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495-501.

Lin, J.R., Zeman, M.K., Chen, J.Y., Yee, M.C., and Cimprich, K.A. (2011). SHPRH and HLTF act in a damage-specific manner to coordinate different forms of postreplication repair and prevent mutagenesis. Mol Cell *42*, 237-249.

Lopez-Saavedra, A., Gomez-Cabello, D., Dominguez-Sanchez, M.S., Mejias-Navarro, F., Fernandez-Avila, M.J., Dinant, C., Martinez-Macias, M.I., Bartek, J., and Huertas, P. (2016). A genome-wide screening uncovers the role of CCAR2 as an antagonist of DNA end resection. Nat Commun *7*, 12364.

Lou, S., Lamfers, P., McGuire, N., and Boerkoel, C.F. (2002). Longevity in Schimke immuno-osseous dysplasia. Journal of Medical Genetics, 922-925.

Machwe, A., Karale, R., Xu, X., Liu, Y., and Orren, D.K. (2011a). The Werner and Bloom syndrome proteins help resolve replication blockage by converting (regressed) holliday junctions to functional replication forks. Biochemistry *50*, 6774-6788.

Machwe, A., Lozada, E., Wold, M.S., Li, G.M., and Orren, D.K. (2011b). Molecular cooperation between the Werner syndrome protein and replication protein A in relation to replication fork blockage. J Biol Chem *286*, 3497-3508.

Machwe, A., Xiao, L., Groden, J., and Orren, D.K. (2006). The Werner and Bloom syndrome proteins catalyze regression of a model replication fork. Biochemistry *45*, 13939-13946.

Machwe, A., Xiao, L., Lloyd, R.G., Bolt, E., and Orren, D.K. (2007). Replication fork regression in vitro by the Werner syndrome protein (WRN): holliday junction formation, the effect of leading arm structure and a potential role for WRN exonuclease activity. Nucleic Acids Res *35*, 5729-5747.

MacKay, C., Toth, R., and Rouse, J. (2009). Biochemical characterisation of the SWI/SNF family member HLTF. Biochem Biophys Res Commun *390*, 187-191.

Mailand, N., Gibbs-Seymour, I., and Bekker-Jensen, S. (2013). Regulation of PCNA-protein interactions for genome stability. Nat Rev Mol Cell Biol *14*, 269-282.

Mason, A.C., Rambo, R.P., Greer, B., Pritchett, M., Tainer, J.A., Cortez, D., and Eichman, B.F. (2014). A structure-specific nucleic acid-binding domain conserved among DNA repair proteins. Proc Natl Acad Sci U S A *111*, 7618-7623.

Massip, L., Caron, P., Iacovoni, J.S., Trouche, D., and Legube, G. (2010). Deciphering the chromatin landscape induced around DNA double strand breaks. Cell Cycle *9*, 2963-2972.

Masuda, Y., Suzuki, M., Kawai, H., Hishiki, A., Hashimoto, H., Masutani, C., Hishida, T., Suzuki, F., and Kamiya, K. (2012). En bloc transfer of polyubiquitin chains to PCNA in vitro is mediated by two different human E2-E3 pairs. Nucleic Acids Res *40*, 10394-10407.

Matsuzaki, K., Borel, V., Adelman, C.A., Schindler, D., and Boulton, S.J. (2015). FANCJ suppresses microsatellite instability and lymphomagenesis independent of the Fanconi anemia pathway. Genes Dev *29*, 2532-2546.

Moinova, H.R., Chen, W.D., Shen, L., Smiraglia, D., Olechnowicz, J., Ravi, L., Kasturi, L., Myeroff, L., Plass, C., Parsons, R*., et al.* (2002). HLTF gene silencing in human colon cancer. Proc Natl Acad Sci U S A *99*, 4562-4567.

Motegi, A., Liaw, H.J., Lee, K.Y., Roest, H.P., Maas, A., Wu, X., Moinova, H., Markowitz, S.D., Ding, H., Hoeijmakers, J.H*., et al.* (2008). Polyubiquitination of proliferating cell nuclear antigen by HLTF and SHPRH prevents genomic instability from stalled replication forks. Proc Natl Acad Sci U S A *105*, 12411-12416.

Muthuswami, R., Truman, P.A., Mesner, L.D., and Hockensmith, J.W. (2000). A eukaryotic SWI2/SNF2 domain, an exquisite detector of double-stranded to single-stranded DNA transition elements. J Biol Chem *275*, 7648-7655.

Neelsen, K.J., and Lopes, M. (2015). Replication fork reversal in eukaryotes: from dead end to dynamic response. Nat Rev Mol Cell Biol *16*, 207-220.

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

O'Sullivan, R.J., Arnoult, N., Lackner, D.H., Oganesian, L., Haggblom, C., Corpet, A., Almouzni, G., and Karlseder, J. (2014). Rapid induction of alternative lengthening of telomeres by depletion of the histone chaperone ASF1. Nat Struct Mol Biol *21*, 167-174.

Oakley, G.O., and Patrick, S.M. (2010). Replication protein A: directing traffic at the intersection of replication and repair. Frontiers in Bioscience *15*, 883-900.

Patne, K., Rakesh, R., Arya, V., Chanana, U.B., Sethy, R., Swer, P.B., and Muthuswami, R. (2017). BRG1 and SMARCAL1 transcriptionally co-regulate DROSHA, DGCR8 and DICER in response to doxorubicin-induced DNA damage. Biochim Biophys Acta *1860*, 936-951.

Poole, L.A., and Cortez, D. (2017). Functions of SMARCAL1, ZRANB3, and HLTF in maintaining genome stability. Crit Rev Biochem Mol Biol, 1-19.

Poole, L.A., Zhao, R., Glick, G.G., Lovejoy, C.A., Eischen, C.M., and Cortez, D. (2015). SMARCAL1 maintains telomere integrity during DNA replication. Proc Natl Acad Sci U S A *112*, 14864-14869.

Popuri, V., Bachrati, C.Z., Muzzolini, L., Mosedale, G., Costantini, S., Giacomini, E., Hickson, I.D., and Vindigni, A. (2008). The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities. J Biol Chem *283*, 17766-17776.

Popuri, V., Hsu, J., Khadka, P., Horvath, K., Liu, Y., Croteau, D.L., and Bohr, V.A. (2014). Human RECQL1 participates in telomere maintenance. Nucleic Acids Res *42*, 5671-5688.

Postow, L., Woo, E.M., Chait, B.T., and Funabiki, H. (2009). Identification of SMARCAL1 as a component of the DNA damage response. J Biol Chem *284*, 35951-35961.

Puccetti, M.V., Fischer, M.A., Arrate, M.P., Boyd, K.L., Duszynski, R.J., Betous, R., Cortez, D., and Eischen, C.M. (2017). Defective replication stress response inhibits lymphomagenesis and impairs lymphocyte reconstitution. Oncogene *36*, 2553-2564.

Ralf, C., Hickson, I.D., and Wu, L. (2006). The Bloom's syndrome helicase can promote the regression of a model replication fork. J Biol Chem *281*, 22839-22846.

Rivera, T., Haggblom, C., Cosconati, S., and Karlseder, J. (2017). A balance between elongation and trimming regulates telomere stability in stem cells. Nat Struct Mol Biol *24*, 30-39.

Sandhu, S., Wu, X., Nabi, Z., Rastegar, M., Kung, S., Mai, S., and Ding, H. (2012). Loss of HLTF function promotes intestinal carcinogenesis. Mol Cancer *11*, 18.

Sebesta, M., Cooper, C.D.O., Ariza, A., Carnie, C.J., and Ahel, D. (2017). Structural insights into the function of ZRANB3 in replication stress response. Nat Commun *8*, 15847.

Sfeir, A., Kosiyatrakul, S.T., Hockemeyer, D., MacRae, S.L., Karlseder, J., Schildkraut, C.L., and de Lange, T. (2009). Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. Cell *138*, 90-103.

Sharma, T., Bansal, R., Haokip, D.T., Goel, I., and Muthuswami, R. (2015). SMARCAL1 Negatively Regulates C-Myc Transcription By Altering The Conformation Of The Promoter Region. Sci Rep *5*, 17910.

Svendsen, J.M., Smogorzewska, A., Sowa, M.E., O'Connell, B.C., Gygi, S.P., Elledge, S.J., and Harper, J.W. (2009). Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. Cell *138*, 63-77.

Takai, K.K., Kibe, T., Donigian, J.R., Frescas, D., and de Lange, T. (2011). Telomere protection by TPP1/POT1 requires tethering to TIN2. Mol Cell *44*, 647-659.

Treangen, T.J., and Salzberg, S.L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet *13*, 36-46.

Unk, I., Hajdu, I., Blastyak, A., and Haracska, L. (2010). Role of yeast Rad5 and its human orthologs, HLTF and SHPRH in DNA damage tolerance. DNA Repair (Amst) *9*, 257-267.

Unk, I., Hajdu, I., Fatyol, K., Hurwitz, J., Yoon, J.H., Prakash, L., Prakash, S., and Haracska, L. (2008). Human HLTF functions as a ubiquitin ligase for proliferating cell nuclear antigen polyubiquitination. Proc Natl Acad Sci U S A *105*, 3768-3773.

Vannier, J.B., Pavicic-Kaltenbrunner, V., Petalcorin, M.I., Ding, H., and Boulton, S.J. (2012). RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. Cell *149*, 795-806.

Vujanovic, M., Krietsch, J., Raso, M.C., Terraneo, N., Zellweger, R., Schmid, J.A., Taglialatela, A., Huang, J.W., Holland, C.L., Zwicky, K*., et al.* (2017). Replication Fork Slowing and Reversal upon DNA Damage Require PCNA Polyubiquitination and ZRANB3 DNA Translocase Activity. Mol Cell *67*, 882-890 e885.

Weston, R., Peeters, H., and Ahel, D. (2012). ZRANB3 is a structure-specific ATP-dependent endonuclease involved in replication stress response. Genes Dev *26*, 1558-1572.

Xue, Y., Li, Y., Guo, R., Ling, C., and Wang, W. (2008). FANCM of the Fanconi anemia core complex is required for both monoubiquitination and DNA repair. Hum Mol Genet *17*, 1641-1652.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat Immunol *12*, 62-69.

Yamane, A., Robbiani, D.F., Resch, W., Bothmer, A., Nakahashi, H., Oliveira, T., Rommel, P.C., Brown, E.J., Nussenzweig, A., Nussenzweig, M.C*., et al.* (2013). RPA accumulation during class switch recombination represents 5'-3' DNA-end resection during the S-G2/M phase of the cell cycle. Cell Rep *3*, 138-147.

Yuan, J., Ghosal, G., and Chen, J. (2009). The annealing helicase HARP protects stalled replication forks. Genes Dev *23*, 2394-2399.

Yuan, J., Ghosal, G., and Chen, J. (2012). The HARP-like domain-containing protein AH2/ZRANB3 binds to PCNA and participates in cellular response to replication stress. Mol Cell *47*, 410-421.

Yusufzai, T., and Kadonaga, J.T. (2008). HARP Is an ATP-Driven Annealing Helicase. Science *322*, 748-750.

Yusufzai, T., and Kadonaga, J.T. (2010). Annealing helicase 2 (AH2), a DNA-rewinding motor with an HNH motif. Proc Natl Acad Sci U S A *107*, 20970-20973.

Yusufzai, T., Kong, X., Yokomori, K., and Kadonaga, J.T. (2009). The annealing helicase HARP is recruited to DNA repair sites via an interaction with RPA. Genes Dev *23*, 2400-2404.

Zellweger, R., Dalcher, D., Mutreja, K., Berti, M., Schmid, J.A., Herrador, R., Vindigni, A., and Lopes, M. (2015). Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. J Cell Biol *208*, 563-579.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. Nat Cell Biol *16*, 2-9.