REGULATION OF DNA REPLICATION INITIATION AND ELONGATION

by

Gemma Louise Regan-Mochrie

A Dissertation

Presented to the Faculty of the

Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences,

Memorial Sloan-Kettering Cancer Center

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

New York, NY

December 2022

_____ _____

Xiaolan Zhao PhD Date

Dissertation Mentor

To my wife, Ellie, who has supported me throughout my Ph.D. and to my family who encouraged me in every way and who inspired my love of science.

ABSTRACT

DNA replication is fundamental to all forms of life. Defects in this process can lead to loss of genomic integrity and a variety of diseases from developmental defects to cancer. As such, it comes as no surprise that DNA replication is highly regulated to ensure that the entire genome is duplicated once and only once per cell cycle. Understanding the regulation of each step of replication is critical to understanding how DNA is faithfully transmitted during cell division, from generation to generation, and how various mistakes can lead to deleterious effects. This dissertation focuses on how DNA replication initiation and elongation are regulated.

Replication initiates at sites across the genome throughout S phase in a replication firing program. The regulation of this firing timing is essential to fully complete genome replication and withstand DNA damage. In this study we aimed to understand how the SUMO-based post-translational modification of a key replication complex, the Origin Recognition Complex (ORC), regulates this firing process. We found that several subunits of this complex are sumoylated and that appropriate levels of ORC sumoylation are key to properly activate early origins. We used two complementary methods to examine the effects of either hyper-sumoylating or hypo-SUMOylating ORC. Hyper-sumoylation of ORC was achieved via the attachment of a specialized SuON tag that increases the local sumoylation near the protein it is covalently bound to. Hypo-sumoylation was achieved by mutating a key sumoylated lysine on Orc2 that we mapped using mass spectrometry. Our data using these two genetic constructs showed that sumoylation of ORC regulates early origin firing to ensure complete replication of the genome and that perturbing this system with improper levels sumoylation results in genomic instability.

After replication has initiated, replication elongation can begin to form the nascent DNA. The leading strand polymerase has the difficult job of replicating large stretches of the genome continuously. The catalytic subunit of the leading strand polymerase contains a unique and highly conserved domain called POPS. Using the budding yeast leading strand polymerase Pol2 as a model, a previous lab member found that this domain was important for replication

initiation and elongation. However, the role of POPS in elongation had not been thoroughly explored. In this work we used cancer associated mutations to impair the function of this domain, resulting greater reliance on DNA repair factors and of loss of symmetry of sister replication forks. These phenotypes were rescued with an additional mutation in the Pol2 exonuclease domain. These data suggested that this domain contributes to maintaining the balance between polymerization and exonuclease activity to ensure replication stability. Collectively my thesis work contributes to the revelation of two mechanisms that regulate DNA replication, both of which are important to maintain genomic integrity.

BIOGRAPHICAL SKETCH

Gemma Louise Regan-Mochrie was born January 7th 1994 in New Haven, CT to Lynne Regan and Simon Mochrie. From a young age her parents inspired her love of science and curiosity about the world around her. In 2011 she entered Smith College where she worked in the laboratory of Christine White-Ziegler where she studied the effects of temperature on gene expression and biofilm formation in uropathogenic and commensal E. coli. She also spent a summer working in the laboratory of Dr. Maria Jasin where she studied the role of a ligase in DNA repair. This experience sparked her interest in DNA replication and repair. In 2015 she graduated with a Bachelor of Arts in Biochemistry and high honors for her thesis work in the White-Ziegler laboratory. She went on to enroll in the Cancer Biology PhD program at the Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences at MSKCC in the same year. She joined Xiaolan Zhao's laboratory where she has focused on the regulation of DNA replication.

ACKNOWLEDGEMENTS

Many people have contributed to my scientific education and personal growth over the past seven years. I'm incredibly thankful for their guidance, encouragement, and time.

First, I would like to thank my thesis mentor Xiaolan. She has always inspired me with her dedication, knowledge, rigor, and boundless excitement for understanding the way the world works. Xiaolan has helped me expand my knowledge and thinking in many ways. One of the most important ideas that she has instilled in me is to have a growth mindset, no matter where we are in life or how old you are there is always more to learn, and I should never be afraid to ask a question or seek advice. Thank you so much for helping me grow both scientifically and personally and providing me with fantastic colleagues in the lab.

I would like to thank my committee members Dirk Remus and Iestyn Whitehouse, who have always offered extremely valuable comments and advice. They also collaborated on both the projects described here, for which I am very thankful. I would also very much like to thank Angel Sfeir for agreeing to be my chair and Duncan Smith for agreeing to be my external examiner.

I also would like to thank all the members of the Zhao lab who I have worked alongside. It has been such a wonderful group of people that I have had the privilege of spending the last six years with and who have helped me immeasurably. I would especially like to thank Michael Meng for his innumerable advice and for his hard work on the Pol2 project; Shibai Li who has been a great friend and has helped me troubleshoot more experiments than I can remember; Nalini Dhingra for being a fantastic friend, mentor and confidant who has been a great help in overcoming setbacks and editing my writing. Jake Bonner and Bingbing Wan for their experimental expertise and friendship, especially when I first joined the lab. Jiayi Fan for being a great friend and bay mate since she joined the lab.

I am very grateful to all my collaborators, especially Timothy Hoggard, Catherine Fox, Nikhil Bhagwat, Neil Hunter, Gerard Lynch and Clemence Claussin whose work has directly

contributed to the work presented in my dissertation. Tim and Nikhil have especially spent a lot of time and energy in this collaboration, which I am very grateful for. I would also like to thank Yuangliang Zhai and Bik Kwoon Tye and their lab members for very helpful discussions.

I would like to thank my family and friends without whose support and love this PhD would not have been possible. I would like all my GSK peers who have always been a joy to spend time with, especially Rui, Paige, and Emily who started in this program with me and have become lifelong friends. Outside of GSK Brandon and Jen, I am so thankful to have met you and for our enduring friendship. Krithika who I have known for over 10 years, and I am so grateful that you decided to do your PhD across the street, I could not have done this without you.

Thank you to my brother James for always taking me out of myself when we spend time together and always being there to listen. To my parents, who inspired my love of science from a very young age and always encouraged me to explore the world around me. Your support of and belief in me has been incredibly important. To Ellie, you have been there for me in ways that I couldn't have imagined and didn't even know I needed. Finally thank you to my cat, Freyja, for sometimes giving me a cuddle when you feel like it.

TABLE OF CONTENTS

List of Figures 3

List of Abbreviations 5

Chapter 1 : Introduction 1

1.1 Overview of DNA replication using Saccharomyces cerevisiae as a model system. 1

1.2 ORC loading of the MCM replicative helicase onto dsDNA 3
1.2.1 The Autonomously Replicating Sequence 4
1.2.2 ORC-DNA interaction 4
1.2.3 MCM loading 6
1.2.4 MCM activation 7

1.3 Genome-wide view of origin firing 9
1.3.1 Origin activation: Origin efficiency 10
1.3.2 Origin Activation: ORC association with origins 11
1.3.3 Origin Activation: Repetitive DNA sequences 11
1.3.4 Origin Activation: Chromatin environment and other external control pathways 13
1.3.5 Preventing Re-Replication 14

1.4 Replication Elongation 15
1.4.1 Overview of three replicative DNA Polymerases 15
1.4.2 Proofreading by Pol ε and Pol δ 16
1.4.3 The POPS domain of polymerase ε 17

1.5 SUMOylation in DNA replication 17
1.5.1 The SUMO modification cycle and enzymes 18
1.5.2 SUMO site selection and additional methods to alter sumoylation levels 19
1.5.3 SUMOylation in DNA replication 19

1.6 Dissertation Research Aims 21

Chapter 2 : The Origin Recognition Complex sumoylation status fine tunes origin activity 23

2.1 Introduction 23

2.2 Results 25
2.2.1 Multiple ORC subunits were sumoylated during normal growth 25
2.2.2 ORC2-SuON caused ORC hyper-sumoylation and cell lethality 26
2.2.3 Lethality caused by ORC hyper-sumoylation was rescued by reducing SUMO E2 function. 27
2.2.4 ORC hyper-sumoylation reduced the firing of a subset of origins 28
2.2.5 ORC hyper-sumoylation reduced MCM recruitment to origin DNA 30
2.2.6 ORC hyper-sumoylation compromised rDNA replication and stability 32
2.2.7 Identification of the major sumoylation site on Orc2 32
2.2.8 Orc2 hypo-sumoylation enhanced early origin firing 33
2.2.9 ORC hypo- or hyper-sumoylation generated a dependence on other genome factors 35

2.3 Discussion 36
2.3.1 Biological consequences of ORC hyper-sumoylation and hypo-sumoylation 37
2.3.2 Proper ORC sumoylation status is required for genomic stability and cell fitness 38
2.3.3 ORC sumoylation acts to modulate the efficiency of origin licensing 39
2.3.4 Models for ORC sumoylation dynamics and future challenges 40

2.4 Materials and Methods 41

2.5 Acknowledgments. 49

2.6 Figures *50*

Chapter 3 : The Balancing act of the Leading Strand Polymerase **66**

3.1 Introduction *66*

3.2 Results *68*
3.2.1 Slow growth of a POPS mutant was rescued by abolishing the Pol2 exonuclease activity. 68
3.2.2 EXO inactivation reduced pol2-REL genome instability and reliance on other genome factors 69
3.2.3 Abolishing Pol2 exonuclease activity improved genome replication of a POPS mutant that showed normal Mrc1 and Csm3 association with the replicative helicase 70
3.2.4 Pol2 exonuclease activity partly accounts for genome-wide sister fork asymmetry and replisome progression defect of a POPS mutant. 71

3.3 Discussion *73*
3.3.1 The division of labor between he leading and lagging strand polymerases 73

3.4 Materials and Methods *74*

3.5 Acknowledgments *77*

Chapter 4 : Conclusions and Future Directions **89**

4.1 Summary *89*

4.2 ORC sumoylation – a new layer of control for origin usage *90*

4.3 Possible mechanisms by which ORC sumoylation interferes with origin usage *93*

4.4 Proper ORC sumoylation is required for optimal genomic stability and cell fitness *94*

4.6 Pol2 POPS contributes to the balance between polymerase and exonuclease function *95*

4.7 Mechanisms of POPS's roles in processive leading strand synthesis? *96*

References **99**

LIST OF FIGURES

Figure 1.1 ORC, Cdc6, and MCM protein domains. 4
Figure 1.2 Seven-steps of pre-replicative complex assembly adapted from (Stillman 2022). 5
Figure 1.3 The main Steps involved in MCM licensing in S phase. 8
Figure 1.4 Genome-wide control of origin firing 12
Figure 1.5 Overview of the sumoylation cycle in budding yeast. 19
Figure 2.1 Increasing ORC sumoylation led to cell lethality that could be rescued by a SUMO E2 mutant. 50
Figure 2.2 Examination of ORC sumoylation. 51
Figure 2.3 ORC2-SuON led to preferential inhibition of a subset of early origins and a temporally homogenized replication profile. 52
Figure 2.4 Copy number analyses of the whole genome sequencing data for ORC2-SuON and ORC2-SuCtrl cells. 53
Figure 2.5 ORC hyper-sumoylation prevented MCM loading. 54
Figure 2.6 ORC hyper-sumoylation prevented MCM loading on DNA. 55
Figure 2.7 ORC2-SuON resulted in inhibition of chromosomal replication with the most severe effects on the rDNA-containing chromosome. 56
Figure 2.8 Mapping the Orc2 sumoylation sites. 57
Figure 2.9 Mass spectrometry data indicate sumoylation at K406 of Orc2. 58
Figure 2.10 Orc2 hypo-sumoylation increased early origin firing. 59
Figure 2.11 Copy number analyses of the whole genome sequencing data for ORC2-HA and orc2-K406R-HA cells. 60
Figure 2.12 Genetic analysis of the ORC hypo- or hyper-sumoylation alleles and a structural model for Orc2 sumoylation. 61
Figure 2.2.13 Genetic analyses of ORC2-SuCtrl and ORC2-HA. 62
Figure 2.14 ORC hyper- and hypo-sumoylation increased checkpoint activation. 63
Figure 2.15 Orc2 sumoylation level progressively increases as S phase progresses and reduces in G2-M phase. . 64
Figure 3.1 Inactivating the Pol2 exonuclease suppressed the growth defect and genomic instability of a POPS mutant. 78
Figure 3.2 Examination of protein levels and growth of pol2 mutants. 79
Figure 3.3 pol2-REL replication defect was improved by inactivating the Pol2 exonuclease and was not associated with reduced CMG-bound Mrc1 and Csm3. 80
Figure 3.4 Examination of a genetic interaction and the FACS profile of pol2 mutants 81
Figure 3.5 FACS profiles for wild-type and pol2-REL cells examined for Mrc1 and Csm3. 82
Figure 3.3.6 Replicon-seq analyses revealed genome-wide sister fork asymmetry of pol2-REL which was improved by EXO inactivation. 83
Figure 3.7 Analysis of DNA replication with Replicon-seq. 84
Figure 3.8 Sister replication fork asymmetry and leading strand DNA synthesis in pol2 mutants 85
Figure 3.9 Analysis of sister replisome progression with Replicon-seq. 86
Figure 3.10 A model for the antagonistic action of the Pol2's POPS and EXO domains in leading strand synthesis. 87
Figure 4.1 ORC sumoylation models 91

LIST OF TABELS

Table 1: Strains used in Chapter 2 ..... *65*
Table 2. Strains used in Chapter 3 ..... *88*

LIST OF ABBREVIATIONS

ARS: Autonomously Replicating Sequence

ACS: ARS consensus sequence

CDK: Cyclin Dependent Kinase

CMG: Cdc45-MCM2-7-GINS

DDK: Dbf4 dependent Kinase

EXO: Exonuclease domain

GINS: Go-ichi-ni-san – complex made of Sld5-Psf1, Psf2, Psf3

MCM: Mini-chromosome maintenance, complex comprised of Mcm2-7

OCCM: ORC-Cdc6-Cdt1-MCM2-7

ORC: Origin Recognition Complex

POL: Polymerase domain

Pol δ: Polymerase δ

Pol ε: Polymerase ε

POPS: Pol2 family-specific catalytic core peripheral subdomain

SUMO: Small ubiquitin-like modifier

CHAPTER 1 : INTRODUCTION

1.1 Overview of DNA replication using *Saccharomyces cerevisiae* as a model system.

DNA replication begins at sites across the genome called origins. The Origin Recognition Complex (ORC) binds to these sites to give them the potential to initiate DNA replication. In budding yeast (*Saccharomyces cerevisiae*), these replication origins are defined by a DNA sequence called the ARS (Autonomously Replicating Sequence). Despite this initially appearing a straightforward mechanism to initiate DNA replication, there is additional complexity that dictates which of these sequences are used during replication. First, not all ARS sequences that have the potential to bind to ORC can recruit ORC in cells. Second, different ARS consensus sequences that do bind ORC *in vivo* have variable affinity for the complex. Third, the differential ability of ARS-ORC interaction in cells can contribute to differences in origin activation or firing.

ORC is responsible for recruiting the MCM (Mini-Chromosome Maintenance protein complex) replicative helicase to begin DNA replication initiation. This recruitment can be impacted by other proteins and the chromatin environment around a particular origin. The MCM recruitment step is highly regulated to ensure that the whole genome is replicated once and only once per cell cycle and that replication is completed by the time cells are ready to divide.

Current studies suggest that origin activation requires two tightly regulated and temporally separated steps to ensure faithful genome replication. The first step is helicase loading, beginning with MCM recruitment, in which MCM is loaded onto origin DNA in G1 phase. The second step is origin activation, in which MCM is activated to begin unwinding dsDNA throughout S phase. The temporal separation of origin licensing (MCM loading) and origin firing (MCM activation) provides cells the opportunity to bestow multi-layered regulation of both. For example, throughout S phase, MCM activation occurs at early and late origins at different times.

Additionally, the same set of origins will not necessarily be activated in each cell during an individual cell cycle. Origin efficiency refers to the likelihood of an origin being activated in each cell cycle. Some origins are highly efficient and are frequently used, while others remain dormant most of the time, only activated to complete replication if cells experience replication stress (Tuduri et al. 2010). The global regulation of origin firing affects genome stability and cell-fate determination, and thus is a vital aspect of genome maintenance (Smith et al. 2001; Watanabe and Maekawa 2010; Ryba et al. 2011; Debatisse et al. 2012; Pozo et al. 2018).

During the transition to S phase MCM is activated through a series of phosphorylation and protein recruitment steps to begin unwinding the parental DNA. This allows for the replicative polymerases, Polymerase ε (Pol ε), Polymerase δ (Pol δ), and Polymerase α (Pol α) to polymerize the nascent strand. Pol α elongates the primer by synthesizing short DNA stretches to allow replication to begin. Pol ε is the leading strand polymerase that must carry out continuous synthesis over large stretches of the genome, while Pol δ is the lagging strand polymerase that acts in a discontinuous manner and produces ~150 nt Okazaki fragments in each of its synthesis cycles. Together with other proteins involved in replication, the replicative helicase and the polymerases constitute the replisome. As replication progresses the replisome must overcome many types of template barriers, such as proteins bound to the DNA, RNA-DNA hybrids, and other forms of stress. To cope with these hurdles, the replisome can undergo post-translational modifications, and can recruit additional proteins to help clear template barriers. Finally, replication terminates when two neighboring replication forks meet. The end goal of replication is to have two identical copies of the genome in the cell that can be passed on to the two daughter cells.

So far eukaryotic replication is best understood in budding yeast for multiple reasons. For example, it permits precise genetic engineering to study the effects of mutation of different replication factors. In addition, many techniques are more feasible in budding yeast as it

contains a much smaller genome compared to higher eukaryotes and can mitotically grow as both haploid and diploid cells. Also importantly, the mechanisms of DNA replication are highly conserved among eukaryotes. Some of the earliest studies found important similarities between eukaryotic DNA replication in all organisms, including the fact that replication initiation is internal to yeast chromosomes and many chromosomes contain more than one replication unit (Newlon et al. 1974). Techniques have also been developed in budding yeast that make it a very desirable system to use, for example large-scale genetic screens have uncovered many replication proteins and interactions (Dumas et al. 1982). At the same time, the biochemical purification of proteins from yeast have allowed for the reconstitution of a minimal *in vitro* replication system (Seki and Diffley 2000; Remus et al. 2009; Yeeles et al. 2015). The budding yeast genome was the first eukaryotic genome to be sequenced, and this sequence has been used to determine coordination of origin firing as well as sites of initiation and termination of all the origins of replication across the genome (Raghuraman et al. 2001). To summarize, given most replication factors are conserved from yeast to higher eukaryotes, budding yeast offers a valuable system for examining multi-faceted regulation of replication, as its origins are well-mapped, and it permits genetic and biochemical examination of both positive and negative regulation (Bell and Labib 2016). The following discussion will be specifically focused on the mechanisms of DNA replication in budding yeast.

1.2 ORC loading of the MCM replicative helicase onto dsDNA

The mechanism by which ORC binds to the DNA, recruits Cdc6 and then loads the MCM replicative helicase has been a topic of much interest for the past several decades. The integration of genetic, biochemical, and structural studies has significantly advance our understanding of multiple steps by which the MCM double hexamer is loaded around double stranded DNA in G1 phase and activated during S phase.

1.2.1 The Autonomously Replicating Sequence

In budding yeast, unlike in higher eukaryotes, ORC binds to a specific DNA sequence known as the autonomously replicating sequence (ARS) which is the minimal DNA consensus sequence necessary to begin DNA replication. This sequence contains four regions, A, B1, B2, and B3, which were named based on their effect on plasmid stability (Clyne and Kelly 1997) (**Figure 1.2A**). The A-domain is highly conserved and necessary for origin function, while the B elements increase origin function but are not essential (Deshpande and Newlon 1992; Theis and Newlon 1994; Theis and Newlon 1997). However, this core 17bp stretch, known as the ARS consensus sequence (ACS), alone is not sufficient for replication initiation as there are thousands of sequences that match this in the genome but only about 400 sites function as origins of replication, work is ongoing to understand exactly how these origins are chosen (Breier et al. 2004; Shah and Krishnamachari 2012).

1.2.2 ORC-DNA interaction

There have been several structures of the budding yeast ORC complex bound to the DNA that help to explain the mechanism by which ORC can recruit the MCM, a hetero-hexamer made of up Mcm2-7. The ORC complex is a hetero-hexamer consisting of Orc1-6 and is stably associated with chromatin throughout the cell cycle (Gibson et al., 2006). Orc1-5 share a similar structure and contain an AAA+ ATPase domain and winged helix domains (WHD) for DNA binding (**Figure 1.1**) (Bleichert et al. 2017). The six subunits are arranged as

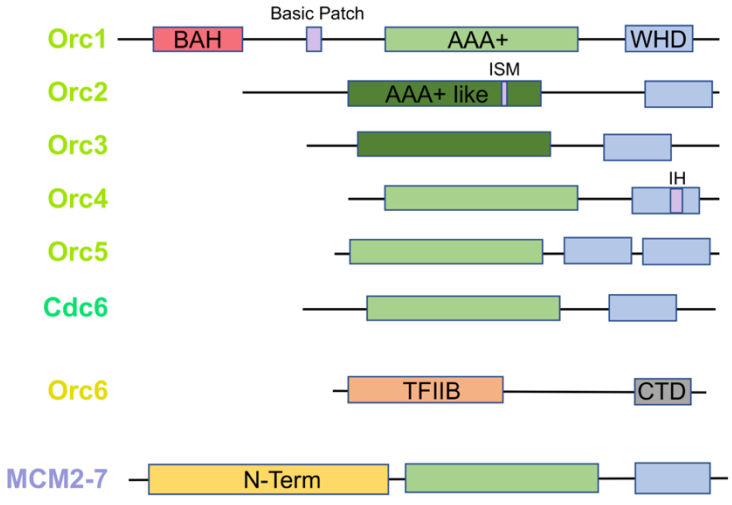


Figure 1.1 ORC, Cdc6, and MCM protein domains.

Orc1:Orc4:Orc5:Orc3:Orc2:Orc6 (Sun et al., 2012). The structure of the yeast ORC bound to a 72-base pair origin DNA sequence that contains the ACS and the B1 element demonstrates how the complex is organized around the DNA and how DNA bending by ORC allows for interaction with the B1 element (**Figure 1.2B**) (Li et al. 2018). Orc1-5 oligomerize using inter-subunit interactions between their AAA+ domains and between the WHDs and the AAA+ domains to encircle the ACS DNA (Li et al. 2018). Orc6 is on the periphery, is not involved with ACS recognition, but interacts with Orc2, Orc3 and Orc5 and the B1 element (Li et al. 2018).

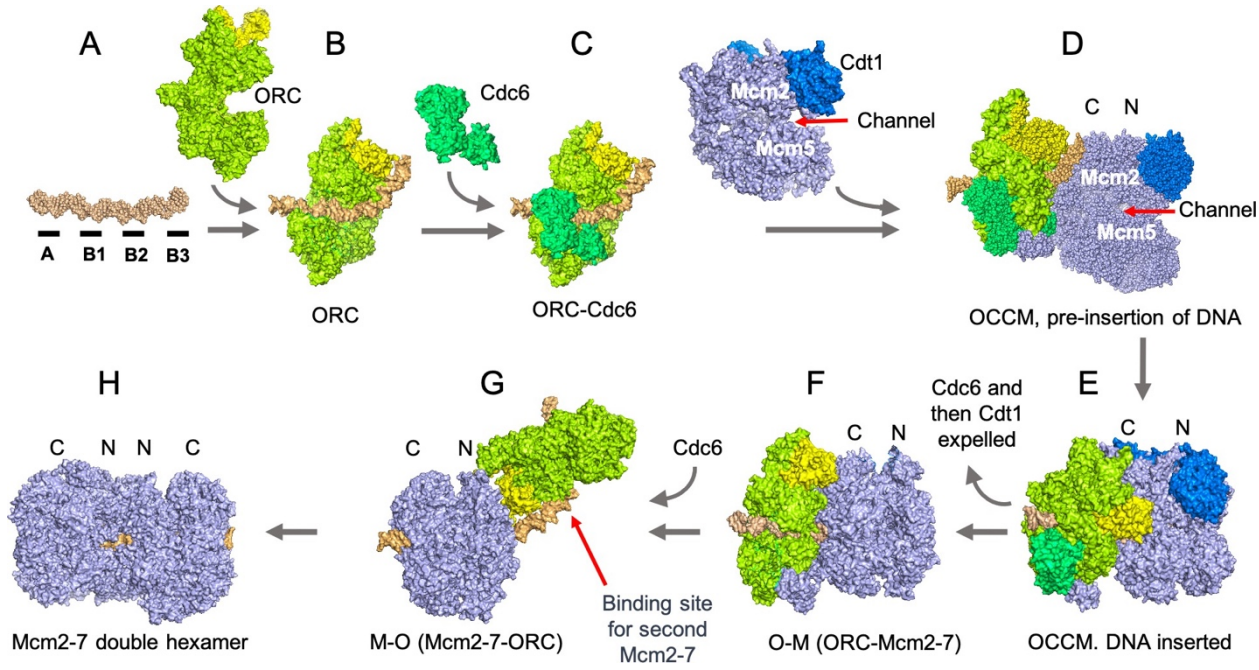


Figure 1.2 Seven-steps of pre-replicative complex assembly adapted from (Stillman 2022).

A) The ARS sequence that is bound by ORC in budding yeast B) ORC (lime green and yellow) bound to the A and B1 elements. C) Cdc6 (green) binding to the DNA bound ORC D) ORC-Ccdc6-Cdt1(blue)-MCM (purple)(OCCM) without DNA E) OCCM, with DNA F) ORC-MMCM G) MCM-ORC ready for another MCM recruitment H) MCM double hexamer

Three motifs from Orc1, Orc2, and Orc4 are important for DNA binding and specific recognition of the ACS sequence (Li et al. 2018; Lee et al. 2021). These critical motifs include a basic patch on Orc1, which is located between the bromo-adjacent homology domain (BAH) domain and the AAA+ domain and interacts with the minor groove, the Orc2- initiation specific motif (ISM) which places a tryptophan into the minor groove, and the Orc4 insertion helix, which

contains mostly hydrophobic residues with four aromatic side chains that inserts into the major groove and is involved in sequence specific interaction with the ACS (**Figure 1.1, purple boxes**) (Li et al. 2018; Lee et al. 2021).

However, the presence of an ARS is not the sole requirement for ORC-origin binding. The surrounding chromatin environment is also important in this process. Proper nucleosome positioning around the ARS is important for origin selection and function (Eaton et al. 2010). A recent study has shown that ORC preferentially binds to nucleosomes, over non-nucleosomal DNA and that this is important for MCM loading, potentially providing a link between ORC binding in yeast and higher eukaryotes (Li et al. 2022). The BAH domain of Orc1 has been shown to promote stable ORC association with DNA without affecting the integrity of the complex (Muller et al. 2010). The Orc1 BAH domain can interact with histones and influences ORC association with most yeast origins (De Ioannes et al. 2019).

1.2.3 MCM loading

In budding yeast ORC is bound to the DNA throughout the cell cycle but can only recruit the MCM helicase with the cooperation of Cdc6 (Liang et al. 1995). After anaphase, Cdc6 binding to ORC completes the ring of Orc1-5 and changes its conformation, making it ready for interaction with MCM associated with its co-factor Cdt1 (Liang et al. 1995) (**Figure 1.2C**). Cdc6 is also a AAA+ protein, which has extensive sequence similarity with Orc1 (**Figure 1.1**) (Perkins and Diffley 1998; Duncker et al. 2009). The structure of this complex (ORC-Cdc6-DNA) revealed that Cdc6 itself contributes to DNA binding through its WHD and AAA+ domains (**Figure 1.1**) (Sun et al. 2012). Cdc6 rearranges a short alpha-helix in the Orc1 AAA+ domain and the Orc2 WHD domain, leading to the activation of the Cdc6 ATPase and the formation of the three sites for the recruitment of MCM (Feng et al. 2021). During the assembly of the ORC-Cdc6-Cdt1-MCM (OCCM) complex, these domains are stabilized by the WHDs of MCM2-7, this

interaction is critical for the formation of this complex (**Figure 1.1** and **1.2D**) (Yuan et al. 2017; Feng et al. 2021).

The DNA enters the helicase through the Mcm2-5 interface, and the final ring closure requires ATP hydrolysis by Orc1 and Cdc6 (**Figure 1.2E**) (Yuan et al. 2017). ATP hydrolysis by MCM is required for Cdt1 release (Coster et al. 2014). After the first MCM helicase is loaded, the ORC complex flips to form an inverted Mcm2-7-ORC (MO) complex to allow a second helicase to be recruited (**Figure 1.2G**) (Miller et al. 2019; Gupta et al. 2021). This MO structure formation relies on the B2 element to serve as a second binding site for ORC. The distances between the ACS and the B2 element at different origins may play a role in determining if re-replication can be prevented through phosphorylation of ORC, as will be discussed in section 1.3.5 (Miller et al. 2019). The final goal of this process is for two MCMs to be loaded around the double stranded DNA in a head-to-head orientation (**Figure 1.2H**). To ensure that initiation does not occur, prematurely, MCM that is loaded onto the DNA is SUMOylated in G1 phase (**Figure 1.3, step 3**) (Wei & Zhao, 2016)

1.2.4 MCM activation

Loaded MCM is licensed at the start of S phase through a series of post-translational and protein association steps. This MCM activation process begins with the docking of DDK (Dbf4 dependent kinase) composed of two subunits, Dbf4 and Cdc7. Dbf4 contacts the N-terminal domains of Mcm2, Mcm4, and Mcm6 so they can be phosphorylated by Cdc7 (**Figure 1.3,** step 4) (Sheu and Stillman 2006; Sheu and Stillman 2010; Cheng et al. 2022). Dbf4 helps to limit the Cdc7 kinase activity to S phase via transcriptional and post-translational means (Forsburg 2004). Cdc7 phosphorylation of Mcm4, and Mcm6 results in helicase activation,

through the relief of an inhibitory function of a domain on Mcm4 and potentially an alteration in in the overall structure of the MCM double hexamer as a mutation in Mcm5 can also bypass the need for DDK (Hoang et al. 2007; Cheng et al. 2022). DDK-mediated MCM phosphorylation also leads to the recruitment of scaffold proteins Sld3, Sld7, and the MCM cofactor Cdc45 (**Figure 1.3**, step 5) (Aparicio et al. 1997; Labib 2010).

A second kinase, CDK (Cyclin dependent kinase), is activated by Clb5 or the shorter lived Clb6, to phosphorylate two additional scaffolding proteins Sld2 and Sld3 (**Figure 1.3,** step 6) (Donaldson et al. 1998; Jackson et al. 2006; Zegerman and Diffley 2007). Phosphorylation Sld2 permits the formation of the pre-loading complex, consisting of Sld2, Dpb11, Polymerase epsilon (Pol ε), and GINS (go-ichi-ni-san), itself a complex composed of Sld5 and Psf1-3 that is a key cofactor for MCM (Takayama et al. 2003). Phosphorylated Sld3 can then interact with Dbp11 to allow the association of the pre-loading complex with the loaded MCM. Finally the recruitment of other replication factors, such as Mcm10, converts the pre-initiation complex into two active replication forks that can move in opposite directions along with the replisome to duplicate the parental DNA (**Figure 1.3,** step 7) (Thu and Bielinsky 2013).

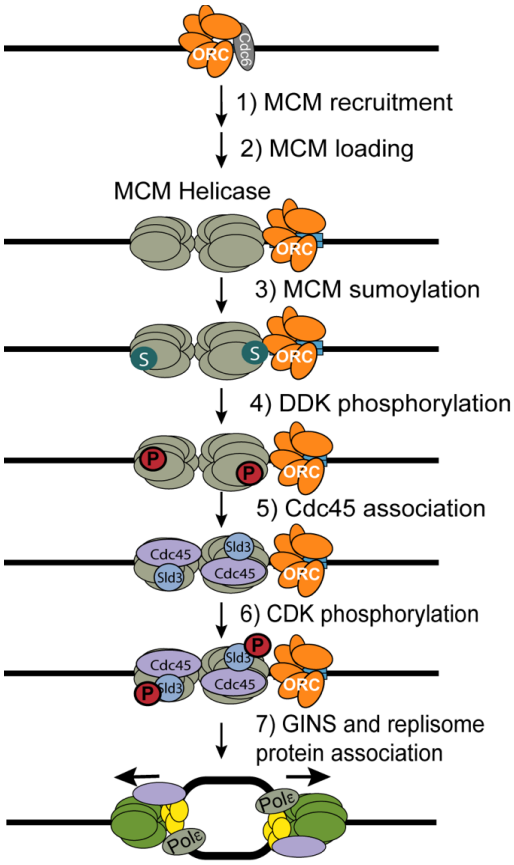


Figure 1.3 The main Steps involved in MCM licensing in S phase.

Steps 1 and 2: MCM is recruited and loaded onto the DNA by ORC. Step 3: MCM is sumoylated in G1 phase to prevent premature replication initiation. Step 4: DDK is upregulated in the transition to S phase and phosphorylates MCM Step 5: Cdc45 associates with p-MCM. Step 6: CDK phosphorylates Sld2 and Sld3 Step 7: GINS and Pol ε association

1.3 Genome-wide view of origin firing

The above sections detailed how an individual origin is activated to initiate replication. Complete replication of the entire genome in a timely fashion requires replication initiation at approximately 400 sites across the yeast genome. The firing of this many origins occur in a well-defined replication timing program in which subsets of origins fire in early, mid, and late S phase (Raghuraman et al. 2001; Yoshida et al. 2013). Origins that fire immediately upon S phase entry are called early, while those activated closer to the end of S phase are called late (Raghuraman et al. 2001). The importance of the staggered activation of origins are multifold: it allows cells to adjust the rate of DNA synthesis to the availability of limiting resources, such as dNTPs and histones, and helps cells to respond to stresses through S phase checkpoints (Friedel et al. 2009). In fact, late origins can also be operationally defined as those that are repressed by the DNA replication checkpoint when cells are exposed to a genotoxin or upon dNTP depletion (Santocanale and Diffley 1998; Crabbe et al. 2010; Zegerman and Diffley 2010; Poli et al. 2012).

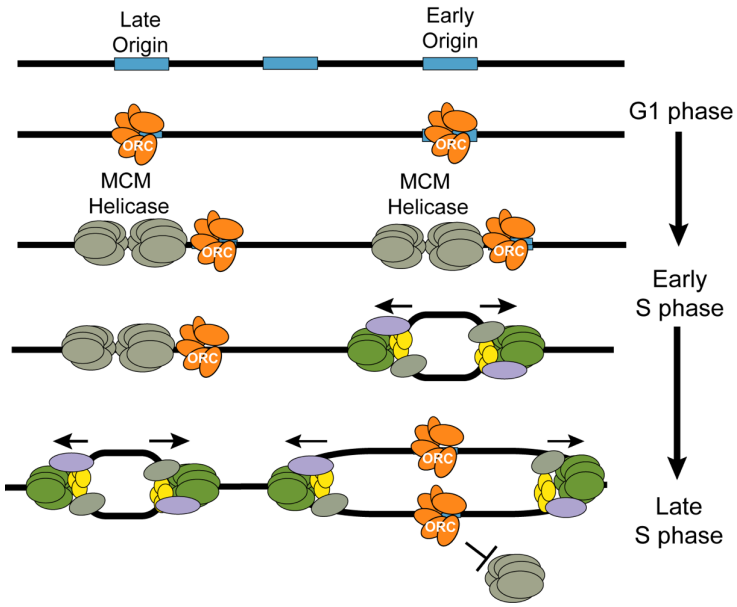


Figure 1.4 The Replication firing program.

Replication origins fire throughout S phase in a replication origin firing program. MCMs are loaded in G1 phase of the cell cycle. These origins are licensed throughout S phase and are activated at different times to complete DNA replication

The establishment of this origin timing program is dependent on multiple *cis*- and *trans*-acting elements. A major determent for an origin's firing timing is its affinity for limiting

replication factors. For example, early origins fire earlier because they have a higher affinity for low-abundance replication proteins, such as Sld3, Sld7, and Cdc45 (Tanaka et al. 2011). A variety of other factors can also influence replication timing. These include ORC-DNA binding efficiency, chromatin environment, and proteins such as Fkh1/2 and the Sir complexes that influence certain sets of origins. Additionally, as all origins compete for a limiting number of initiation factors, the activation of early origins can indirectly affect late origins. I note that no positive correlation between transcriptional activity and early origin firing in budding yeast has been found, unlike in multicellular organisms (Ferguson and Fangman 1992; Friedman et al. 1996; Raghuraman et al. 2001). Below, I review some of the main factors influencing origin usage and timing.

1.3.1 Origin activation: Origin efficiency

Origin firing timing relates to origin efficiency which can vary greatly between origins. Some origins are used in almost every cell cycle (>80%) while others are rarely active (<10%) (Friedman et al. 1997; Raghuraman et al. 2001). There is a general trend that efficient origins firing early, though there are some late firing origins that are very efficient (Ferguson et al. 1991; Das et al. 2015). Dormant origins that do not often fire in an unperturbed S phase can help to complete replication if cells are subject to DNA damage that can cause fork stalling (Santocanale et al. 1999; Vujcic et al. 1999; Tuduri et al. 2010). Interestingly, when the limited level of a specific set of initiation factors (Sld2, Sld3, Dpb11, Dbf4, Sld7, and Cdc45) is increased, the efficiency of dormant origins can be increased (Lynch et al. 2019). Single molecule studies have shown that origin usage can be extremely variable between individual cells, though the replication program is carefully maintained at the population level, suggesting that the replication program is a result of both deterministic and stochastic events (Czajkowsky et al. 2008; Rhind and Gilbert 2013; Yoshida et al. 2013).

Origin efficiency and timing may be related to the amount of MCM loaded at an origin. The so-called multiple-MCM model helps to explain how licensing events in G1 could affect the timing of origin firing in S phase (Das and Rhind 2016). It has been proposed that multiple MCMs could be loaded at individual origins, and that the number of MCMs loaded could influence firing efficiency and timing. This would suggest that early origins could have more MCMs loaded, and given MCM complexes are stochastically activated across the genome, origins with an increased number of MCMs are more likely to fire earlier (Das et al. 2015). The loading of multiple MCMs onto early origins could provide a memory of ORC activity in G1 and could stably establish the replication timing program (Das et al. 2015).

1.3.2 Origin Activation: ORC association with origins

The timing and affinity of ORC binding to DNA during G1 phase affects the timing of origin activation (Wu and Nurse 2009; Hoggard et al. 2013). Hoggard et al. revealed a link between ORC-origin binding mechanisms and the regulation of origin activation timing. They showed that some origins rely on strong ORC-DNA binding while others rely on ORC-'chromatin' interactions. Those origins that rely on chromatin interactions are enriched for early firing origins, while those that are 'DNA dependent' are enriched for later firing origins (Hoggard et al. 2013). As described above, the Orc1-BAH could help explain these 'chromatin' dependent origins. Indeed, the activity of a subset of origin activation shows Orc1-BAH dependency (Muller et al. 2010; Hoggard et al. 2013).

1.3.3 Origin Activation: Repetitive DNA sequences

Specific regions of the genome, particularly centromeres, telomeres, and the ribosomal DNA (rDNA) have particular requirements for replication. The replication of centromeric regions and approximately 20kb around them occurs earlier than the rest of the genome (Pohl et al. 2012). The replication of telomeres and approximately 35kb near the telomere occurs later on average

than the rest of the genome (Raghuraman et al. 2001). The early replication of the centromeric regions in budding yeast is due to the Ctf19 kinetochore complex, which can recruit DDK in telophase (**Figure 1.4**) (Natsume et al. 2013). This early firing is important to facilitate strong pericentromeric cohesion in S phase (Natsume et al. 2013). In contrast, replication origins located close to telomeres or within sub-telomeric repeat sequences are replicated late, partly due to the binding of the Rif1 protein that antagonizes DDK (Cosgrove et al. 2002).

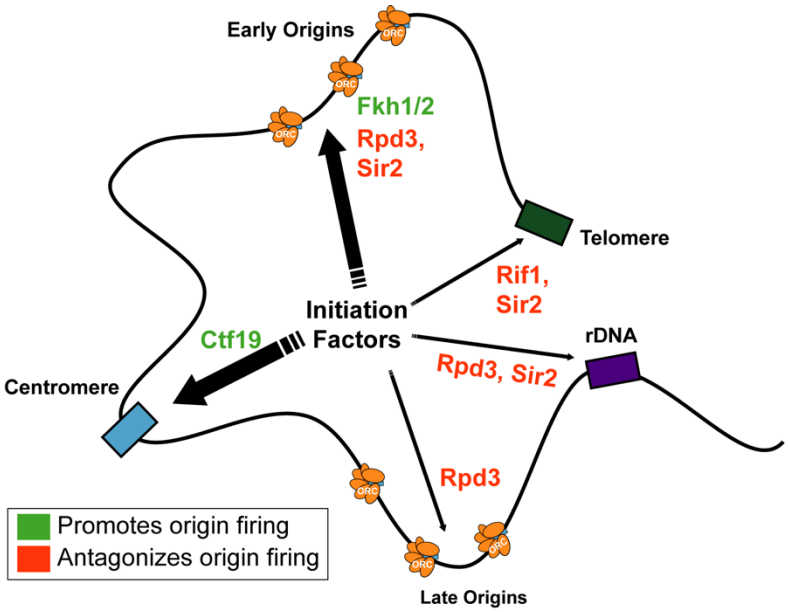


Figure 1.4 Genome-wide control of origin firing

As the source of ribosomal RNA, the rDNA, is a critical part of the genome whose maintenance is crucial for survival. It is highly repetitive, contains programmed replication fork stalls, and is highly transcribed thus can be a very unstable region of the genome (Kobayashi and Sasaki 2017). In budding yeast the rDNA are comprised of about 150 repeats on chromosome 12 that encode the 35S rRNA precursor, an rARS, a 5SrDNA, a non-coding promoter, and a replication fork blocking site (Kobayashi 2011). This number of repeats can be highly variable but generally accounts for approximately 10% of the entire genome (Kobayashi 2011). Because it is so large and as mentioned above replication factors are limiting, origin firing or copy number changes in this region can affect the replication program of the rest of the genome (Kwan et al. 2020). Replication initiation defects in this region lead to chromosome breakage and severely reduced copy number (Sanchez et al. 2017). In addition, rDNA can be a sensitive indicator for abnormalities of ORC binding (Ide et al., 2006). In fact, a mutation in

human Orc4, that results in in Meier Gorlin syndrome, leads to a prolonged S phase due to compromised initiation in the rDNA.

1.3.4 Origin Activation: Chromatin environment and other external control pathways

Local chromatin environment can influence origin firing timing via histone modification. Histone acetylation increases chromatin accessibility allowing for easier access for proteins involved in initiating replication (Vogelauer et al. 2002). Nucleosomes flanking early origins are acetylated on H4K16, which increases their mobility and facilitates MCM loading (Lengronne and Pasero 2020). The Sir2 histone de-acetylase that counters this modification has been shown to play a role in establishing the normal distribution of MCM complexes across budding yeast chromosomes (Hoggard et al. 2020; Lengronne and Pasero 2020). By deacetylating histones near early origins Sir2 reduces the advantage of early origins over late origins for pre-replication complex loading, ensuring a more even distribution of activated origins (**Figure 1.4**) (Hoggard et al. 2020). Loss of Sir2 lead to an increase in early replication of early origins and the delay of replication of the late-replicating euchromatin (Hoggard et al. 2020). At the same time Sir2 also reduces the activation of ribosomal DNA origins (Yoshida et al. 2014).

Rpd3 is another histone deacetylase that can dampen the activity of replication origins; however, it affects the firing of both early and late origins. Rpd3 regulates the time at which Cdc45 associates with origins, preventing this association at both early and late origins, though it affects late firing origins more strongly (**Figure 1.4**) (Vogelauer et al. 2002). There is also some evidence that Rpd3 additionally regulates origin firing at the rDNA locus to ensure enough replication factors to complete replication across the genome (Yoshida et al. 2014).

Examples of positive origin regulators uncovered in yeast thus far include the forkhead family transcription factors, Fkh1 and Fkh2, which bind adjacent to a subset of origins to promote ORC-origin binding and consequently origin licensing as well as DDK recruitment to

origins and consequently origin activation (**Figure 1.4**) (Lei et al. 1997; Knott et al. 2012b; Fang et al. 2017; Hoggard et al. 2021).

Finally, there are also some regions of the genome that have a physiological requirement for their replication timing. For example, genes that encode histones are among the earliest replicating loci in all species to allow for enough histone production during DNA replication (Muller and Nieduszynski 2017).

1.3.5 Preventing Re-Replication

As ORC binds to the DNA throughout the cell cycle in yeast, cells need to prevent MCM re-loading onto already fired origins to prevent re-initiation and re-replication, which can lead to genomic instability. The main mechanisms for preventing re-replication rely on the CDK kinase. These mechanisms work together to ensure that replication occurs once and only once per cell cycle (Green et al. 2006). *First*, CDK phosphorylates Orc2 and Orc6 to inhibit MCM loading, potentially through preventing MO formation required to recruit a second MCM dependent on the distance between the ACS and B1 element (Miller et al. 2019). *Second*, a hydrophobic patch of Clb5 in CDK binds to an “RXL” cyclin binding motif in Orc6 to sterically reduce the recruitment of Cdt1-MCM to ORC (Chen and Bell 2011). Mutation of the RXL motif prevents the N-terminal binding of Cdt1 to Orc6 and shows the larger increase in re-replication compared to the ORC phosphorylation mutants (Chen and Bell 2011). *Third*, CDK phosphorylation of MCM promotes its export from the nucleus to prevent it from being loaded during S phase (Nguyen et al. 2000; Early et al. 2004). *Fourth*, CDK phosphorylation of Cdc6 leads to its ubiquitin mediated proteolysis via SCF^{CDC4} (Early et al. 2004). Mutations of all or some these mechansims additively lead to re-replicaiton, however not every liscenced origin re-replicates when all these are mutated suggesting there many be further mechansims that prevent reinitiation (Nguyen et al. 2001; Green et al. 2006; Tanny et al. 2006).

1.4 Replication Elongation

Up until this point we have focused on MCM activation or so-called origin firing, but we will now move to the synthesis of the nascent strand. Once a replication origin has fired the helicase unwinds parental DNA and polymerization of the nascent strands can begin. This step of replication is known as elongation. During elongation three DNA polymerases add DNA nucleotides to the 3' end of the newly synthesized strand. DNA polymerases only synthesize in this direction, thus synthesis proceeds continuously on the leading strand and discontinuously on the lagging strand (Ogawa and Okazaki 1980).

1.4.1 Overview of three replicative DNA Polymerases

The three DNA polymerases involved in the bulk of DNA synthesis are polymerase α, polymerase δ, and polymerase ε (Yuan et al. 2020). Polα-primase, which has the lowest fidelity among the three, synthesizes short DNA-DNA primers for extension by Pol δ or Pol ε (Kunkel et al. 1991). Pol δ synthesizes the lagging strand DNA, in the form of Okazaki fragments, while Pol ε replicates the leading strand (Pursell et al. 2007; Nick McElhinny et al. 2008).

In this replication scheme, there is significant demand for Pol ε to be processive and to overcome various forms of template barriers (Shcherbakova et al. 2003). Pol ε is composed of four subunits: Pol2, which contains both a polymerase and exonuclease domain, and three accessory subunits Dbp2, Dbp3, and Dpb4. In wild-type cells, sister replication forks have been seen to progress at similar rates suggests that Pol ε is capable of efficient DNA synthesis despite template impediments (Claussin et al. 2022). This capacity stems from both its intrinsic attributes and the assistance of extrinsic factors. For example, multiple replication-promoting factors can remove template barriers enabling Pol ε's resumption of synthesis (Hizume and Araki 2019). In addition, unlike Pol δ, the catalytic subunit of Pol ε contains a structural domain

that stabilizes Pol ε-template association (Langston et al. 2014; Goswami et al. 2018). Pol ε contains two flexibly tethered lobes, the non-catalytic lobe of which is anchored to the motor of the helicase, whereas the polymerization domain extends toward the side of the helicase (Zhou et al. 2017). The eukaryotic polymerases are well conserved in their architecture and sequence, especially their catalytic domains.

1.4.2 Proofreading by Pol ε and Pol δ

Both Pol δ and Pol ε contain not only polymerase (POL) domains but also exonuclease (EXO) domains for proofreading wherein the enzyme can remove terminal mis-incorporated nucleotides from nascent DNA. The EXO domain of Pol δ is also involved in Okazaki fragment maturation and mismatch repair (Reha-Krantz 2010). In Pol ε, the POL and EXO domains are in the N-terminal half of its catalytic subunit Pol2, whose C-terminal half contains inactive polymerase/exonuclease domains that are structurally important (Hogg et al. 2014; Jain et al. 2014). The proofreading activity of these B-family polymerases is triggered by the presence of a base-pair mismatch between the template and newly synthesized DNA at the primer-template junction (Brutlag and Kornberg 1972; Khare and Eckert 2001). In these situations the 3' end of the primer is melted, and moved to the spatially separate proofreading domain where one or more nucleotides are exonucleolytically degraded (Bernad et al. 1989). The primer end can then be returned to the polymerase domain where DNA synthesis can continue (Capson et al. 1992). It is crucial that the exonuclease activity is only used when necessary, otherwise the processivity of the enzyme could be impeded (Spacciapoli and Nossal 1994a; Spacciapoli and Nossal 1994b). DNA synthesis is therefore a careful

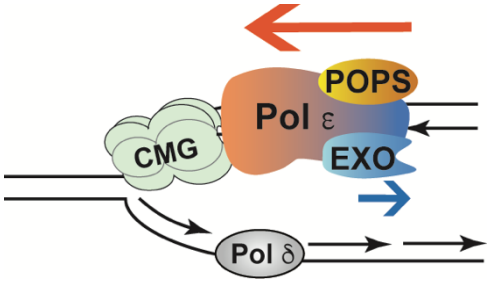


Figure 1.5 Replication fork overview.

The CMG helicase unwinds the double strand. Pol ε is the leading strand polymerase while pol δ synthesizes the lagging strand.

balance between polymerization and proofreading activity to ensure complete and careful replication of the genome.

1.4.3 The POPS domain of polymerase ε

The N-terminal of POL region of Pol2 contains several highly conserved domains that are absent in other DNA polymerases. One is the P-domain that contributes to Pol ε processivity by aiding Pol2-DNA binding (Hogg et al., 2014). A previous lab member has examined another domain, called the POPS (Pol2 family-specific catalytic core peripheral subdomain) domain. He and coworkers have shown that POPS mutation induces large genomic rearrangements, without affecting the mutation rate. In addition, POPs domain mutations in the human Pol2 counterpart (POLE) have been found in cancer patients, five of which occur with moderate to high frequency (Cerami et al. 2012; Gao et al. 2013). Modeling three of these five cancer mutations in the Pol2 POPS suggests that POPS affected both replication initiation via regulating pre-LC formation and replication elongation (Meng et al. 2020).

Interestingly, Pol2 is sumoylated on a single lysine in POPS to aid in replication fork progression (Meng et al. 2019). The purified Pol ε complex without a functional POPS domain reduced DNA synthesis *in vitro,* which supports a role for this domain in leading strand synthesis, however a more in depth understanding of this role has yet to be uncovered (Meng et al. 2020).

1.5 Sumoylation in DNA replication

Posttranslational modifications play key roles in regulating DNA replication via their ability to change various properties of substrates quickly and reversibly such as their activity, localization, stability, and interaction with other cellular molecules. These modifications are also

widely used to adjust cellular functions or in response to endogenous or exogenous stress. Modification by the small ubiquitin-like modifier (SUMO) plays important roles in regulating eukaryotic DNA replication and repair and reviews of this process and its main effects in yeast replication are described below (Sarangi and Zhao 2015).

1.5.1 The SUMO modification cycle and enzymes

SUMO is a highly conserved member of the ubiquitin family of protein modifiers. It is an approximately 100 amino acid protein that forms a ubiquitin fold despite only having 18% sequence identity with ubiquitin and different surface charge distribution (Bayer et al. 1998). The process by which SUMO is attached to the substrate is similar to that of ubiquitin (**Figure 1.4**). *First*, the single SUMO E1 enzyme carries out an ATP-dependent activation of the SUMO C-terminus by conjugating SUMO to itself. *Second*, this E1 transfers the active SUMO to the single SUMO E2 (Ubc9). *Third*, SUMO is transferred from Ubc9 to the substrate so that SUMO is covalently linked to its substrate through an isopeptide bond between the C-terminal carboxyl group of SUMO and the ε-amino group of a lysine in the substrate (Johnson 2004). This last step of the process requires the help of E3s. E3s enhance the specificity of sumoylation by interacting with the substrate and can lead to sumoylation at specific genomic loci (**Figure 1.5**).

Sumoylation is a labile process and desumoylation is a necessary part of the cycle that removes SUMO from substrates. In yeast, desumoylation is carried out by two Ulp family proteins that specifically cleave the C-terminal of SUMO (**Figure 1.5**). Ulp1 is additionally required to generate mature SUMO from the precursor. In yeast there is only one SUMO protein, while vertebrates contain three and plants contain many (Johnson 2004).

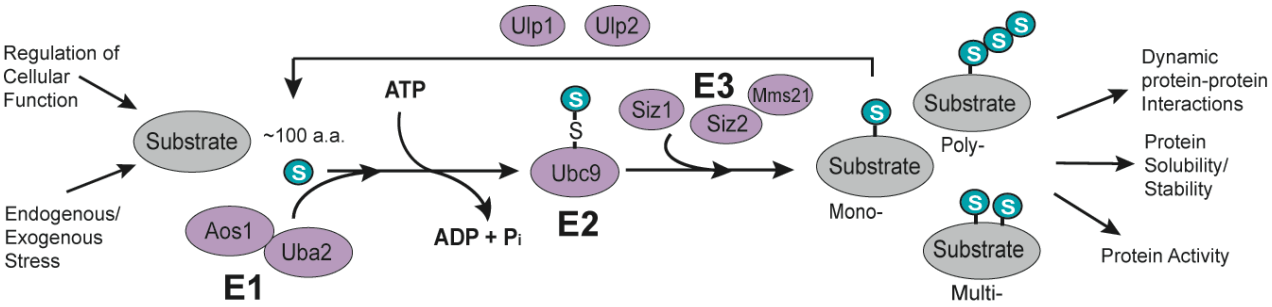


Figure 1.5 Overview of the sumoylation cycle in budding yeast.
SUMO is added to substrates via an E1, E2, and E3 mechanism. In budding yeast the SUMO E2 is Ubc9. Ulp proteins are responsible for desumoylation. Proteins can be mono-, poly-, or multi-sumoylated. Sumoylation can have a variety of consequences for target proteins.

1.5.2 SUMO site selection and additional methods to alter sumoylation levels

Several recent SUMO proteomic studies have estimated that 50-75% of sumoylation occurs on a consensus site (Xu et al. 2008). The SUMO consensus site consists of ψKXE where ψ is a large hydrophobic amino acid (I, L, V) and X is any residue. Sumoylation can be prevented by mutating the modified lysine residue to arginine, which is also a polar positively charged amino acid. Sumoylation can also be increased using a tag termed SuON. SuON is a high-affinity domain from a region of Ulp1 (418-621aa) wherein the catalytic cysteine 580 was mutated to serine to eliminate the desumoylation function. A control tag, called SuCtrl, has the same sequence as SuON except containing an additional point mutation (F474A) that prevents SUMO binding (Almedawar et al. 2012; Wei and Zhao 2016a). Like other post-translational modifications SUMO can stabilize proteins by protecting them from degradation, changing their cellular localization - targeting them to the nucleus for example - altering protein-protein interaction or inducing DNA binding.

1.5.3 Sumoylation in DNA replication

Sumoylation has been shown to regulate many DNA transactions including replication. Previous lab members have shown that many proteins are sumoylated during DNA replication and in response to DNA damage. In budding yeast multiple ORC subunits have been shown to be

sumoylated in response to DNA damage (Cremona et al. 2012). This work will show that this modification also occurs in untreated cells and will address the biological functions of this modification. In human cells Orc2 sumoylation has been shown to help prevent re-replication, polyploidy, and DNA damage at centromeric chromatin through the recruitment of a histone deacetylase to the centromere (Huang et al. 2016). Although the precise mechanisms may differ between yeast and humans, it is possible sumoylation of ORC in both these organisms plays a role in regulating replication initiation.

MCM has been shown to be sumoylated under normal growth conditions in a distinct spatial and temporal pattern relative to DNA replication to prevent premature MCM activation. After Cdc6-mediated MCM loading at replication origins, a fraction of MCM subunits are sumoylated. This sumoylation occurs before DDK-mediated Mcm4 phosphorylation and CMG formation. This modification can aid in the recruitment of the PP1 phosphatase, to counteract Mcm4 phosphorylation and preventing premature CMG formation (Wei and Zhao 2016a). MCM has also been seen to be sumoylated in higher eukaryotes, in which it may also provide a regulatory mechanism to restrain origin firing (Schimmel et al. 2014).

It has also been shown that the DDK kinase that helps to initiate replication is sumoylated on chromatin to be degraded. The SUMO and ubiquitin pathways can be interrelated and poly-SUMO chains can be recognized by SUMO-targeted ubiquitin ligases (STUbLs) that can mediate protein ubiquitylation (Sriramachandran and Dohmen 2014). It was shown that DDK is sumoylated, and that the SUMO protease Ulp2 protects DDK from chain formation early in S phase but allows for DDK degradation in late M phase, thus timing protein turnover (Psakhye et al. 2019).

Several proteins involved in replication elongation have also been shown to be sumoylated including the Pol2 subunit of Pol ε. Sumoylation of these proteins is often in response to DNA damage or fork stalling. Sumoylation of Pol2 for example occurs specifically in S phase and is increased under conditions of replication fork blockade (Meng et al. 2019). In

this case Pol2 sumoylation is important for proper replication progression and Pol2-mediated DNA synthesis (Meng et al. 2019). Overall, sumoylation is emerging as an important post-translational modification that can regulate DNA transaction processes, though the extent and function of sumoylation is still being explored.

1.6 Dissertation Research Aims

In summary, although there is a wealth of knowledge of replication and its regulation, there are still large gaps in our understanding of some important choices that are made during the DNA replication program. Open questions in replication initiation include why ORCs loaded on different origins have different abilities to recruit MCM, what causes differences in origin efficiency, and if there are additional regulators of origin firing timing. Although we have some understanding of how post-translation modifications regulate these processes, there is still more to be determined. One aim of this thesis is to enhance our understanding of how sumoylation of the ORC complex can be used as a means to regulate replication initiation. As described in **Chapter 2,** I found that the sumoylation of ORC is crucial to achieve the optimal origin firing program and consequently genome stability. I found that either hyper- or hypo- sumoylation led to changes in early origin firing in opposite directions, but both led to an increase in genomic instability.

In replication elongation there is a careful balance between polymerization and exonuclease function, however the mechanisms of this balance are not completely understood. In **Chapter 3,** I examine the effect of the POPS domain mutation on the symmetry of replicon synthesis. My work contributed to our conclusions that mutation of the POPS domain can increase sister replication fork asymmetry which can be rescued by an additional mutation in the Pol2 exonuclease domain. This suggests that the POPS domain counterbalances the exonuclease mediated polymerase backtracking to favor forward polymerase movement though

the precise mechanism is unclear. Taken together these findings uncover some important regulatory mechanisms in replication initiation and elongation respectively.

CHAPTER 2 : THE ORIGIN RECOGNITION COMPLEX SUMOYLATION STATUS FINE TUNES ORIGIN ACTIVITY

2.1 Introduction

Faithful duplication of the genome is critical for the development and health of organisms. The initiation of this process is carefully orchestrated and relies on the conserved origin recognition complex (ORC) (Bell and Stillman 1992). Composed of the Orc1-6 subunits, ORC marks replication initiation sites (origins) by binding directly to DNA and recruiting other replication initiation factors. To date, ORC is best examined in budding yeast, wherein it defines ~400 specific origins in G1-phase and collaborates with the Cdc6 protein to recruit and load the MCM replicative helicase complex onto origin DNA, a process referred to as origin licensing (Bell and Labib 2016). In the subsequent S-phase, origins are activated by several replication initiation factors including the DDK kinase so that MCM can direct replisome assembly and unwind the parental DNA duplex to permit nascent strand synthesis (Bell and Labib 2016).

The global regulation of origin function affects genome stability and cell-fate determination, thus is a vital aspect of genome maintenance (Smith et al. 2001; Watanabe and Maekawa 2010; Ryba et al. 2011; Debatisse et al. 2012; Pozo et al. 2018). The use or firing of individual origins is affected by both origin licensing and origin activation, and each of these distinct steps can be controlled at multiple levels. Budding yeast offers a valuable system for examining the complexities of the regulation of origin function, because its origins are mapped, and genetic, genomic, and biochemical tools are available to probe both positive and negative modes of regulation (Bell and Labib 2016). Examples of positive origin regulators uncovered in yeast thus far include the Fkh1 and Fkh2 proteins that bind adjacent to a subset of origins to promote ORC-origin binding and consequently origin licensing as well as DDK recruitment to

---

Chapter 3 is adapted from Regan-Mochrie G, Hoggard T, Bhagwat N, Lynch G, Hunter N, Remus D, Fox CA, Zhao X. 2022. Yeast ORC sumoylation status fine-tunes origin licensing. *Genes Dev.*
Reproduced with permission Cold Spring Harbor Laboratory Press

origins and consequently origin activation (Lei et al. 1997; Knott et al. 2012a; Fang et al. 2017; Hoggard et al. 2020). Countering positive regulation, the Sir2 and 3 proteins promote repressive chromatin at a subset of origins to reduce their licensing probability (Hiraga et al. 2014; Hoggard et al. 2018; Hoggard et al. 2020). In addition, replication initiation factors such as DDK and the MCM co-factor Cdc45 are limiting in cells, thus their affinity for specific origins positively correlates with origin firing timing and overall origin function (Mantiero et al. 2011; Tanaka and Araki 2011; Saner et al. 2013; Tanaka and Araki 2013). Generally, origins that fire in early S phase (early origins) tend to be more effective at competing for the limiting origin activation factors (Mantiero et al. 2011; Tanaka and Araki 2011). As such, altering early origin firing can indirectly alter the use of origins that fire later in S phase (Knott et al. 2012a; Hoggard et al. 2020). Collectively, studies thus far have revealed multi-layered regulation of origin function at the level of both licensing and activation; however, a full picture has yet to be achieved.

Multiple origin-binding factors are sumoylated in both yeast and humans, suggesting that this protein modification is relevant to the regulation of origin function (Golebiowski et al. 2009; Cremona et al. 2012). Sumoylation is a highly dynamic protein modification that targets a large array of proteins and influences substrate functions by diverse mechanisms (Zhao 2018). Understanding the roles of sumoylation on any substrate requires establishing approaches where its sumoylation status can be specifically altered. Here we focus on the ORC complex that is sumoylated in both yeast and human cells as revealed by proteomic screens under stress conditions (Golebiowski et al. 2009; Cremona et al. 2012). Currently the biological consequences of ORC sumoylation are poorly understood.

To address the above question, we established that multiple yeast ORC subunits are sumoylated during unperturbed growth, indicating that this modification is not confined to stress conditions. We then generated both hyper- and hypo-sumoylated ORC and examined the consequences to ORC function in chromosome replication. Genome-scale assessment of origin firing in an ORC hyper-sumoylation situation revealed selective repression of origin function,

particularly that of early origins, and reduced chromatin-associated MCM during G1 phase. Consistent with these *in vivo* data, MCM loading reactions performed with reconstituted proteins revealed that sumoylated ORC inhibited MCM recruitment to origin DNA. Further, determination of a specific endogenous sumoylation site on Orc2 using mass-spectrometry allowed the generation of a single-residue substitution mutant that led to reduced Orc2 sumoylation. This hypo-sumoylated ORC mutant had the opposite origin firing phenotype from that caused by hyper-sumoylated ORC, namely enhanced firing of a subset of early origins. Finally, though hyper- and hypo-sumoylated ORC showed opposing effects on origin function, both resulted in genomic instability and required other genome stability factors for optimal growth. These data provide evidence that appropriate ORC sumoylation levels are required to promote the normal spatiotemporal dynamics of origin function that in turn is important for genome stability.

2.2 Results

2.2.1 Multiple ORC subunits were sumoylated during normal growth

Our previous work identified ORC sumoylation under DNA damage conditions in yeast (Cremona et al. 2012). Here we tested whether ORC subunits were sumoylated during normal growth. We employed an established method for sumoylation detection, wherein sumoylated proteins are enriched on Ni-NTA resin due to binding of endogenously expressed 8His-tagged yeast SUMO (Smt3), referred to as Nickel pull-down or Ni-PD (Ulrich and Davies 2009). The procedure was carried out under denaturing conditions to limit desumoylation during extraction and contamination of other proteins through protein-protein interactions. Each endogenous ORC subunit was tagged at its C-terminus with a 6HA epitope that is lysine-free to avoid potential sumoylation of the tag itself. Viable cells were obtained for four out of the six ORC subunits. While HA-tagged Orc2, Orc4, and Orc5 supported normal growth, similarly tagged

Orc1 showed slower growth (**Figure 2.2a**). The sumoylation levels in all four HA-tagged strains were examined.

For each ORC subunit examined, Ni-PD eluate was evaluated by immunoblotting using an anti-HA antibody. Unmodified forms of HA-tagged ORC subunits that migrated at their expected molecular weight were detected at low levels due to non-specific histidine-mediated binding to Ni-NTA resin, as previously documented for other proteins (**Figure 2.1a** orange dots) (Cremona et al. 2012; Wei and Zhao 2016b; Dhingra et al. 2019). Importantly, modified forms unique to each ORC protein were detected as slower migrating bands (Fig. 1A, teal triangle). These modified forms exhibited the typical reduced migration in SDS-PAGE caused by sumoylation, as a single SUMO moiety causes a ~20 kD protein band upshift (Ulrich 2009). These sumoylated forms were further verified as they were only detected in the presence of 8His-SUMO, but not with untagged SUMO, whereas unmodified forms were detected in both conditions (**Figure 2.2b**).

We observed that sumoylation levels for Orc1, Orc2, and Orc4 were higher than for Orc5 (**Figure 2.1a**). In addition, mono-sumoylated and multi- or poly-sumoylated forms were seen for Orc1, Orc2, and Orc4, while Orc5 mainly exhibited a mono-sumoylated form (**Figure 2.1a**). Two distinct mono-sumoylated Orc4 forms were observed likely due to two different sites of mono-sumoylation leading to different migration rates, as seen for other sumoylated targets such as PCNA or Rfa1 (**Figure 2.1a**) (Papouli et al. 2005; Dhingra et al. 2019). Taken together, these data provided evidence that multiple ORC subunits were sumoylated during normal cell growth, and that each sumoylated ORC subunit had a distinct sumoylation status.

2.2.2 ORC2-SuON caused ORC hyper-sumoylation and cell lethality

Next, we assessed the consequences of altering ORC sumoylation levels. First, we generated an ORC hyper-sumoylation construct. The strategy used exploits a high affinity SUMO interaction domain to promote sumoylation of its fusion partner and possibly closely

associated subunits, presumably by increasing local SUMO concentrations (Almedawar et al. 2012). This tag, referred to as SuON, has been used to efficiently enhance the sumoylation of several proteins (Almedawar et al. 2012; Bonner et al. 2016; Wei and Zhao 2016b). The same tag containing a single point mutation at the key SUMO binding residue was used as a control and referred to as SuCtrl (Almedawar et al. 2012).

We chose to tag Orc2 as it tolerated C-terminal tags well (**Figure 2.2a**), and its sumoylation forms were easily detectable (**Figure 2.1a**). The SuON or SuCtrl tag sequence was inserted at the 3'-end of one copy of the *ORC2* gene in diploid cells, followed by sporulation to produce untagged and tagged versions of *ORC2*. *ORC2-SuCtrl* fusion supported wild-type growth, suggesting the tag was tolerated by Orc2 (**Figure 2.1b**, lower panels). In contrast, *ORC2-SuON* fusion-containing cells were not recovered (**Figure 2.1c**, lower panel), providing evidence that hyper-sumoylation of ORC caused cell lethality.

Given that SuON does not cause cell lethality when fused to other proteins, despite increasing sumoylation of its fusion partners (Almedawar et al. 2012; Bonner et al. 2016; Wei and Zhao 2016b), we reasoned that the *ORC2-SuON* lethality was due to hyper-sumoylation of Orc2 and/or other ORC subunits. Indeed, examination of protein extracts showed higher Orc2 sumoylation levels in diploid cells containing *ORC2-SuON* compared to *ORC2-SuCtrl* (**Figure 2.1d**). Global sumoylation levels were similar between these diploid cells, suggesting the specificity of *ORC2-SuON* (**Figure 2.2c**). Thus *ORC2-SuON* cell inviability was most likely caused by excessive ORC sumoylation.

2.2.3 Lethality caused by ORC hyper-sumoylation was rescued by reducing SUMO E2 function.

To further test the notion that *ORC2-SuON* cell lethality was due to increased ORC sumoylation, we repeated the SuON tagging of Orc2 in diploid cells heterozygous for a mutant allele of the SUMO E2 enzyme Ubc9 (*ubc9-10*) (Cremona et al. 2012; Wei and Zhao 2017). If excessive ORC sumoylation were causing lethality, then the *ubc9-10* mutant would be predicted

to reduce ORC sumoylation and rescue the cell lethality caused by *ORC2-SuON*. While Ubc9 is essential, *ubc9-10* supports normal growth at its permissive temperature of 24ºC with moderately reduced levels of global sumoylation (Cremona et al. 2012; Wei and Zhao 2017).

At 24ºC, *ubc9-10* cells exhibited reduced Orc2 sumoylation levels in cells containing *ORC2-SuON* (**Figure 2.1e**). Furthermore, haploid spore clones containing both *ORC2-SuON* and *ubc9-10* were viable, albeit slow growing, in contrast to the inviable *ORC2-SuON* spore clones (**Figure 2.1f**). These data provided additional evidence that the inviability of *ORC2-SuON* cells was due to increased sumoylation of ORC.

As *ubc9-10* allows for *ORC2-SuON* cell viability, haploid cells that contain either *ORC2-SuON* or *ORC2-SuCtrl* as the only copy of *ORC2* could be compared in a *ubc9-10* background. *ORC2-SuON* led to a higher level of sumoylated Orc2 even in a *ubc9-10* background as compared to *ORC2-SuCtrl* (**Figure 2.2d**). Thus, this system allowed us to further examine the effects of ORC hyper-sumoylation. Because the SuON tag can increase the sumoylation of other subunits within a complex when fused to a single subunit (Almedawar et al. 2012; Bonner et al. 2016; Wei and Zhao 2016b), we examined sumoylation of Orc1 and Orc4 in *ORC2-SuON ubc9-10* cells. *ORC2-SuON* enhanced Orc1 and Orc4 sumoylation compared to the *ORC2-SuCtrl* cells (**Figures 2.2e and 2.2f**). Thus, the *Orc2-SuON* fusion caused hyper-sumoylation of multiple ORC subunits. However, Orc2-SuON and Orc2-SuCtrl cells showed similar levels of sumoylation of the Mcm2 and Mcm3 subunits of the replicative helicase that binds to ORC at origins during origin licensing (**Figure 2.2g**). Thus, Orc2-SuON likely limited hyper-sumoylation only to ORC itself.

2.2.4 ORC hyper-sumoylation reduced the firing of a subset of origins

We exploited the viability of *ORC2-SuON ubc9-10* cells to examine the consequences of excessive ORC sumoylation on origin function. Specifically, whole genome sequencing of G1- and S-phase cells was used to generate high-resolution replication profiles for both *ORC2-*

SuON ubc9-10 and congenic *ORC2-SuCtrl ubc9-10* cells. In these experiments, DNA copy number changes (S/G1 ratios) were used to derive replication profiles for each chromosome (Hawkins et al. 2013). As shown in the example of chromosome VII, compared with *ORC2-SuCtrl* cells, cells containing *ORC2-SuON* exhibited an alteration of S-phase copy number at some origins (**Figure 2.3a and Figure 2.4**). These data provided evidence that the function of specific origins was sensitive to ORC hyper-sumoylation.

Next, a quantitative approach was used to identify the groups of origins affected by Orc2-SuON. To this end, each origin was assigned a Z-score represented by its S-to-G1 ratio of sequenced reads to reflect its replication behavior (Hoggard et al. 2021). A kernel density estimation was performed to generate a graph of origin density (y-axis) versus Z-scores (x-axis) (**Figure 2.3b**). This graph revealed the genome-scale effects of ORC hyper-sumoylation on origin firing. The *ORC2-SuCtrl* data generated a plot with a shoulder and two clearly distinguished peaks, corresponding to the known 'early', 'mid', and 'late' origin waves of replication (**Figure 2.3b**, grey trace). In contrast, the *ORC2-SuON* profile generated only a single peak centered between the mid-S and late-S origins (**Figure 2.3b**, teal trace). Thus, ORC hyper-sumoylation altered the normal pattern of temporal control of origin firing, with origins normally activated in distinct halves of S-phase (first-half, early; second-half, late) acting more similarly.

The graph in Figure 2.3b provided evidence that the replication of a significant number of early origins was relatively delayed and/or that of late origins was accelerated. To test this notion further, the Z-score ratio of *ORC2-SuON*/*ORC2-SuCtrl* was determined for each origin. A negative ratio (cut-off of < 0.8) indicated that an origin's firing was relatively reduced in *ORC2-SuON* cells whereas a positive value (cut-off of >1.2) suggested the opposite. By these criteria the firing of 35 origins was reduced (negative) and 43 origins enhanced (positive) by ORC hyper-sumoylation (**Figure 2.3c**). These origins were further parsed by their experimentally determined replication times and the data displayed in stacked histograms (**Figure 2.3c**)

(Yabuki et al. 2002). These analyses revealed that among the affected origins, ORC hyper-sumoylation preferentially reduced the firing of early origins relative to mid- and late- acting origins, and, conversely, enhanced that of late origins. This conclusion was consistent with the current model that reducing the ability of early origins to compete for limited replication factors enhances the availability of such factors to late origins. Thus, we posit that ORC hyper-sumoylation directly reduced the firing of a subset of early origins likely via dampening origin licensing and, in doing so, allowed for a relative increase in the firing of some late origins.

2.2.5 ORC hyper-sumoylation reduced MCM recruitment to origin DNA

ORC's established biochemical role at replication origins is to recruit and load the MCM complex in the origin licensing process, and therefore reduced MCM loading onto chromosomal DNA is a likely mechanism by which *ORC2-SuON* reduced origin function. To test this idea, the effects of ORC sumoylation were examined in a reconstituted MCM recruitment and loading system (Remus et al. 2009). In this system, purified yeast ORC was first loaded onto origin-containing DNA immobilized on paramagnetic beads (**Figure 2.5a**, step 1). Then one half of the DNA-bound ORC was subjected to a reconstituted sumoylation reaction, which employed purified sumoylation machinery and ATP (Zhao and Blobel 2005), while the other half served as the control reaction by omitting SUMO (**Figure 2.5a**, step 2). Robust ORC sumoylation was achieved only in the reactions that included SUMO, and sumoylation did not reduce ORC retention on DNA-beads (**Figure 2.5b and Figure 2.6a**, lane 2 and 6). We note that ORC subunits were sumoylated to varying degrees similar to observations *in vivo* under the conditions used here. The analyses indicate that Orc1-3 were abundantly sumoylated because only slower migrating sumoylated forms were detected, while Orc4-6 were sumoylated to a lesser degree because unmodified forms of these subunits were easily detected (**Figure 2.6b**).

The sumoylation machinery and ATP in both the reaction and the control were removed from the DNA-beads (**Figure 2.5a**, step 3). This step was followed by the addition of purified

Cdc6 and MCM–Cdt1 in the presence of either ATPγS or ATP to initiate MCM recruitment or to allow for a complete loading reaction, respectively (**Figure 2.5a**, step 4). ATP-hydrolysis dependent MCM loading resulted in the formation of a MCM double hexamer topologically bound to the DNA that was resistant to high-salt wash, whereas MCM that was only recruited to ORC-Cdc6-DNA but not loaded, as occurs in the presence of ATPγS, was washed off with high-salt buffer (**Figure 2.5a**, step 5) (Remus et al. 2009; Sun et al. 2013). As such, this experimental set up can assess both 'MCM loading' and 'MCM recruitment' (**Figure 2.5a**, step 6).

In the presence of ATPγS and a low salt wash, 'MCM recruitment' was efficient in the absence of ORC sumoylation, as seen previously (**Figure 2.5b and Figure 2.6a**, lane 1). However, MCM recruitment was greatly reduced upon ORC sumoylation (**Figure 2.5b and Figure 2.6a**, lane 2). As expected, no MCM was recovered on DNA in the presence of ATPγS after a high-salt wash regardless of ORC sumoylation (**Figure 2.5b and Figure 2.6a**, lane 3-4). In the presence of ATP and a high-salt wash, MCM loading was observed in the absence, but not the presence of ORC sumoylation (**Figure 2.5b and Figure 2.6a**; lane 7-8). Importantly, defective MCM recruitment associated with ORC sumoylation was reversed by removing sumoylation using the purified desumoylation enzyme Ulp1 prior to the addition of MCM (**Figure 2.5c and 2.6c**). Thus, the ORC-DNA complex was fully capable of recruiting and loading MCM once sumoylation was removed. Therefore, ORC hyper-sumoylation directly inhibited MCM loading *in vitro* by impeding MCM recruitment to the ORC-origin DNA complex.

To test whether ORC sumoylation also reduced MCM association with chromatin *in vivo*, yeast cells arrested in G1 were examined for chromatin bound MCM levels using a standard chromatin fractionation assay (Schepers and Diffley 2001). The Mcm2 signal from the chromatin fraction in *ORC2-SuON* cells was reduced approximately two-fold compared to *ORC2-SuCtrl* cells, while the level in the supernatant was concomitantly increased (**Figure 2.5d**). Taken together, the *in vivo* and *in vitro* data provided evidence that ORC hyper-sumoylation inhibits

MCM recruitment to origins. Accordingly, the alterations in replication origin firing observed in *ORC2-SuON* cells could be explained by reduced origin licensing efficiency.

2.2.6 ORC hyper-sumoylation compromised rDNA replication and stability

Thus far, we have focused on non-repetitive regions of the genome. It is established that replication and stability of the repetitive ribosomal DNA (rDNA) is particularly sensitive to defects in replication initiation factors including ORC mutants (Kwan et al. 2013; Salim et al. 2017; Sanchez et al. 2017). We therefore examined whether *ORC2-SuON* compromised the replication and stability of the rDNA. First, pulsed field gel electrophoresis (PFGE) was used to examine chromosomal replication efficiency. In this assay only fully replicated chromosomes can enter the gel. While replication completion of all chromosomes was compromised by *ORC2-SuON*, chromosome XII, which contains the rDNA repeats, showed the most striking defects (**Figure 2.7a**). Quantitative comparison of gel entry between the two longest chromosomes in yeast, chromosomes IV and XII, confirmed that the *ORC2-SuON* associated replication defects were more severe for chromosome XII (**Figure 2.7b**). Second, we used a rDNA marker-loss assay to assess rDNA stability (**Figure 2.7c**). *ORC2-SuON* increased rDNA marker loss by approximately 4-fold compared to *ORC2-SuCtrl* (**Figure 2.7c**). Together, these data provided evidence that ORC hyper-sumoylation caused defects in rDNA replication and stability.

2.2.7 Identification of the major sumoylation site on Orc2

Our examination of *ORC2-SuON* cells provided evidence that ORC hyper-sumoylation reduced the function of a subset of origins. To complement this analysis, a mutation in ORC that reduced ORC sumoylation was required. To this end, an endogenous sumoylation site on at least one ORC subunit first had to be identified. As biochemical mapping of sumoylation sites of endogenously expressed proteins is challenging due to the low level of sumoylated forms, we concentrated on mapping the sumoylation site(s) on the Orc2 subunit due to its relatively high

level of sumoylation. Using a two-step protocol to enrich endogenously sumoylated Orc2 forms, four Orc2 sumoylation sites were identified by mass spectrometry (**Figure 2.8a and Figure 2.9**). Three of the sites, K406, K434, and K419, were within the conserved AAA+ like domain of Orc2, while the fourth site, K592 was within the Orc2 winged-helix domain (WHD) (**Figure 2.8b**).

To assess which of these four sites were most relevant for Orc2 sumoylation, gene replacement was used to convert each lysine codon to an arginine codon, alone or in combination, at the endogenous Orc2 locus. Substitution of all four lysine residues with arginine reduced Orc2 sumoylation substantially (**Figure 2.8c**, lane 4). Notably, the single K406R substitution caused a similar reduction in sumoylation (**Figure 2.8c**, lane 5), whereas K592R alone or the combination of K419R and K434R mildly reduced Orc2 sumoylation (**Figure 2.8c**, lane 1-2). Therefore, K406 was the primary site responsible for Orc2 sumoylation *in vivo*. Importantly*, orc2-K406R* had no effect on Orc2 protein levels or Orc2 stability on chromatin, consistent with the *in vitro* data showing that ORC-DNA stability was unaffected by ORC sumoylation (**Figures 2.8d and 2.8e**). Thus, the *orc2-K406R* mutant protein reduced Orc2 sumoylation substantially but retained the most fundamental activity of ORC, binding to chromosomal DNA.

2.2.8 Orc2 hypo-sumoylation enhanced early origin firing

To assess the effects of the Orc2 sumoylation mutant on DNA replication, cell-cycle progression upon release from G1-arrest was measured by FACS analyses. Compared with Orc2-6HA*,* the mutant *orc2-K406R-6HA* cells exhibited a two-mode alteration in S phase progression (**Figure 2.10a**). At early timepoints (25 and 30 min), more *orc2-K406R* cells were in S-phase compared to *ORC2* WT cells, while at 50 and 60 minutes, when most *ORC2* wild-type cells had reached the 2C peak, a substantial fraction of *orc2-K406R* cells had not yet reached that point (**Figure 2.10a**). These data provided evidence that Orc2 hypo-sumoylation caused DNA replication defects

To compare the consequences of hypo- vs hyper-sumoylation of ORC on origin function, high-resolution replication profiles as described for *ORC2-SuON* above were generated for *ORC2-6HA,* and *orc2-K406R-6HA*. As exemplified by the replication profile scan of chromosome VII, *orc2-K406R* altered the firing of subsets of origins relative to *ORC2* (**Figure 2.10b and Figure 2.11**). *Z*-score analyses for origins with different replication times revealed enhancement of early origin firing by *orc2-K406R* (**Figure 2.10c**). This phenotype was generally opposite to that observed for *ORC2-SuON* as described above. To further challenge this idea on a genome-scale, a kernel density estimation was performed to generate graphs of the origin density (y-axis) versus the log2 of Z-score ratios (x-axis) for each *ORC2-SuON*/*ORC2-SuCtrl* (teal) and *orc2-K406R*/*ORC2* (purple) datasets (**Figure 2.10d**). This approach provided a direct comparison of the genome-scale origin firing of hyper- vs hypo-sumoylation of ORC, which showed the two had opposing effects on a significant number of origins (**Figure 2.10d**). Finally, as for *ORC2-SuON*/*ORC2-SuCtrl* analyses in Figure 2C, *orc2-K406R*/*ORC2* Z-score ratios were determined for each origin, affected origins divided into negatively or positively affected categories, and the origins were parsed again by their experimentally determined replication times (**Figure 2.10e**). These analyses revealed that among affected origins, ORC hypo-sumoylation caused by *orc2-K406R* preferentially enhanced the firing of early origins and reduced that of late origins, showing generally the opposite consequences on affected origins compared to ORC hyper-sumoylation (**Figure 2.3c**). Consistent with the dual-shifts of origin function observed in *orc2-K406R-6HA* cells, gross assessment of MCM chromosomal association did not reveal a clear difference when compared with ORC2-6HA cells (**Figure 2.13a**).

In summary, ORC hypo-sumoylation caused by *orc2-K406R* enhanced the function of a subset early origins while reducing that of a subset of late origins, the latter outcome likely an indirect consequence of enhanced early origin function. Thus, ORC hyper- and hypo-

sumoylation had opposing effects on origin function, consistent with a negative role for ORC sumoylation on origin licensing of some early S-phase origins.

2.2.9 ORC hypo- or hyper-sumoylation generated a dependence on other genome factors

Our analyses revealed that *ORC2-SuON* and *orc2-K406R* had generally opposing effects on origin firing, yet both types of effects might be expected to cause chromosomal replication imbalances that compromise genome stability and cell fitness. To test this prediction, genetic interactions of the two different ORC alleles were assessed against other relevant genome factors. First, both *ORC2-SuON* and *orc2-K406R,* but not their corresponding controls showed negative interactions with the null allele affecting the DNA damage checkpoint protein Rad9 (**Figure 2.12a and 2.12b; Figure 2.13b**). These data provided genetic evidence that alterations in ORC sumoylation status increased DNA lesions. Consistent with this finding, time course experiments showed that *orc2-K406R* cells exhibited increased levels of checkpoint activation in late S/G2 phase as evidenced by the increased phosphorylation of the checkpoint kinase Rad53, as well as increased levels of γH2A, an indicator for DNA lesions (**Figure 2.14a**). In addition, increased levels of Rad53 activation were also seen in *ORC2-SuON* cells compared with *ORC2-SuCtrl* (**Figure 2.14b**).

Another genetic indicator for the replication defect associated with both *ORC2-SuON* and *orc2-K406R* was slower growth upon the removal of the replication promoting factor Rrm3, which is implicated in regulating both fork movement and ORC function (Syed et al 2016). Both ORC alleles reduced the growth of *rrm3*Δ cells, while their corresponding controls did not (**Figure 2.12a and 2.12b; Figure 2.13a-2.13c**). Reduced cell growth of the *rrm3*Δ *ORC2-SuON* double mutant compared with either single mutant was clear from their smaller spore clone size, while the negative effect of *rrm3*Δ on *orc2-K406R* was more evident based on the cell doubling time measurement (**Figure 2.12a and Figure 2.13c**). These genetic data provided evidence that either ORC hyper- or hypo-sumoylation enhanced the yeast's dependence on other

important genome maintenance factors. Finally, rDNA stability was examined in cells containing *orc2-K406R* and those containing untagged or HA-tagged wild-type *ORC2*, as described above for *ORC2-SuON* (**Figure 2.7c**). Similar to *ORC2-SuON, orc2-K406R* led to an increase in rDNA marker loss, providing evidence that ORC hypo-sumoylation affected genomic stability (**Figure 2.12c**).

2.3 Discussion

Sumoylation regulates genome maintenance via targeting multiple substrates. While proteomic studies have established that yeast and human ORC subunits are sumoylated under stress conditions, the work here provided evidence that yeast Orc1, 2, 4 and 5 subunits are sumoylated during normal growth. We constructed two complementary ORC alleles resulting in hyper-sumoylated and hypo-sumoylated ORC to address the biological consequences of ORC sumoylation on genome replication and stability. High-resolution genome-scale analyses of chromosome replication revealed that ORC hyper- and hypo-sumoylation had opposing effects, with the former showing attenuating and the latter showing enhancing effects on subsets of early origins. These opposing phenotypic outcomes support the conclusion that sumoylation inhibited ORC activity. The role of ORC hyper-sumoylation was further assessed using biochemically reconstituted sumoylation and MCM loading systems, which revealed an inhibitory effect of ORC sumoylation on origin licensing. Taken together, our findings supported a model wherein ORC sumoylation preferentially impairs the use of a subset of early origins via disfavoring MCM recruitment onto origins (**Figure 2.12d**). Despite opposing effects on origin function, both hyper- and hypo-sumoylated ORC reduced genome stability and increased the dependence on genome maintenance factors for growth. We thus propose that tipping the balance of ORC sumoylation status in one direction or the other influences the balance of origin

function distributed across chromosomes and that ORC sumoylation status fine tunes the level and distribution of origin licensing required for genome stability (**Figure 2.12d**).

2.3.1 Biological consequences of ORC hyper-sumoylation and hypo-sumoylation

Alteration of substrate sumoylation status provides an effective strategy for assessing the functions of specific sumoylation. The *ORC2-SuON* allele enhanced sumoylation of Orc2 as well as Orc1 and Orc4, without changing global sumoylation levels or the sumoylation status of Mcm2 or 3, which directly interacts with ORC. Though it is difficult to completely exclude off-target effects, these data provided evidence that *ORC2-SuON* limited upregulation of sumoylation to ORC itself. Notably, *ORC2-SuON* led to yeast inviability, which was suppressed by a mild SUMO E2 mutant (*ubc9-10)* with concomitant reduction in Orc2 sumoylation. Thus, haploid *ORC2-SuON ubc9-10* cells could be used for experiments and compared to *ORC2-SuCtrl* in congenic *ubc9-10* backgrounds. Chromosomal replication analyses revealed that *ORC2-SuON* reduced the quantitative replication values of a subset of early origins. In addition, *ORC2-SuON* led to rDNA stability defects, consistent with reduced ORC activity. Thus, ORC hyper-sumoylation attenuates the essential role for ORC in origin function.

We also generated an ORC hypo-sumoylation mutant by focusing on Orc2. To achieve this feat, the sumoylation sites on endogenous Orc2 were identified using biochemical enrichment and mass spectrometry. While four sites were identified, systematic mutational analyses revealed that one (K406) was a major site for Orc2 sumoylation. K406 is conserved only within the *Saccharomyces* genus (**Figure 2.8b**) and *orc2-K406R* exhibited normal Orc2 levels and Orc2 association with chromatin, as well as cell viability. However, *orc2-K406R* enhanced the firing of a subset of early origins, thus producing an origin phenotype that was generally the opposite of that produced by *ORC2-SuON*. These results provided additional support to the notion that ORC sumoylation inhibited early origin licensing and that its levels were tuned to optimize ORC functions for normal yeast growth.

We note that in the two examined ORC mutants, early and late origins were still fired early and late in S phase, consistent with the theory that replication timing is controlled at the level of origin activation, not licensing. With this said, loaded MCM is the substrate for origin activation factors, thus alterations in origin licensing efficiencies affect origins' competitiveness for activation factors. Our data suggest that ORC sumoylation status influences the relative replication probability of origins at the levels of origin licensing efficiency.

2.3.2 Proper ORC sumoylation status is required for genomic stability and cell fitness

While ORC hyper- and hypo-sumoylation affected origin function in opposing directions, they both compromised genome stability and growth. Both types of alleles led to rDNA instability. In addition, even though *orc2-K406R* enhanced the firing of early origins and accelerated entry into S-phase, the mutant cells were delayed in completing S-phase compared to wild-type controls. Genetic data provided additional evidence that both ORC hyper-and hypo-sumoylation compromised cell fitness as these ORC states sensitized the growth of cells lacking two genome maintenance proteins, including the checkpoint protein Rad9 and the Rrm3 helicase that is implicated in regulating ORC and promoting replication fork progression through challenging chromosomal structures (Syed et al. 2016). The genetic interactions with *rrm3*Δ could reflect an enhanced requirement of Rrm3 when the origin firing program is altered. In the case of *orc2-K406R*, checkpoint activation and an increase γH2A levels were detected in G2-phase, suggesting that genome lesions are generated. The simplest interpretation is that excessive activation of early origins due to *orc2-K406R* limited the ability of late regions to complete replication by the end of S-phase, thus leaving DNA gaps that can induce the Rad9 checkpoint and produce genomic instability (**Figure 2.12d**). Other possibilities cannot be excluded such as those yet-to-be defined genome maintenance defects caused by two opposing Orc2 alleles. For example, the enhanced origin licensing associated with Orc2 hypo-sumoylation might cause early origin re-firing and partial DNA re-replication. In addition, given

that stress conditions can alter both protein sumoylation and origin firing, ORC sumoylation may change origin licensing in response to some of these conditions to avoid replication fork-transcription conflicts that might arise due to transcriptional changes in response to stress. Changes in origin licensing could also be used to help promote mutagenesis, which in turn could help cells adapt to new stressful environments.

2.3.3 ORC sumoylation acts to modulate the efficiency of origin licensing

ORC must perform multiple biochemical steps to load the double-hexameric MCM onto DNA (origin licensing) (Stillman 2022). Hyper-sumoylated ORC associated with DNA *in vitro* and hypo-sumoylated ORC did not reduce its chromatin association, suggesting that sumoylation does not significantly affect ORC association with DNA. In contrast, both *in vivo* and *in vitro* data suggest that hyper-sumoylated ORC impaired MCM recruitment to DNA. We note that higher *in vitro* ORC sumoylation levels compared with *in vivo* levels may lead to stronger effects on MCM loading than occurs *in vivo*. Nevertheless, the consistency between the biochemical data and the *in vivo* data supported a model wherein origin licensing efficiency is regulated by sumoylated ORC on origin DNA, with ORC sumoylated forms attenuating the efficiency of the MCM recruitment step (**Figure 2.12d**). This mechanistic explanation was adequate to explain the genome-scale replication data generated for this report, but further experiments defining the dynamics of ORC sumoylation *in vivo* could inform us of the more precise role of this modification. For example, examining Orc2 sumoylation during the cell cycle showed that its level increased as cells progressed through S phase and decreased in G2-M phase (**Figure 2.15**). These data raise the possibility that Orc2 sumoylation might play a role in preventing re-licensing of origins in late S phase, a hypothesis that will be examined in the future. In the meantime, additional examination of ORC sumoylation *in vitro* can inform us how sumoylated ORC impedes MCM recruitment *in vitro* and this would benefit from careful titration experiments of ORC sumoylation to establish conditions that match *in vivo* ORC sumoylation. While a more

complete understanding of the molecular roles of ORC sumoylation require further investigation, recent structural data offer potential insight. Specifically, in a recent cryo-EM structure of the ORC-Cdc6 complex, Orc2-K406 is located on the surface of the ORC complex proximal to both the Cdc6 binding site and Orc2-WHD (Feng et al. 2021). As SUMO is flexibly tethered to its modification site in other substrates (Powers et al. 2018; Attali et al. 2021), it is probable that sumoylation at this site can sterically impede one or more MCM recruitment steps (**Figure 2.12e**). While this model can provide an explanation for the negative effects of ORC sumoylation on MCM loading, the current data cannot yet indicate the full set of sumoylation site(s) on ORC most relevant to its functional regulation.

2.3.4 Models for ORC sumoylation dynamics and future challenges

The sumoylated forms of most substrates represent only a small fraction of the total amount of corresponding proteins (Zhao 2018). This observation holds for replication initiation factors such as Dbf4 and Mcm2-7 subunits and for the ORC subunits described in this report (Wei and Zhao 2016b; Psakhye et al. 2019). The small fraction of sumoylated proteins present at a given moment might be explained, at least in part, by the highly dynamic sumoylation and desumoylation cycles. However, it remains unclear how such a low-level of sumoylation on a given protein, for this discussion ORC, leads to a functional outcome. We consider three non-exclusive possibilities. First, sumoylation of ORC may occur efficiently only at specific loci. Perhaps different levels of desumoylation and/or sumoylation enzymes exist at different origins, such that only ORC at a subset of origins maintains a level of sumoylation sufficient to inhibit origin licensing. Second, we consider a “SUMO flicker” model, wherein cycles of sumoylation and desumoylation reduce the time windows for unmodified ORC to complete the multistep MCM loading reaction. Indeed, controlling the probability of a complete MCM loading reaction has been proposed as a mechanism for balancing the distribution of licensed origins across yeast chromosomes (Das et al. 2015; Hoggard et al. 2020). Third, a “collective effect” model

suggests that sumoylation of any one of the six ORC subunits would be sufficient to attenuate origin licensing, with sumoylation on more subunits leading to stronger inhibition. This model could explain the stronger genetic interactions seen for Orc2-SuON than *orc2-K406R* (**Figure 2.12a-2.12b**). Future investigations, including mapping additional sumoylated sites on ORC subunits, will be required to address these models.

While yeast and human origin control differ in many respects, it is notable that a small fraction of human Orc2 is sumoylated where it promotes recruitment of the KDM5A histone deacetylase that in turn generates a heterochromatin state that prevents re-replication of centromeric DNA (Huang et al. 2016). Thus, despite the differences between yeast and human ORC regulation as well as yeast and human origins, in both organisms sumoylated Orc2 helps to restrain origin function.

2.4 Materials and Methods

Yeast strains and procedures. Standard procedures were used for cell growth, medium preparation, epitope-tagging at endogenous loci, and tetrad dissection, unless otherwise indicated. All strains are isogenic to W1588-4C, a *RAD5* derivative of W303 (*MATa ade2-1 can1-100 ura3-1 his3-11,15 leu2-3,112 trp1-1 rad5-535*) and are listed in Table 1, only one strain is listed for each genotype, but at least two independent isolates of each genotype were used in the experiments. Mutations were introduced using a standard one-step integration PCR-based method. Correct tagging and mutations were verified by sequencing. The SuON tag is comprised of amino acids 418-621 of the SUMO binding domain of Ulp1 with the C580S substitution to abolish enzymatic activity, while the SuCtrl tag contains one additional substitution that abolished the high-affinity SUMO binding site (F474A) (Wei and Zhao 2017). At least two biological isolates of the same genotype were examined for each experiment. Strains containing *ubc9-10* were examined at its permissive temperature 24ºC. Doubling time of yeast

cells were measured as described (Hung et al. 2018). Briefly, cells were grown in a 96 plate at 30˚C with OD600 measured every 15 mins for 24 h, and doubling times were calculated from these data using a custom R script.

Synchronization and FACS analyses. Standard methods were used to synchronize yeast cells in G1 phase using alpha factor (Dhingra et al. 2019). *ubc9-10* log-phase cells were treated for 3 h with three doses of 5 ug/ml alpha factor each added at 1 h interval. Wild-type *UBC9* log-phase cells were treated for 1 h with 5 ug/ml alpha factor followed by the addition of 2.5 ug/ml alpha factor for another 30mins. When 95% cells had reached G1 arrest based on the presence of unbudded cells, cells were washed and released into fresh media without alpha factor to allow for entry into S phase. For PFGE gel analyses, 15 ug/ml nocodazole was added 30 mins after G1 release to prevent cells exiting the first cell cycle. Cell cycle stages were confirmed by Flow cytometry analyses (FACS) using standard procedures on a FACS Calibur flow cytometer, and data were analyzed with FCS7 software as described previously (Dhingra et al. 2019).

Detection of protein sumoylation by Ni-PD. Standard Ni-NTA pulldown was performed as previously described (Ulrich and Davies 2009). In brief, protein extracts prepared in 55% TCA were incubated in buffer A (6 M guanidine HCl, 100 mM sodium phosphate, pH 8.0, 10 mM Tris-HCl, pH 8.0) with rotation for 1 h at room temperature. The cleared supernatant was obtained after centrifugation for 20 min and was then incubated with Ni-NTA resin (Qiagen) in the presence of 0.05% Tween 20 and 4.4 mM imidazole at room temperature overnight with rotation. Beads were washed twice with buffer A supplemented with 0.05% Tween 20 and then four times with buffer C (8 M urea, 100 nM sodium phosphate, pH 6.3, 10 mM Tris-HCl, pH 6.3) supplemented with 0.05% Tween 20. HU buffer (8 M urea, 200 mM Tris-HCl, pH 6.8, 1 mM EDTA, 5% SDS, 0.1% bromophenol blue, 1.5% DTT, 200 mM imidazole) was used to elute proteins from the beads. Samples were loaded onto a 4-12% gradient Tris-Glycine gel (Bio-

Rad). Western blotting probed with antibodies recognizing the tagged proteins detected both sumoylated and unmodified substrates. The unmodified forms of the substrates were detected due to non-specific binding to the Ni-NTA beads and were not enriched in samples from cells expressing 8His-tagged yeast SUMO (Smt3). Denaturing conditions during protein extraction minimized desumoylation.

Enrichment of endogenously sumoylated Orc2 for mass spectrometry. Eight liters of cell cultures from strains containing 8His-Smt3 and Orc2-HA were harvested. Note that in order to facilitate SUMO (Smt3) identification in Mass spectrometry, the most C-terminal Ile of Smt3 was replaced by Arg to enable trypsin cleavage at this site. This variant of Smt3 (I96R) has been shown to support normal SUMOylation function (de Albuquerque et al. 2016). Cells were frozen in droplets with liquid nitrogen, and then ground into powder in a freezer mill (Freezer/Mill 6875D). The powder was thawed on ice, and cell extracts brought to a final concentration of 14% TCA and pelleted by centrifugation. After removing TCA, the pellets were washed with acetone and then resolubilized in buffer A as described above with shaking at 30˚C. Cleared supernatant supplemented with 0.05 % Tween 20 and 4.4 mM imidazole was then incubated with Ni-NTA resin (Qiagen) at room temperature overnight. Beads were washed once with buffer A containing 0.05% Tween 20 and once with buffer C as described above with 0.05% Tween 20. HU buffer described above was used to elute proteins from the beads. Eluted protein was then dialyzed for 2 h at room temperature against the RIPA buffer (150 mM NaCl, 50 mM Tris, pH7.4, 5 mM EDTA, 1.25% Triton X-100, 10 mM N-ethylmaleimide). Protein extract was then incubated with agarose beads crosslinked to HA antibodies and rotated overnight at 4˚C. Beads were washed once with RIPA buffer and with triethylammonium bicarbonate (TEAB) buffer (Sigma) before elution with 5% SDS in TEAB. Samples loaded onto a 4-12% gradient Tris-Glycine gel (BioRad) were examined by immunoblotting with anti-HA antibody and confirmed the presence of Orc2 sumoylated forms prior to mass spectrometry.

Immunoblotting analysis and antibodies. Protein samples were examined by SDS-PAGE and transferred to a 0.2-μm nitrocellulose membrane (G5678144, GE) for immunoblotting. Antibodies used were anti-HA (3F10), anti-V5 (R960-25, Invitrogen), PAP (P1291, Sigma), anti-Rad53 (yC-19, Santa Cruz), anti-Pgk1 (22C5D8, Invitrogen), anti-Orc1 (3E9) (Gabrielse et al. 2006), anti-Orc4 (1B1) (Gabrielse et al. 2006), and anti-Smt3 (Zhao and Blobel 2005). Validation of antibodies was provided either on the manufacturers' websites or in the cited references. For quantification purposes, membranes were scanned with a Fujifilm LAS-3000 luminescent image analyzer, which has a linear dynamic range of 10^4. Quantification of blots and generation of figures was performed with ImageJ and Photoshop.

Trichloroacetic acid (TCA) protein extraction. To examine protein levels and Rad53 phosphorylation, cell extracts were prepared as reported (Dhingra et al. 2019). Cell pellets were resuspended in 20% TCA and lysed by glass bead beating. The lysate was centrifuged to remove supernatant. Precipitated proteins were resuspended in Laemmli buffer (65 mM Tris-HCl pH 6.8, 2% SDS, 5% 2-mercaptoethanol, 10% glycerol, 0.025% bromophenol blue) with 2 M Tris to neutralize the solution. Prior to loading, samples were boiled for 5 mins and spun down for 10 min at 12,000rpm to remove insoluble materials.

Chromatin fractionation. Chromatin fractionation was performed as described previously with minor modifications (Schepers and Diffley 2001). Spheroplasts were lysed in lysis buffer containing 1% Triton X-100 and were laid upon a 30% sucrose cushion and centrifuged at 13,000 rpm for 20 min to separate the supernatant and chromatin fractions. The chromatin-bound fraction was washed with lysis buffer and resuspended in the same buffer. Equal volumes of samples from lysate, supernatant and chromatin fractions were precipitated with 20% TCA and resuspended in Laemmli buffer with the addition of 2 M Tris to neutralize TCA.

Pulsed field gel electrophoresis (PFGE). PFGE was performed as previously described (Cremona et al. 2012). Cells were addressed in G1 phase and then released into the cell cycle. Cells harvested from the indicated time points in Figure 4A were embedded in agarose plugs, spheroplasted, and deproteinized. Plugs were loaded into 0.5X TBE gels and run on a CHEF-DR III Pulsed Field Electrophoresis System (Bio-Rad) for 12 h to achieve chromosome separation. Gels were stained by ethidium bromide. Chromosome signal was measured using ImageJ and normalized to the G1 signal. The position of each chromosome was derived as described in (Lai 1993).

rDNA marker loss frequency. The loss frequency of the *ADE2-CAN1* cassette inside the rDNA array was measured as previously described (Fritze et al. 1997). Cells were grown to stationary phase over equal doubling times and plated on synthetic complete (SC) media for cell count totals. Cells were additionally plated on media containing canavanine (SC+Can) and incubated at 30˚C for 2 d after which colonies were counted. The frequency of marker loss was calculated as previously described using the formula $F_R = N_{can}/N_C$, where N_{Can} = number of colonies on SC+Can plates and N_C = number of cells plated on SC plates (Bernstein et al., 2011).

Whole-genome sequencing and copy number calculation. G1-arrested cells were released into cell cycle at 24ºC and S-phase progression was monitored by FACS. For *ORC2-SuON ubc9-10* and *ORC2-SuCtrl ubc9-10* cells growing at 24ºC, S-phase samples were collected 40 min after G1, whereas 30 min samples after G1-release were examined for *ORC2-HA* and *orc2-K406R-HA* cells growing at 30ºC. For both G1- and S-phase samples, 1.5 µg genomic DNA was used to generate libraries using a KAPA library kit at the iGO facility (Memorial Sloan Kettering Cancer Center, MSKCC) and sequenced with a HiSeq 4000 (Illumina, MSKCC) or a NextSeq1000 (Illumina) in the Department of Biomolecular Chemistry, UW Madison. At least 10

million 50-bp paired-end reads were generated per sample. Copy number calculation and chromosome map were derived following the general procedure as reported (Batrakou et al. 2020) In brief, reads were first mapped to the S288c reference genome (SGD, SacCer3), excluding repetitive sequences, and then summed in 0.5-kb bins with Genome Browser. Bins containing fewer than 600 reads were excluded. For each strain, the binned reads from S-phase samples at a given locus were divided by those from the G1 sample and normalized to the ratio of total reads. The normalized S/G1 read ratio was then adjusted to be between 1 and 2 based on the percentage of replication of the sample to derive a relative copy number of the particular locus. The maps of adjusted copy numbers were smoothed with the LOESS function. The sequencing data were deposited in the SRA database. The datasets were assigned the Biosample accession number PRJNA821839

Z-score analysis. Z-scores were calculated as described in (Hoggard et al. 2021). Briefly, sequencing coverages for each genomic nucleotide within *ORC2-SuON*, *ORC2-SuCtrl*, *orc2-KR-HA*, and *ORC2-HA* S-phase were determined. The G1 sample was normalized for sequencing depth and breadth (Skene and Henikoff 2015) and then mapped to 1015 windows, defined in (Hoggard et al. 2021), containing both origin and non-origin loci and spanning 10,001 bp (ORC site start ± 5 kbp). Thus mapped, coverages within each window were summed, and then converted to S/G1 ratios. Z-scores for each window within a particular strain were calculated using the following equation: Z-score = $(S/G1_{ratio} – \mu)/\sigma$, where μ and σ are the mean and standard deviation of all S/G1 ratios in a population, respectively. Only windows associated with confirmed origins were shown in figures. Confirmed origins were placed in three cohorts based on the Trep value measured in (Yabuki et al. 2002). Origins with the lowest 1/3 of Trep values, consistent with the earliest replicating origins were considered early, origins with the highest 1/3 of Trep values, consistent with latest replicating origins, were considered late. Origins with values between the two considered mid-firing.

Mass spectrometry and data analysis. Samples enriched for sumoylated Orc2 were reduced in 4 mM TCEP (Pierce), alkylated by 10 mM iodoacetamide (Sigma), and then quenched using 10 mM DTT (Fisher) with 30 min reaction time at each step at room temperature in the dark. The resulting sample was then loaded onto S-traps (Protifi) according to the manufacturer's instructions, and then digested overnight at 37°C with trypsin (Promega). Digested peptides were eluted from the S-trap and lyophilized for 48 h. Lyophilized peptides were resuspended in 0.1% trifluoroacetic acid (TFA, Fisher) and fractionated with high-pH reversed-phase chromatography (Pierce). Fractionated peptides were lyophilized and resuspended in 0.1% TFA. Peptide concentrations of the fractions were estimated with Fluorometric Quantitative Peptide Assay (Pierce). 250 ng peptides from each fraction were loaded onto a Bruker nano-elute UPLC in line with a Bruker TIMS-TOF pro mass spectrometer and eluted with a 45 min acetonitrile gradient. Raw TOF data were searched with PEAKS Studio X+ (Bioinformatic Solutions Inc.) against a yeast proteome downloaded from Uniprot (uniprot.org), with precursor and fragment mass accuracy at 10ppm, two missed cleavages, carbamidomethyl cysteine as fixed modification, methionine oxidation, peptide N-terminal acetylation and diglygly lysine (GlyGly) as variable modifications, and five modifications allowed per peptide. Identified sites were verified by manual inspection of MS2 spectra. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Deutsch et al. 2020) via the PRIDE (Perez-Riverol et al. 2022) partner repository with the dataset identifier PXD032977

Structure model. The structure of the yeast ORC bound to Cdc6 and origin DNA (pdb: 7mca) as described in (Feng et al. 2021) was analyzed in the Chimera software (Pettersen et al. 2004) with the addition of yeast SUMO (pdb:2EKE) as described by (Duda et al. 2007). One orientation of the SUMO was depicted in Figure 7E to highlight its potential to interfere with Cdc6 binding to ORC and/or the conformational change of the Orc2 WHD domain.

Protein purification. ORC, Cdc6 and Cdt1·Mcm2-7 were purified as previously described (Gros et al. 2015). SUMO, SUMO E1, SUMO E2, and the Siz1 and Siz2 SUMO E3s were purified as described (Zhao and Blobel 2005). Ulp1 was expressed in E. coli strain BL21 DE3 codon + RIL. Cells were grown at 37ºC to each 0.7 OD600 in LB supplemented with 50 mg/ml kanamycin and 34 mg/ml chloramphenicol, followed by cooling in an ice bath for 10 min. Induction was carried out for 4 h at 30ºC by the addition of 1 mM IPTG. Cells were harvested, resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.05 % NP-40, 10 % glycerol, 2 mM 2-mercaptoethanol) supplemented with protease inhibitor cocktail and lysed by sonication. Extracts were centrifuged for 20 min at 15,000 rpm in an SS34 rotor. The resulting soluble phase was recovered, supplemented with 10 mM imidazole and passed over a 5 ml His-trap column. The resin was washed with 10 column volumes of lysis buffer supplemented with 30 mM imidazole, and bound proteins then eluted with a gradient of 30 – 400 mM imidazole in lysis buffer over 10 column volumes.

ORC sumoylation and MCM loading assays. A linear 1kb DNA fragment containing ARS305 was generated by PCR and bound to streptavidin-coated M-280 Dynabeads as described (Remus et al., 2009). 1 pmol DNA beads (25 nM) was incubated with 82.5 nM ORC. ORC sumoylation reactions were performed in 40 µl reaction buffer (25 mM HEPES-KOH (pH 7.6), 0.1 M K-acetate, 0.02% NP-40, 10 mM Mg(OAc)2, 5% glycerol, 1 mM DTT and 5 mM ATP) containing 25 nM SUMO E1 (Uba2-Aos1), 240 nM SUMO E2 (Ubc9), 170 nM SUMO E3s (Siz1 and Siz2) and 3.5 µM Smt3 that contains thee lysine mutations to reduce SUMO chain formation. Reactions were mixed on ice and incubated at 30 °C for 30 min using a thermoshaker (Eppendorf) under constant agitation at 1,200 rpm. Beads were then washed once with 0.4 ml low-salt wash buffer (25 mM HEPESKOH [pH 7.6], 0.3 M K-acetate, 0.02% NP-40, 5 mM Mg(OAc)2, 1 mM EDTA, 1 mM EGTA, 10% glycerol, 1 mM DTT) before being

resuspended in 40 µl fresh reaction buffer supplemented with or without 0.9 µM Ulp1 and incubated again at 30 °C for 30 min at 1,200 rpm. The reaction was once again washed with 0.4 ml low-salt wash buffer before being resuspended in 40 µl fresh reaction buffer supplemented with 85 nM Cdc6 and 275 nM Cdt1·Mcm2-7 and either 5 mM ATP or ATPγS. The reactions were again incubated at 30 °C for 30 min at 1,200 rpm. Beads were then washed once with 0.4 ml low-salt wash buffer (25 mM HEPESKOH [pH 7.6], 0.3 M K-acetate, 0.02% NP-40, 5 mM Mg(OAc)2, 1 mM EDTA, 1 mM EGTA, 10% glycerol, 1 mM DTT) and once with either 0.4 ml low-salt buffer or high-salt wash buffer (as low-salt buffer, but 0.5 M NaCl instead of 0.3 M K-acetate). The beads were finally resuspended in 1x Laemmli buffer (2% SDS, 60 mM Tris-Cl pH 6.8, 5% 2-mercaptoethanol, 0.003% w/v bromophenol blue, 10% glycerol) before being analyzed by either Western blot or silver-stained SDS-PAGE.

2.5 Acknowledgments. We thank the Zhao lab members for discussions, Dr. Meng-Qiu Dong at the National Institute of Biological Sciences for initial help with mass spectrometry, and Dr. Yuangliang Zhai at the University of Hong Kong for help with structural modeling. This work was supported by NIH R01GM131058 and R35 GM145260 to X.Z., NIH R35GM141641 to C.A.F. and NIH R01-GM107239 to D. R.

Author contributions. Gemma Regan-Mochrie, Gerard Lynch, Dirk Remus, and Xiaolan Zhao designed experiments. Gemma Regan-Mochrie performed in vivo experiments, Gerard Lynch performed in vitro tests. Timothy Hoggard and Gemma Regan-Mochrie performed analysis of the whole genome sequencing experiments. Nikhil Bhagwat and Gemma Regan-Mochrie performed Mass Spectrometry experiments and data analysis. Gemma Regan-Mochrie and Xiaolan Zhao wrote the manuscript with editing from Dirk Remus and Catherine Fox. Specific contributions to individual figure panels are noted below each legend.

2.6 Figures

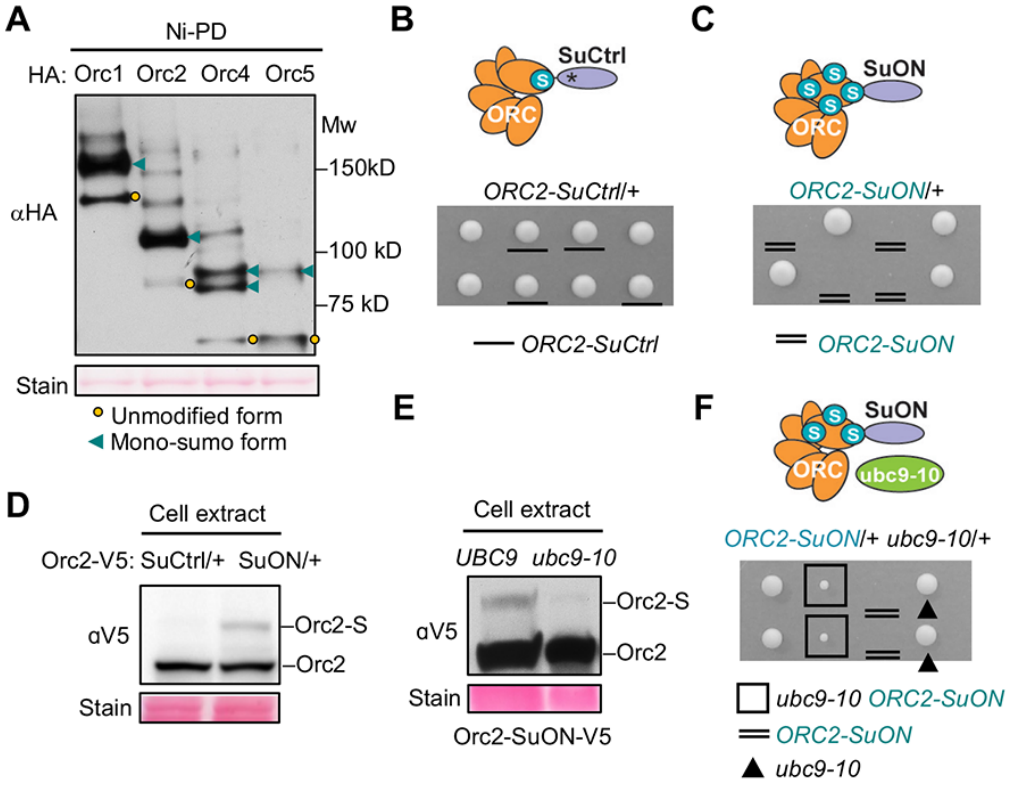


Figure 2.1 Increasing ORC sumoylation led to cell lethality that could be rescued by a SUMO E2 mutant.
A. *Sumoylation of Orc1, Orc2, Orc4, and Orc5 were detected under normal growth conditions. Sumoylated forms of each HA-tagged protein were enriched by Ni-PD. Unmodified and sumoylated bands indicated by orange circles and filled teal arrows, respectively. Equal loading indicated by ponceau stain (Stain).* ***B.*** *The SuCtrl tag fused with Orc2 supported normal growth. Top: a cartoon to depict the fusion of the SuCtrl to the Orc2 subunit of the ORC complex. S: SUMO. Asterisks indicate a point mutation in the SUMO binding domain (See text). Bottom: representative tetrads from diploid strain heterozygous for ORC2-SuCtrl. Cells were grown at 30˚C for 2 days. Strains were also heterozygous 8His-SMT3, which did not affect growth.* ***C.*** *Orc2 fused with SuON caused cell lethality. Top: a cartoon depicting the SuON tag fusion to the Orc2 subunit to increase ORC sumoylation. Symbols as in panel B. Bottom: representative tetrads were shown as in panel B.* ***D.*** *ORC2-SuON led to increased Orc2 sumoylation. Whole cell extracts were examined by immunoblotting; the unmodified and sumoylated forms are indicated. Equal loading is indicated by the ponceau stain.* ***E.*** *ubc9-10 reduced Orc2 sumoylation caused by ORC2-SuON. Orc2 sumoylation levels were examined in cells containing a single copy of wild-type or mutant UBC9.* ***F.*** *Cell lethality caused by ORC2-SuON was rescued by ubc9-10. Top: a cartoon to depict that in principle ubc9-10 could reduce ORC sumoylation. Bottom: representative tetrads were shown as in panel B except that cells were grown at 24˚C for 3 days.*

This figure was generated by the author of this thesis (adapted from Regan-Mochrie et al., Genes and Dev 2022)

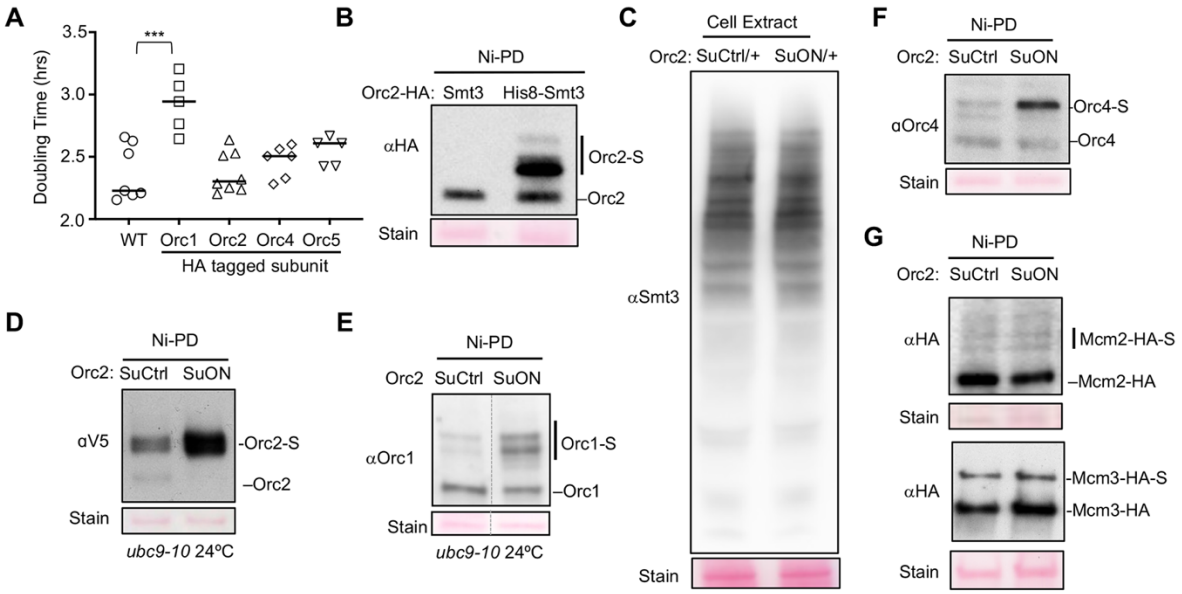


Figure 2.2 Examination of ORC sumoylation.
A. *Cell growth for HA-tagged ORC strains. Orc1-HA increased the doubling time of cells compared to wild-type and other tagged strains. Unpaired two-tailed t-test was used '*' P <0.05.* ***B.*** *Orc2 sumoylation bands were detectable only when SUMO (Smt3) was tagged by the 8His-tag when using the Ni-PD method.* ***C.*** *Examination of cell extract showed that total sumoylation levels as detected by anti-Smt3 antibody on immune blots were unaffected by ORC2-SuON.* ***D-F.*** *ORC2-SuON led to increased sumoylation levels for Orc2 (D), Orc1 (E), and Orc4 (F). Experiments were conducted as in B.* ***G.*** *Mcm2 and Mcm3 sumoylation levels were not altered in ORC2-SuON ubc9-10 compared to ORC2-SuCtrl ubc9-10 cells. Cells were arrested in G1 phase at 24˚C (top panels) and treated with 0.03% MMS for 2 hrs at 24˚C to better detect Mcm3 sumoylation (bottom panels).*

Panels ***a-g*** *were generated by the author of this thesis with assistance from Timothy Hoggard for panel* ***a*** *(adapted from Regan-Mochrie et al., Genes and Dev 2022)*

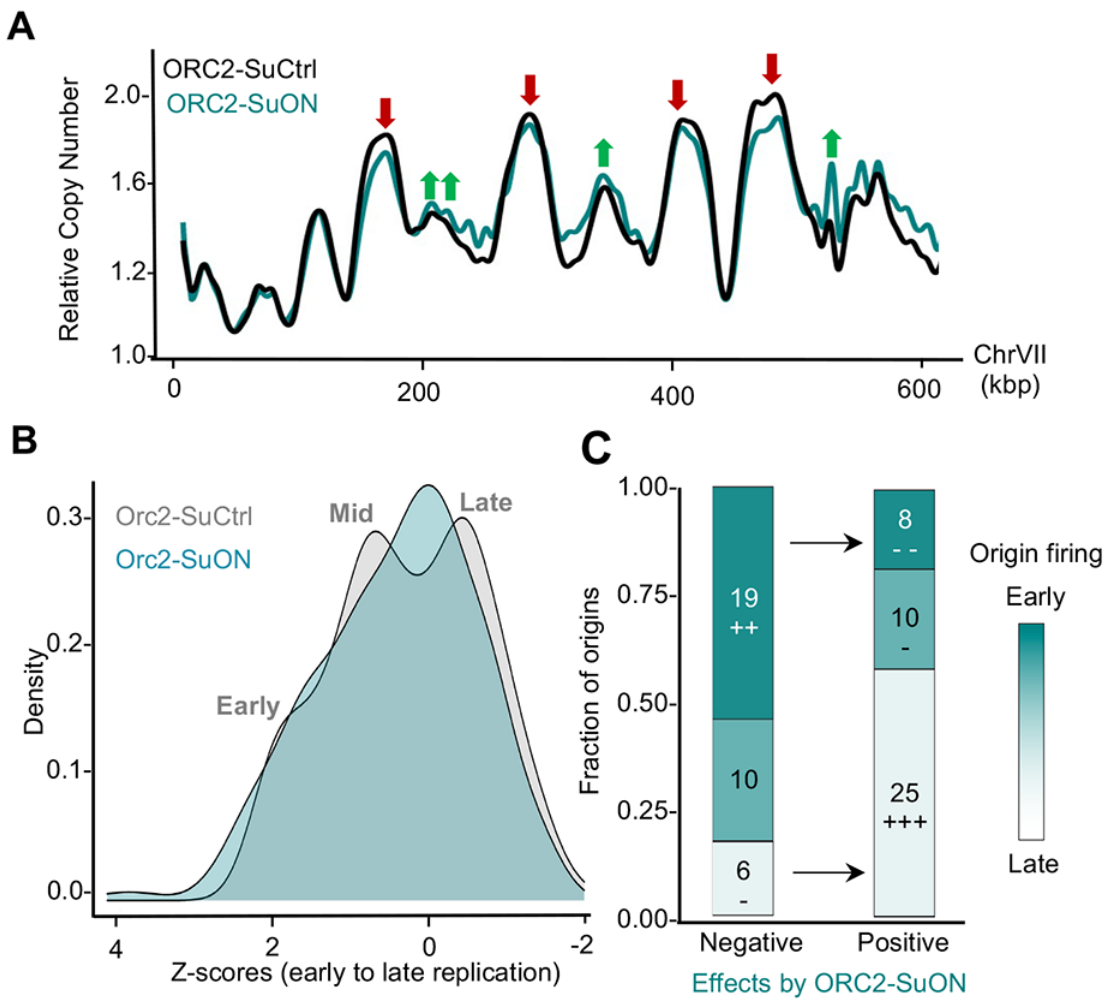


Figure 2.3 ORC2-SuON led to preferential inhibition of a subset of early origins and a temporally homogenized replication profile.
A. *Copy number analysis based on genome sequencing data depicted as a chromosomal replication scan for chromosome VII for ORC2-SuCtrl (black) and ORC2-SuON (teal) cells. Red and green arrows mark the origins that show decreased and increased firing in ORC2-SuON compared with ORC2-SuCtrl cells, respectively.* ***B.*** *ORC2-SuON altered the replication timing program. Kernel density estimation was used to generate a plot of origin Z scores vs. density of replication origins in ORC2-SuON (teal) and ORC2-SuCtrl (grey).* ***C.*** *The ORC2-SuON/ORC2-SuCtrl Z-score ratio was determined for each origin, and the origins negatively affected (ratio cut-off of < 0.8) or positively affected (ration cut-off of >1.2) by ORC2-SuON were parsed by their experimentally determined Trep value (Yabuki et al. 2002). The P-values for enrichment (+) or depletion (-) of the various types of Trep origins among affected origins are indicated (+++/--- </= 0.001; ++/-- </= 0.01; +/- </= 0.05).*

Panels ***a-c*** *were generated by the author of this thesis and Timothy Hoggard (adapted from Regan-Mochrie et al., Genes and Dev 2022)*

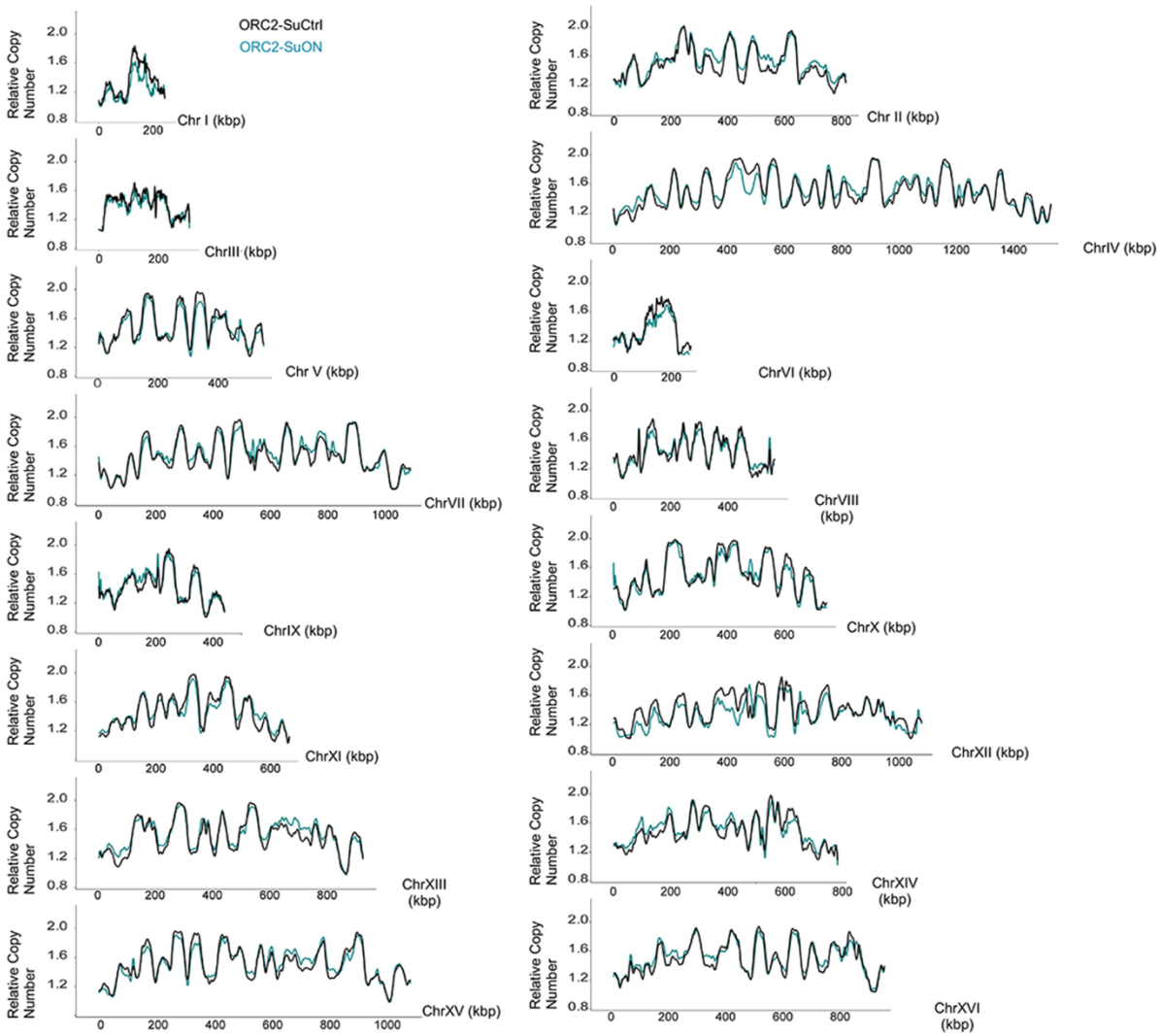


Figure 2.4 Copy number analyses of the whole genome sequencing data for ORC2-SuON and ORC2-SuCtrl cells.
Chromosomal scans for all 16 yeast chromosomes. Cells were arrested in G1 at 24˚C and subsequently released into cell cycle. Samples of G1 arrested and 40 mins after release were subjected to whole-genome sequencing and relative copy number changes were plotted.

This figure was generated by the author of this thesis and Timothy Hoggard (adapted from Regan-Mochrie et al., Genes and Dev 2022)

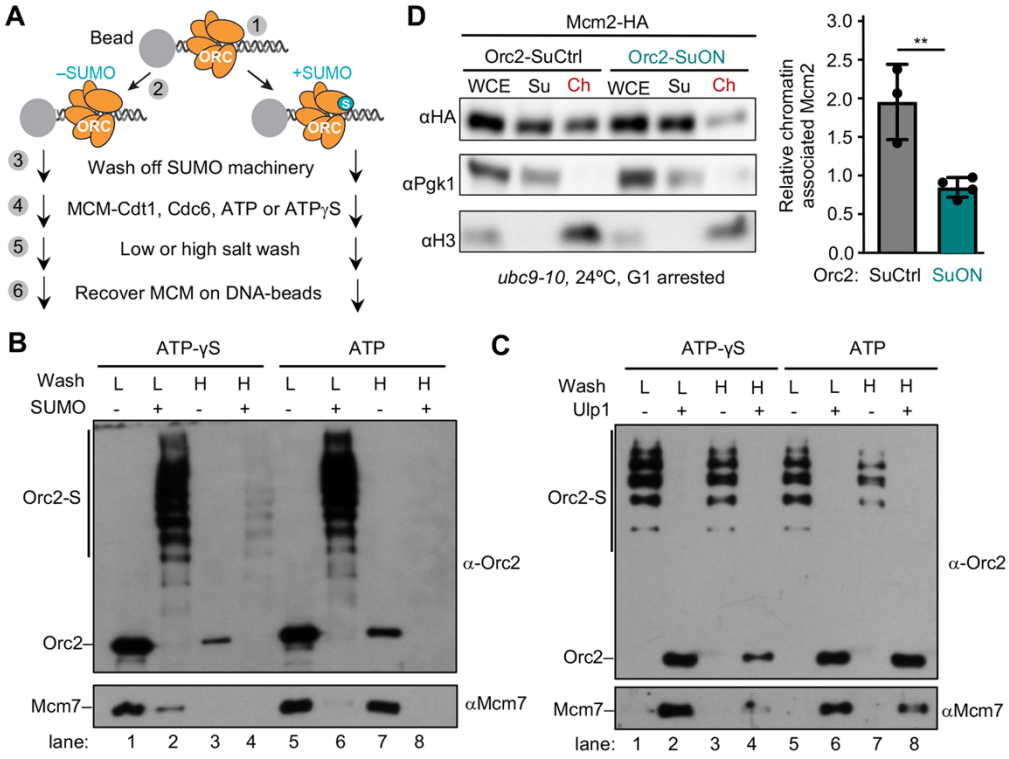


Figure 2.5 ORC hyper-sumoylation prevented MCM loading.
A. *Reaction scheme to determine the impact of ORC hyper-sumoylation on MCM recruitment and loading in vitro (see text for details).* ***B.*** *Immunoblot analysis of DNA-bound fractions to assess Orc2 and Mcm7 DNA-association in vitro. Following ORC sumoylation (+) or mock treatment (-), MCM loading reactions were performed in the presence of ATPγS (lanes 1-4) or ATP (lanes 5-8), as indicated. DNA beads were washed with low-salt (L) or high-salt (H) buffer, as indicated.* ***C*** *Immunoblot analysis of Orc2 and Mcm7 association with DNA in vitro. Following ORC sumoylation, MCM loading reactions were performed in the absence (-) or presence (+) of Ulp1 and either ATPγS (lanes 1-4) or ATP (lanes 5-8). DNA beads were washed either with low-salt (L) or high-salt (H) buffer.* ***D.*** *Assessment of chromatin-bound and soluble MCM from G1-arrested ORC2-SuCtrl and ORC2-SuON cells. Left: Immunoblot showing Mcm2 in the soluble and chromatin fractions in G1 cells. Pgk1 served as the soluble control, while histone H3 served as the chromatin-bound control. Right: quantification of the ratio of Mcm2 to H3 recovered in the chromatin fraction. An unpaired two-tailed t-test generated a P-value of <0.01 (**) for the difference between Mcm2 recovery in ORC2-SuCtrl and ORC2-SuON cells.*

Panels ***a*** *and* ***d*** *were generated by the author of this thesis, panels* ***b*** *and* ***c*** *were generated by Gerard Lynch (adapted from Regan-Mochrie et al., Genes and Dev 2022)*

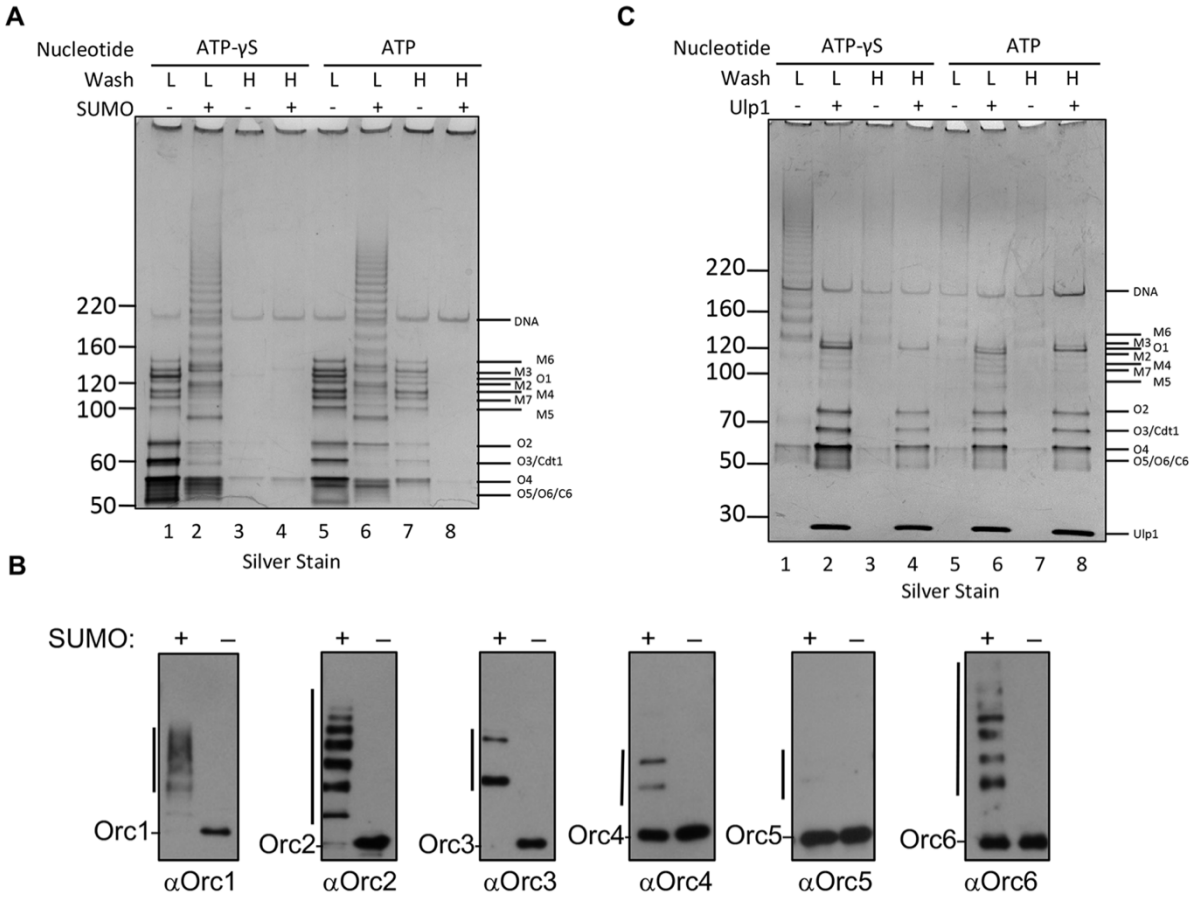


Figure 2.6 ORC hyper-sumoylation prevented MCM loading on DNA.
A. *DNA-bound fractions from experiment in Figure 3B were analyzed by SDS-PAGE and silver stain. O1-6: Orc1-6; M2-7: Mcm2-7.* ***B.*** *The six ORC subunits are sumoylated to different degrees in the in vitro sumoylation reactions. The sumoylation reactions were carried out as Figure 3A-3B. All reactions contain SUMO E1, E2, Siz2 SUMO E3. SUMO was only omitted in the reactions loaded in the right lane. The reactions were examined by immunoblotting probed using antibody against each ORC subunits as indicated. Line indicates the sumoylated forms of each subunit while unmodified forms are labeled.* ***C.*** *DNA-bound fractions from experiment in Figure 3C were analyzed by SDS-PAGE and silver stain.*

Panels ***a-c*** *were generated by Gerard Lynch (adapted from Regan-Mochrie et al., Genes and Dev 2022)*

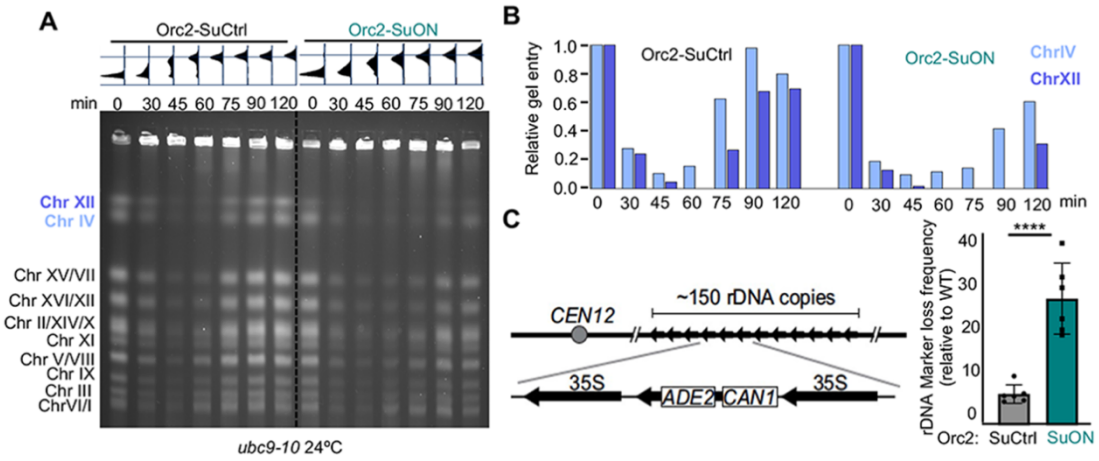


Figure 2.7 ORC2-SuON resulted in inhibition of chromosomal replication with the most severe effects on the rDNA-containing chromosome.
A. *PFGE analysis of ORC2-SuCtrl and ORC2-SuON cells during an S-phase time course initiated after release of cells from G1-arrest. Chromosomal DNA was stained with Ethidium Bromide after PFGE. Dotted line demarks the two different strains that were analyzed on the same gel. The cell-cycle progress for each time point analyzed was determined by FACS and is shown above the gel.* ***B.*** *DNA gel entry signal normalized to G1 for chromosome IV and chromosome XII.* ***C.*** *The effect of ORC2-SuON on rDNA marker loss was determined. Left: schematic of the rDNA marker loss frequency assay. Right: rDNA marker loss frequency between ORC2-SuCtrl and ORC2-SuON cells. An unpaired two-tailed t-test was used to generate a P-value of <0.0001 (****) for the difference between two types of cells.*

This figure was generated by the author of this thesis (adapted from Regan-Mochrie et al., Genes and Dev 2022)

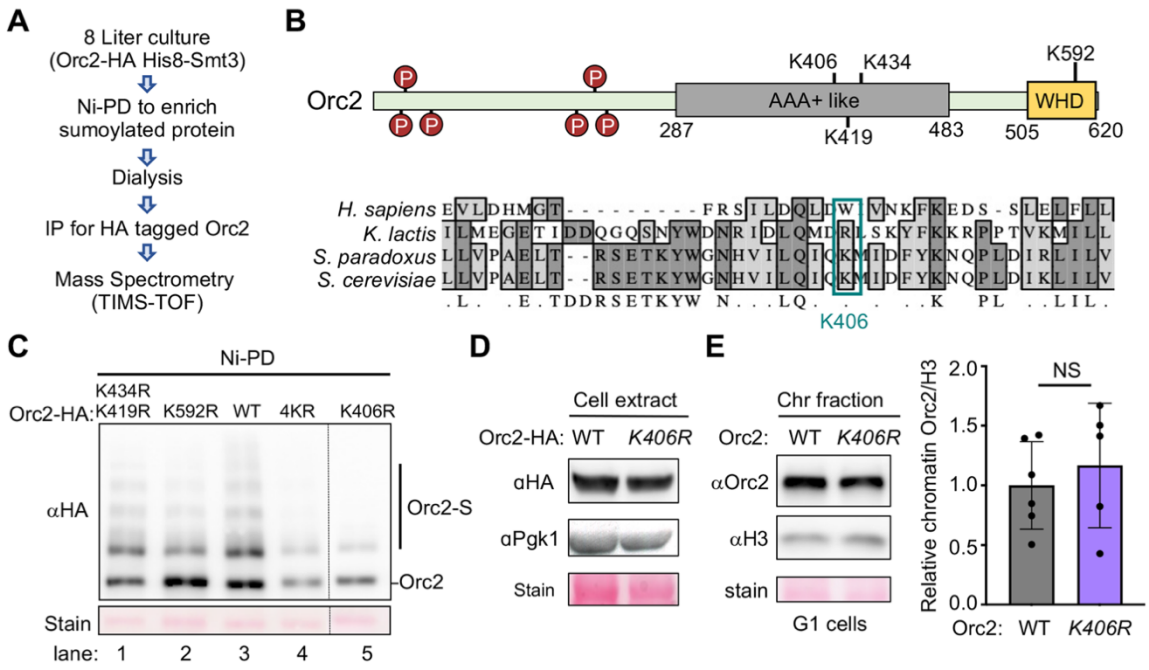
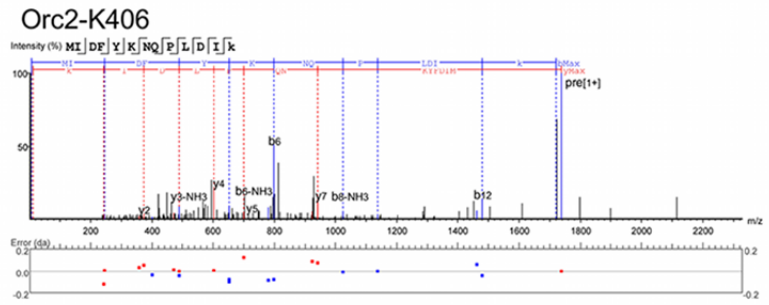


Figure 2.8 Mapping the Orc2 sumoylation sites.
A. *Experimental outline for mapping endogenous Orc2 sumoylation site(s) (details in Methods).* ***B.*** *Schematics of Orc2 domains and modification sites. Top: CDK phosphorylation sites (P) and four lysine residues mapped via mass spectrometry. Bottom: conservation analyses of Orc2 orthologs. K406 is only conserved within the Saccharomyces genus.* ***C.*** *The effects of mutating four lysine residues on Orc2 sumoylation. Orc2 sumoylation levels were compared among congenic cells that differed by the indicated ORC2 genotypes on the same gel with dotted line indicating the removal of superfluous lanes. Equal loading indicated by Ponceau-S stain (Stain).* ***D.*** *Levels of Orc2 determined in whole cell extracts from cells with the indicated genotypes. Pgk1 served as a loading control. The Ponceau S-stained portion of the blot also shown.* ***E.*** *Orc2 in the chromatin fractionation was examined for the indicated genotype. Immunoblot results and a Ponceau S-stained section of the membrane are on left and quantification of independent experiments are on the right. Two tailed t-tests generated P-values indicating no difference in Orc2 chromatin-association between the two examined genotypes.*

Panels ***a****,* ***c-e*** *were generated by the author of this thesis, panel* ***b*** *was generated by the author of this thesis with assistance from Timothy Hoggard (adapted from Regan-Mochrie et al., Genes and Dev 2022)*

A



B

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	132.05	114.04	115.02	66.52	M					13
2	245.13	227.18	228.11	123.07	I	1607.84	1589.83	1590.82	804.42	12
3	360.16	342.15	343.14	180.58	D	1494.78	1476.75	1477.75	747.88	11
4	507.23	489.26	490.20	254.11	F	1379.73	1361.72	1362.70	690.37	10
5	670.29	652.36	653.26	335.65	Y	1232.66	1214.65	1215.64	616.83	9
6	798.47	780.46	781.36	399.69	K	1069.60	1051.59	1052.57	535.24	8
7	912.43	894.42	895.40	456.71	N	941.42	923.43	924.38	471.25	7
8	1040.49	1022.48	1023.46	520.74	Q	827.46	809.45	810.44	414.23	6
9	1137.54	1119.53	1120.51	569.28	P	699.40	681.43	682.38	350.20	5
10	1250.61	1232.61	1233.58	625.81	L	602.35	584.34	585.32	301.68	4
11	1365.68	1347.63	1348.62	683.33	D	489.26	471.26	472.19	245.13	3
12	1478.76	1460.70	1461.71	739.87	I	374.13	356.23	357.21	187.62	2
13					K(+114.04)	261.16	243.15	244.13	131.08	1

Figure 2.9 Mass spectrometry data indicate sumoylation at K406 of Orc2.
A. *Example of MS2 scan of an Orc2 peptide from the three that were found identifying K406 SUMOylation. Lower panel shows fragment mass error.* ***B.*** *Table of the b and y ions matching this peptide and the modification.*

*Panels **a** and **b** were generated by Nikhil Bhagwat (adapted from Regan-Mochrie et al., Genes and Dev 2022)*

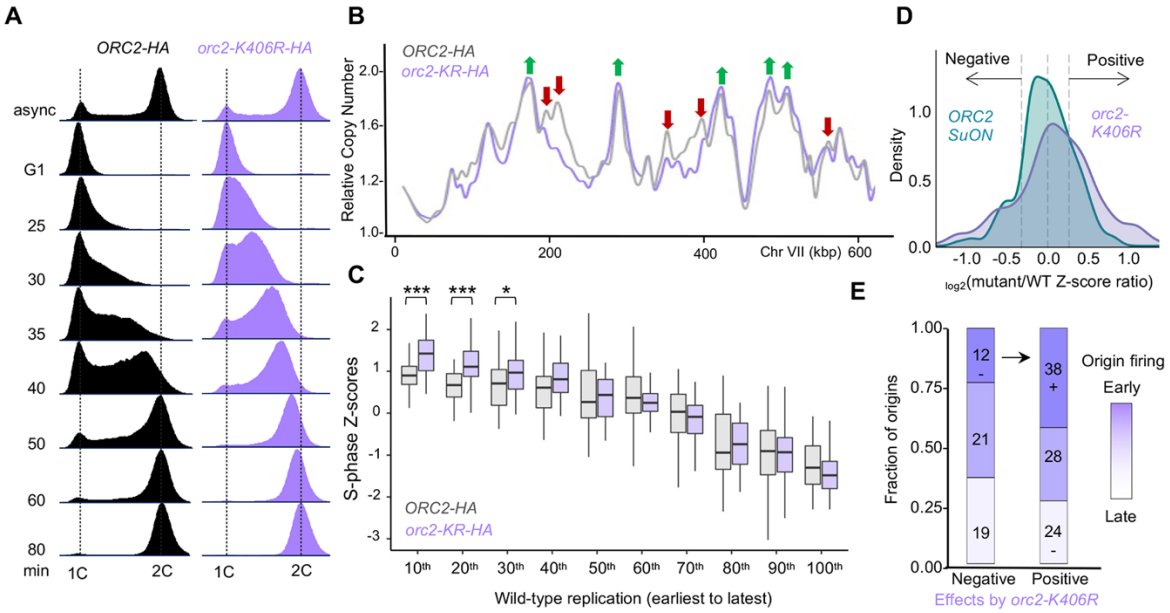


Figure 2.10 Orc2 hypo-sumoylation increased early origin firing.
A. *orc2-K406R cells exhibited a bi-phasic replication profile as shown by flow cytometry analysis at 24˚C. At 25 and 30 mins orc2-K406R entered S phase more rapidly, whereas at 50 and 60 mins orc2-K406R appeared to be lagging behind wild-type cells.* ***B.*** *Copy number analysis based on genome sequencing data depicted as a chromosomal replication scan for chromosome VII for ORC2-HA (black) and orc2-K406-HA (light purple) cells. Red and green arrows mark the origins that show decreased and increased firing in orc2-K406 compared with ORC2 cells, respectively.* ***C****. Z-score distribution analyses. Z scores were plotted for the earliest to the latest fired origins. Origins were parsed into 10 distinct cohorts by their firing time. The Z scores for origins within each decile were presented as box-and-whiskers plots from the earliest 10% (10th) to the latest 10% (100th) decile. Wilcoxon rank sum P values for the differences between the two genotypes are indicted (**<0.01 and ***<0.001).* ***D.*** *To compare the effects of hypermorphic Orc2 sumoylation (ORC2-SuOn, green) to hypomorphic Orc2 sumoylation (orc2-K406R) on the origin replication timing program, the log2 of the ratio between the Z-score determined for each origin in the relevant mutant to the Z-score for that origin in the wild-type ORC2 strain was determined. Kernel Density Estimation was then used to generate a plot of log2 of the origin Z-score rations versus density of replication origins.* ***E.*** *The orc2-K406R/ORC2 Z-score ratio for each origin was determine. The origins negatively affected (ratio cut-off of < 0.8) or positively affected (ratio cut-off of >1.2) by orc2-K406R were then parsed by their experimentally determined Trep values and the data presented in histograms as in Figure 2C. The P-values for enrichment or depletion were as described in Figure 2C.*

Panel ***a*** *was generated by the author of this thesis, panels* ***b-e*** *were generated by the author of this thesis and Timothy Hoggard (adapted from Regan-Mochrie et al., Genes and Dev 2022)*

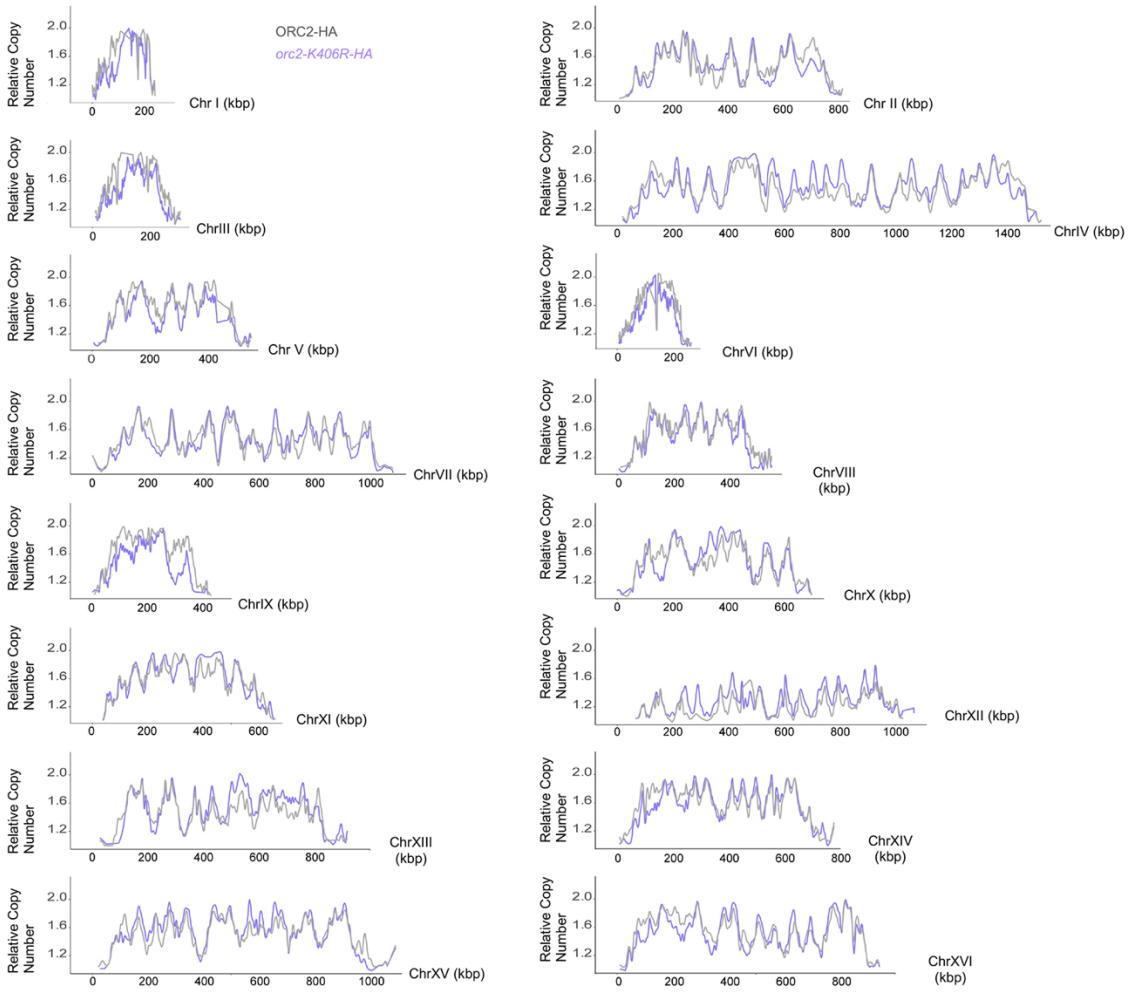


Figure 2.11 Copy number analyses of the whole genome sequencing data for ORC2-HA and orc2-K406R-HA cells.
Experiments and analyses were conducted, and data are presented as Figure S2, except cells were grown at 30ºC and 30 min post-release samples were analyzed.

This figure was generated by the author of this thesis and Timothy Hoggard (adapted from Regan-Mochrie et al., Genes and Dev 2022)

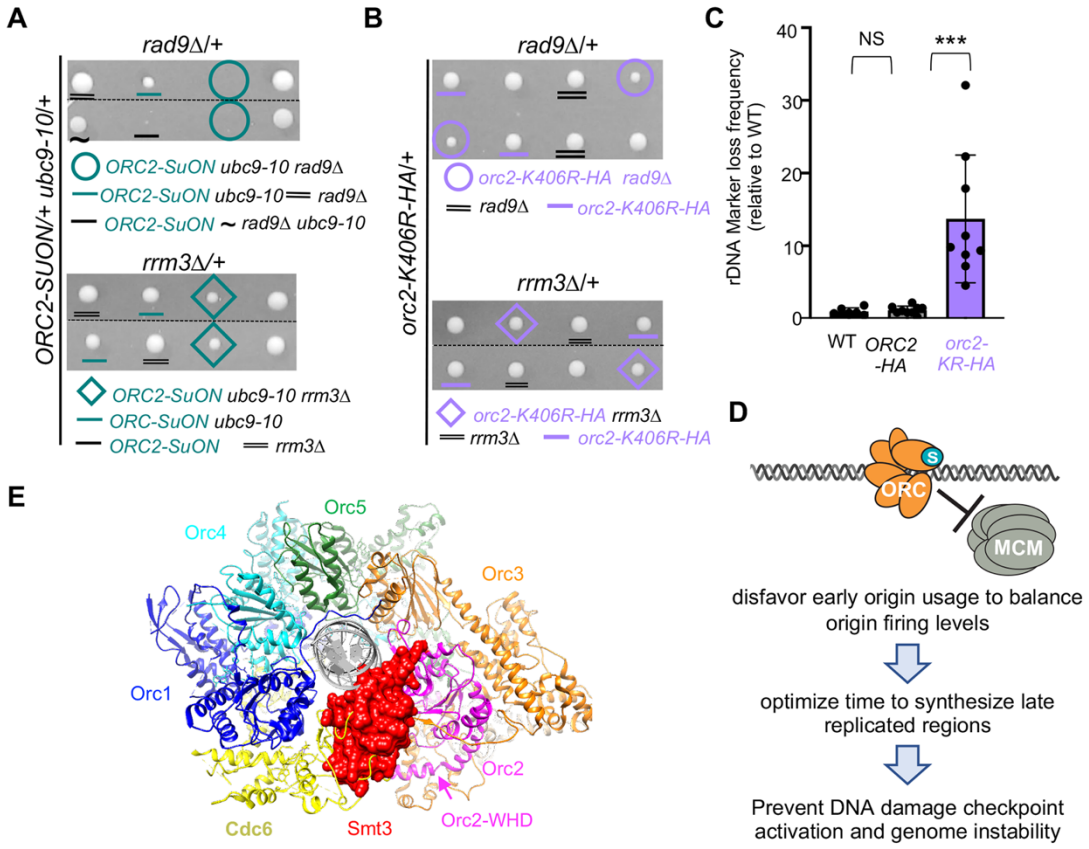


Figure 2.12 Genetic analysis of the ORC hypo- or hyper-sumoylation alleles and a structural model for Orc2 sumoylation.

A-B. *Genetic interactions between ORC2-SuON (A) and orc2-K406R (B) with rad9Δ or rrm3Δ. Controls for ORC2-SuON and orc2-K406R did not show negative interactions with either rad9Δ or rrm3Δ (Fig. S6). Representative tetrads heterozygous of indicated genotype are shown. Cells were grown at 24˚C for 3 d in A and at 30˚C for 2 d in B.* ***C.*** *The effect of orc2-K406R on rDNA marker loss was determined as in Figure 4C. An unpaired two-tailed t-test was used to generate a P-value of <0.001 (***) for the difference between two types of cells.* ***D.*** *A model for how ORC sumoylation affected replication initiation and genome stability consisted with the data. ORC sumoylation disfavors MCM association at a subset of early origins. This regulation allowed for a more balanced distribution of licensed origins across chromosomes that promoted timely replication of genomic regions duplicating late in S-phase. In the absence of this balance, checkpoint activation mechanisms were stimulated, and rDNA instability ensued.* ***E.*** *A structural view of how Orc2 K406 sumoylation (SUMO, red) may sterically interfere with Orc2-WHD domain conformational changes that aid MCM recruitment and/or Cdc6 binding to ORC that is essential for MCM recruitment (Feng et al. 2021).*

Panels ***a-c*** *were generated by the author of this thesis, panel* ***d*** *was generated by the author of this thesis and Xiaolan Zhao and panel* ***e*** *was generated by Dr. Yuangliang Zhai (adapted from Regan-Mochrie et al., Genes and Dev 2022)*

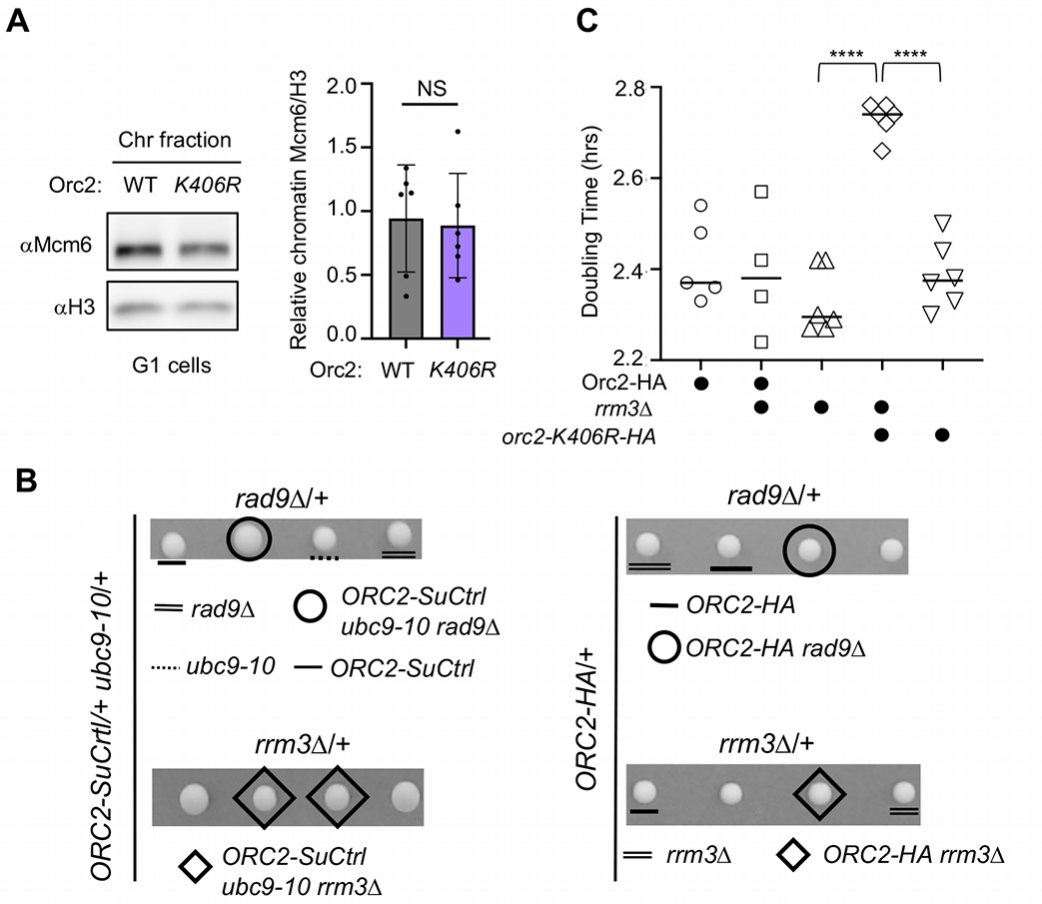


Figure 2.2.13 Genetic analyses of ORC2-SuCtrl and ORC2-HA.
A. *Assessment of chromatin-bound Mcm6 from G1-arrested ORC2-HA and orc2-K406R-HA cells. Left: Immunoblot showing Mcm6 in the chromatin fractions. Histone H3 served as the chromatin-bound control. Right: quantification of the ratio of Mcm6 to H3 recovered in the chromatin fraction. An unpaired two-tailed t-test found no significant difference (NS) between Mcm6 recovery in ORC2-HA and orc2-K406R-HA cells.* ***B.*** *ORC2-SuCtrl (left) and ORC2-HA (right) did not show negative genetic interactions with rad9Δ (top) and rrm3Δ (bottom). Experiments and analyses were conducted, and data are presented as Figure 7A-7B.* ***C****. Cell growth measured by doubling time for strains with indicated genotype. Unpaired two-tailed t-test was used; '****' p<0.0001.*

*Panels **a** and **b** were generated by the author of this thesis with assistance from Timothy Hoggard for panel **c** (adapted from Regan-Mochrie et al., Genes and Dev 2022)*

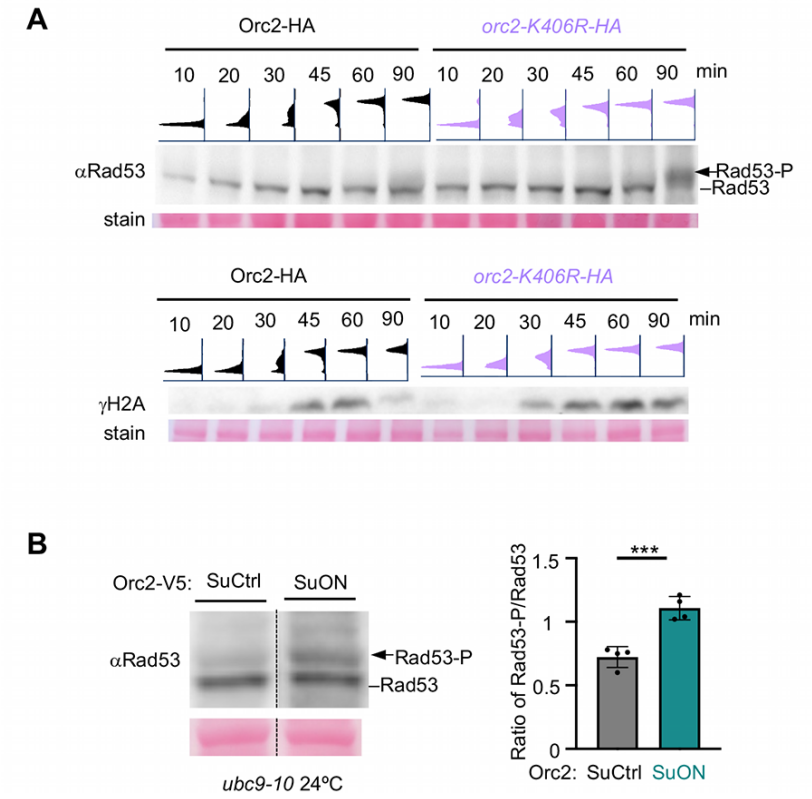


Figure 2.14 ORC hyper- and hypo-sumoylation increased checkpoint activation.
A*. Top: orc2-K406R cells showed Rad53 activation in G2, not during S phase. FACS analysis of cells arrested in G1 and released into YPD at 24˚C and immunoblot of Rad53 wherein an upshift in Rad53 band indicating its phosphorylation. Bottom: orc2-K406R-HA cells exhibited increased levels of gH2A levels compared with Orc22-HA strains in G2/M phase.* ***B.*** *Asynchronized ORC2-SuON cells exhibited increased levels of Rad53 phosphorylation compared with ORC2-SuCtrl cells when grew at 24ºC in the ubc9-10 background.*

This figure was generated by the author of this thesis (adapted from Regan-Mochrie et al., Genes and Dev 2022)

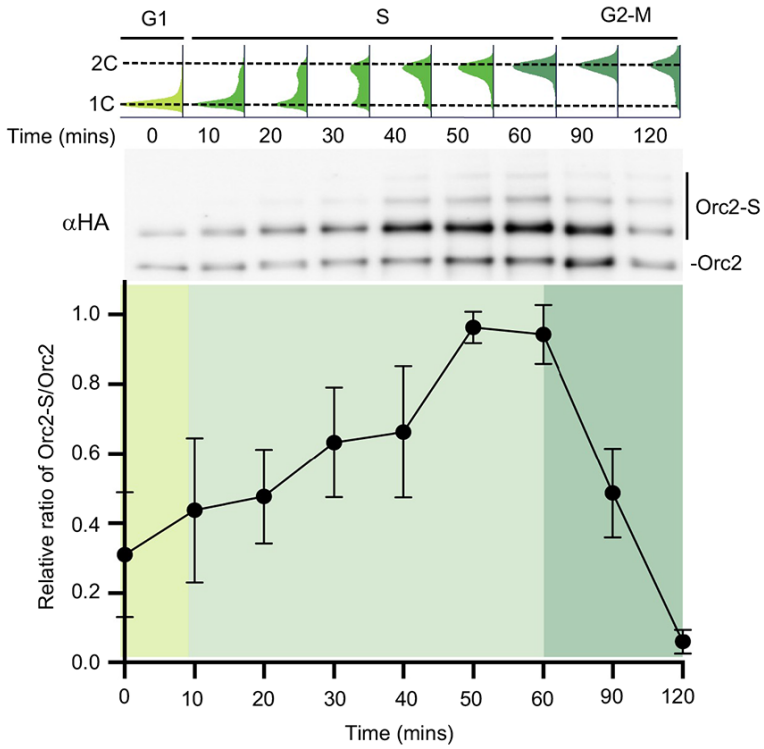


Figure 2.15 Orc2 sumoylation level progressively increases as S phase progresses and reduces in G2-M phase.
Top: *FACS shows cell progression through the cell cycle.* ***Middle:*** *A representative immunoblot displays the sumoylated and unmodified forms of HA-tagged Orc2 at each indicated time point. Ni-PD experiments were performed to enrich sumoylated proteins, as described in the main text.* ***Bottom:*** *The ratio of sumoylated to unmodified Orc2 signals on the immunoblots was quantified for three biological repeats. The highest ratio determined for each replicate was set to 1; the averages and standard deviations are shown.*

This figure was generated by the author of this thesis (adapted from Regan-Mochrie et al., Genes and Dev 2022)

Table 1: Strains used in Chapter 2

Strain	Genotype
T2071-1	*MATa ORC1-6HA::NAT 8His-SMT3::TRP1*
T2072-1	*MATa ORC2-6HA::NAT 8His-SMT3::TRP1*
T2073-1	*MATa ORC4-6HA::HYG 8His-SMT3::TRP1*
T2074-1	*MATa ORC5-6HA::HYG 8His-SMT3::TRP1*
T2105	*ORC2-3V5-SuON::KAN/+ 8His-SMT3::TRP1/+*
T2106	*ORC2-3V5-SuCtrl::KAN/+ 8His-SMT3::TRP1/+*
T2134	*ORC2-3V5-SuON::KAN/+ 8His-SMT3::TRP1/+ ubc9-10::NAT/+*
T2134-9d	*ORC2-3V5-SuON::KAN 8His-SMT3::TRP1 ubc9-10::NAT*
T2134-10a	*MATa ORC2-3V5-SuON::KAN ubc9-10::NAT*
T2135-3a	*ORC2-3V5-SuCtrl::KAN 8His-SMT3::TRP1 ubc9-10::NAT*
T2135-2b	*MATa ORC2-3V5-SuCtrl::KAN ubc9-10::NAT*
X8175-11a	*MATa ORC2-3V5-SuON::KAN MCM2-3HA::HIS3 ubc9-10::NAT*
X8176-17a	*MATa ORC2-3V5-SuCtrl::KAN MCM2-3HA::HIS3 ubc9-10::NAT*
X8536-4c	*MATa ORC2-3V5-SuON::KAN MCM3-3HA::HIS3 ubc9-10::NAT*
X8536-10b	*MATa ORC2-3V5-SuCtrl::KAN MCM3-3HA::HIS3 ubc9-10::NAT*
X8176-11d	*MATa MCM2-3HA::HIS3*
X8527-3d	*ORC2-6HA::NAT 8His-smt3-I96R::KAN*
X8799-9b	*ORC2-3V5-SuCtrl::KAN ubc9-10::NAT rDNA::ADE2-CAN1*
X8800-1d	*ORC2-3V5-SuON::KAN ubc9-10::NAT rDNA::ADE2-CAN1*
X8817	*ORC2-3V5-SuON::KAN/+ ubc9-10::NAT/+ rad9Δ::HIS3/+*
X8818	*ORC2-3V5-SuCtrl::KAN/+ ubc9-10::NAT/+ rad9Δ::HIS3/+*
X8819	*ORC2-3V5-SuON::KAN/+ ubc9-10::NAT/+ rrm3Δ::KAN/+*
X8820	*ORC2-3V5-SuCtrl::KAN/+ ubc9-10::NAT/+ rrm3Δ::HIS/+*
T2224-4	*orc2-K406R, K419R, K434R, K592R-6HA::NAT 8His-Smt3::TRP1*
T2255-13c	*orc2-K592R-6HA::NAT 8His- SMT3::TRP1*
T2256-2	*orc2-K419R, K363R-6HA::NAT 8His-SMT3::TRP1*
X8701-9d	*ORC2-HA::NAT*
X8699-13d	*orc2-K406R-HA::NAT*
X8797	*orc2-K406R-HA::NAT/+ rrm3Δ::KAN/+*
X8798	*ORC2-HA::NAT/+ rrm3Δ::KAN/+*
X8797-3C	*orc2-K406R-HA::NAT rrm3Δ::KAN*
X8798-7B	*ORC2-HA::NAT rrm3Δ::KAN*
X8798-2D	*rrm3Δ::KAN*
X8814	*orc2-K406R-HA::NAT/+ rad9Δ::HIS3/+*
X8815	*ORC2-HA::NAT/+ rad9Δ::HIS3/+*

CHAPTER 3 : THE BALANCING ACT OF THE LEADING STRAND POLYMERASE

3.1 Introduction

The eukaryotic genome is replicated through numerous sequence units referred to as replicons. Within each replicon, a pair of replisomes assemble at the replication initiation site (origin) to establish twin replication forks that travel outwards until merging with neighboring replicons (Burgers and Kunkel 2017). Within the replisome, the leading and the lagging strand DNA polymerases, Pol ε and Pol δ, respectively, are primarily responsible for copying the two parental strands (Pursell et al. 2007; Daigaku et al. 2015). While Pol δ acts in a discontinuous manner, Pol ε carries out continuous synthesis that extends many kilobases and is a primary driver of replisome progression by promoting the rapid progression of the replicative helicase (Gan et al. 2017; Taylor and Yeeles 2019; Devbhandari and Remus 2020). The observation that sister replication forks emanating from the same replication origin progress at highly similar rates suggests that Pol ε is capable of continuous DNA synthesis through a complex chromatin environment (Claussin et al. 2022). Robust synthesis by Pol ε stems from intrinsic attributes and the assistance of extrinsic factors. For example, multiple replication-promoting factors can remove template barriers that impede Pol ε (Hizume and Araki 2019). In addition, unlike Pol δ, the catalytic subunit of Pol ε binds directly to the replicative helicase, which encircles the leading strand template, thus stabilizing Pol ε-template association (Langston et al. 2014; Goswami et al. 2018).

Besides being tethered to the replicative helicase, Pol ε has additional features that may contribute to the enzyme's intrinsic ability to continuously synthesize DNA. In particular, the catalytic core of Pol ε possesses several conserved domains not found in Pol δ or other DNA

Chapter 3 is adapted from Meng X[*], Claussin C[*], Regan-Mochrie G[*], Whitehouse I[#], Zhao X[#]. Balancing act of a leading strand DNA polymerase specific domain and its exonuclease domain promotes genome-wide replication fork symmetry (*submitted*)

polymerases (Hogg et al. 2014). Thus far, only two of these domains were examined in detail using the yeast catalytic subunit of Pol ε, Pol2, as the prototype of this family of enzymes. In one study, it was shown the P-domain of Pol2 enhances the Pol ε-template association and processive leading strand synthesis (Hogg et al. 2014). Recently, we charactered another Pol ε family domain, referred to as POPS (POl2 family-specific catalytic core Peripheral Subdomain), which is located at the periphery of the Pol2 catalytic (CAT) core (Meng et al. 2020) (**Figure 3.2A**). This sixty-eight-residue region shows a high degree of sequence homology between the yeast Pol2 and its human counterpart POLE (Meng et al. 2020). Modeling three recurrent cancer-associated mutations found in this region of the POLE enzyme in yeast generated the *pol2-R567C, E611K, L621F* (*pol2-REL*) allele (**Figure 3.2A**), which exhibited gross chromosomal re-arrangement (GCR) but normal mutation rates, distinguishing POPS domain mutations from other POLE cancer mutants that cause hyper-mutation (Meng et al. 2020). Replication progression defects were seen for *pol2-REL* cells and the purified Pol ε^{REL} complex reduced DNA synthesis *in vitro,* showing that POPS plays an important role in leading strand synthesis (Meng et al. 2020). However, how POPS contributes to Pol2-mediated replication elongation at the molecular level is unclear.

In this work, we addressed the molecular function of the POPS domain, aiming to gain insight into the intrinsic Pol ε features that promote processive DNA synthesis. To this end, we looked for suppression mutations that could rescue *pol2-REL* defects. Surprisingly, we found that abolishing the activity of the Pol2 exonuclease (EXO) domain improved the growth of *pol2-REL* cells. The suppression caused improvement in S phase progression and genomic stability as well as reduced reliance on other genome maintenance pathways of *pol2-REL* cells. Single-molecule analysis using replicon-seq demonstrated that *pol2-REL* led to genome-wide fork slowing or stalling resulting in asymmetric progression of sister replication forks. Importantly, replication fork progression was significantly improved by abolishing the EXO activity of Pol2.

Though the Pol2 EXO domain is best known for editing mis-incorporated nucleotides (Pursell and Kunkel 2008), our data provide evidence for another role for this domain in reducing fork movement and that such role is balanced by POPS to permit processive leading strand synthesis and reduce the risk of chromosomal re-arrangements.

3.2 Results

3.2.1 Slow growth of a POPS mutant was rescued by abolishing the Pol2 exonuclease activity.

Mutation of the highly conserved Pol ε–specific POPS domain results in slow growth and defective DNA replication (**Figure 3.1A**) (Meng et al. 2020). We sought to address whether the slow growth of the POPS mutant allele, *pol2-REL*, could be suppressed by mutating other Pol2 regions implicated in the negative regulation of the enzyme.

One recent study showed that mutating the PCNA binding motif (PIPm, F1119A, F1200A) in Pol2 rescued the poor growth of cells carrying a *pol2* allele lacking its polymerase domain, presumably because this mutation allowed Pol δ to replace Pol ε for leading strand synthesis (Devbhandari and Remus 2020). However, PIPm did not improve the growth of *pol2-REL* at its semi-permissive temperature (34ºC), rather it worsened *pol2-REL* growth, suggesting that when POPS was defective, weakening PCNA binding further reduced DNA synthesis (Fig. 1A). We confirmed that the slow growth of *pol2-REL* or *pol2-REL-PIPm* mutant cells was not caused by abnormal Pol2 protein levels (**Figure 3.2B**). As PCNA binding enhances Pol ε processivity (Chilkova et al. 2007), the negative genetic interaction between REL and PIPm mutations may suggest that POPS could also be involved processive DNA synthesis.

Another study found that abolishing the Pol2 exonuclease activity by mutating two catalytic residues (EXOcd, as in the *pol2-4* allele) improved the growth of a *pol2* mutant that is defective in DNA damage checkpoint mediated phosphorylation at a residue within its EXO

domain (Pellicano et al. 2021). Though *pol2-REL* did not contain mutations in the EXO domain, its growth defect was also suppressed by the EXOcd mutation as spore clones of *pol2-EXOcd-REL* grew better than those of *pol2-REL* (**Figure 3.1A-3.1B**). We confirmed that *pol2-EXOcd-REL* supported normal protein level (**Figure 3.2B**). Thus, we concluded that abolishing the exonuclease activity of Pol2 can improve the growth of a POPS mutant.

The exonuclease domain of Pol2 is best studied for its proofreading ability, which corrects mis-incorporated nucleotides at the replication fork (Pursell and Kunkel 2008). However, a recent study suggested that the EXO domain can promote nascent strand degradation under conditions of replication stress, similar to the effect exerted by the Exo1 nuclease (Pellicano et al. 2021). Given this similarity, we explored if *exo1*∆ could also suppress the *pol2-REL* growth defect. In contrast to the lack of exonuclease activity of Pol2, Exo1 loss did not improve *pol2-REL* growth, nor did it affect the growth of *pol2-EXOcd-REL* cells (**Figure 3.2C**). Thus, the suppression of *pol2-REL* was specifically achieved by the EXO domain mutation and not by *exo1*∆*,* and so was unlike the situation reported in replication stress (Pellicano et al. 2021). Our findings indicated a direct functional interplay between POPS and EXO domains. Since *pol2-REL* exhibited a wild-type level of mutation rates (Meng et al. 2020), such functional interplay is unlikely related the mutation editing function of Pol2's EXO. Rather this can be best explained by a role for the EXO domain in nascent strand degradation or polymerase backtracking in the absence of mis-incorporated nucleoside even during normal growth conditions.

3.2.2 EXO inactivation reduced pol2-REL genome instability and reliance on other genome factors

Next, we addressed whether better growth of *pol2-EXOcd-REL* compared with *pol2-REL* correlated with improvement of replication-associated defects of *pol2-REL*. The growth of cells suffering from replisome deficiency increases reliance on replication-promoting factors, such as

the Rrm3 helicase that removes protein barriers or the RNase H2 nuclease that removes RNA-DNA hybrids from DNA (Hizume and Araki 2019). Consistent with compromised replication in *pol2-REL* cells, their growth strongly relies on the presence Rrm3 and the Rhn201 subunit of RNase H2 (Meng et al. 2020). Significantly, inactivating the Pol2 exonuclease improved the growth of *pol2-REL* in the absence of either *rrm3* or *rhn201* (**Figure 3.1C**). Further, simply overexpressing the gene encoding the RNase H1 protein capable of removing RNA-DNA hybrids, improved the growth of *pol2-REL* (**Figure 3.1D**). Taken together, these data provided evidence that *pol2-REL* has a reduced ability to replicate past template barriers, such as RNA-DNA hybrids, and this deficiency can be improved by inactivation of the Pol2 exonuclease.

Another hallmark of defective DNA replication is a reliance on recombinational repair to fill un-replicated gaps, and on the DNA damage checkpoint that allows more time for replication. We found that *pol2-REL* indeed showed synthetic sickness with mutations of the key recombination repair factor Rad52 and the DNA checkpoint mediator Rad9 (**Figure 3.1C**; **Figure 3.4A**). Significantly, both negative genetic interactions were partly rescued by inactivating the exonuclease activity of Pol2 (**Figure 3.1C**; **Figure 3.4A**).

A third hallmark of impaired genome replication examined here is increased level of GCR. We have previously shown that *pol2-EXOcd* did not change GCR rate as compared with wild-type (Meng et al. 2020). However, the *pol2-EXOcd* was able to significantly reduce the elevated GCR levels seen in *pol2-REL* cells (**Figure 3.1E**). Collectively, our genetic analysis demonstrates that improving *pol2-REL* growth by Pol2 exonuclease inactivation correlates with enhanced genome replication and stability.

3.2.3 Abolishing Pol2 exonuclease activity improved genome replication of a POPS mutant that showed normal Mrc1 and Csm3 association with the replicative helicase

Next, we directly examined whether eliminating the exonuclease activity of Pol2 could correct genome-wide replication defects in *pol2-REL* cells. Flow cytometry analysis of G1-

synchronized cells released into S phase at the semi-permissive temperature confirmed slower DNA replication of *pol2-REL* cells (**Figure 3.3A**) (Meng et al. 2020). While wild-type cells mostly completed S phase 60 min after release from G1 arrest, most *pol2-REL* cells remained in S phase. In contrast to *pol2-REL* cells, DNA synthesis in *pol2-EXOcd-REL* cells was far more advanced at 50 and 60 minutes, suggesting faster genome replication (**Figure 3.3A**, **red arrows**; **Figure 3.4B**).

Replication speed can be affected by the association of replication progression factors, Mrc1 and Csm3-Tof1, with the CMG replicative helicase, which is formed upon the association of the GINS complex and Cdc45 to the MCM helicase (Yeeles et al. 2017). We thus examined the association of Mrc1 and Csm3 with the GINS complex member Psf1 by co-immunoprecipitation. Since *pol2-REL* and wild-type cells progressed through S phase at different rates (**Figure 3.3A**) (Meng et al. 2020), averaging three S phase time points reflected the overall S phase situation of each genotype. Using this method, we found that the amounts of Mrc1 and Csm3 co-immunoprecipitated with the MCM-associated Psf1 were not reduced in *pol2-REL* compared with wild-type cells (**Figure 3.4B-3.42C**; **Figure 3.5**). This observation provided evidence that the replication elongation defect caused by *pol2-REL* was not due to lack of physical association of fork progression promoting factors with the CMG complex.

3.2.4 Pol2 exonuclease activity partly accounts for genome-wide sister fork asymmetry and replisome progression defect of a POPS mutant.

Our data thus far suggested that reduced DNA synthesis caused by the Pol2 POPS mutation could be partly due to Pol2 exonuclease activity impairing replisome movement. To test this model more directly, we utilized the Replicon-seq method to derive genome-wide, high-resolution maps of the movement of sister replication forks in *pol2-REL* and *pol2-EXOcd-REL* compared to wild-type data (Claussin et al. 2022) (**Figure 3.7A**). Sister replication forks move at

a constant rate through tens of kilobases of chromatin in wild-type cells, which gave rise to a highly symmetric “V” shaped pattern when replicons were plotted according to their position and length (**Figure 3.6**). By contrast, replicons in *pol2-REL* mutants showed extensive asymmetry, suggesting that sister replication forks moved at unequal rates due to frequent replisome stalling (**Figure 3.6**). Such defects appeared to be genome-wide, seen at both known fork barriers sites, such as tRNA genes, centromeres and highly transcribed genes, and non-barrier sites (**Figure 3.7B, 3.9A**). Significantly, progression of sister replisomes in *pol2-EXOcd-REL* was closer to the wild-type pattern than that of *pol2-REL* across the genome (**Figure 3.6**). We note that read frequencies across a large range of read lengths were comparable between *pol2-EXOcd-REL* and *pol2-REL* datasets (**Figure 3.7C**).

A quantitative analysis of replicons across the genome was performed by first aligning the mid-points of replicons to their respective ARS Consensus Sequence (ACS), replicons were then grouped by length (Claussin et al. 2022). In all groups, read midpoint in wild-type cells aligned closely with the ACS, indicating that the left and right traveling sister replication forks and replisomes progressed at similar rates (**Figure 3.8A, Top**). In *pol2-REL* cells, the midpoints of the reads were broadly scattered, suggesting that the two replisomes were not progressing symmetrically from their origin (**Figure 3.8A, Middle**). This asymmetry was partially corrected in *pol2-EXOcd-REL* cells (**Figure 3.8A, Bottom**). Calculation of the standard deviation of the data for different length replicons confirmed that the *pol2-EXOcd-REL* partially rescues the broad increase in standard deviation seen in *pol2-REL* cells (**Figure 3.8B**). This suppression was also visible when data were plotted in violin plots (**Figure 3.9B**). Collectively, these results indicated that a POPS mutation impeded replisome progression across all lengths of replicons, and this was partially corrected by disabling the Pol2 exonuclease activity.

Replicon-seq data can also inform the coordination of leading and lagging strand synthesis at the replication forks. In wild-type cells, leading strand is typically ~300-400 nt longer than the contiguous lagging strands (**Figure 3.8C**) (Claussin et al. 2022). However, in *pol2-REL*

cells, the two strands were more closely aligned and differed in position by ~30nt. Significantly, inactivating Pol2 exonuclease activity in *pol2-REL* cells increased the relative length of the leading strand compared to the lagging to ~80nt, which more closely resembles the wild-type pattern (Fig. 4C). This finding suggests that the POPS mutation interferes with leading strand synthesis likely by limiting Pol ε processivity.

3.3 Discussion

3.3.1 The division of labor between the leading and lagging strand polymerases

The division of labor between the leading and lagging strand polymerases during genome duplication is fundamentally important for the stable inheritance of genetic information. The striking difference in the length of nascent DNA synthesized by the two polymerases argues that they must adopt distinct strategies optimized for their specific tasks. In particular, the leading strand polymerase Pol ε faces the challenge of processive DNA synthesis over long stretch of template DNA that contains frequent barriers. Previous studies suggest that this challenge can be mitigated by several unique features of Pol ε that foster its association with template (Sengupta et al. 2013; Hogg et al. 2014; Langston et al. 2014; Goswami et al. 2018). Here we reveal a different strategy that also contributes to Pol ε mediated leading strand synthesis. Our Replicon-seq data suggest that POPS aids processive synthesis at least partly by countering the Pol2 exonuclease function. This conclusion is further supported as replication-related defects caused by POPS mutation were improved by EXO inactivation, including slow growth and S phase progression, increased GCR rates, and synthetic sickness with mutants affecting replication promoting factors. Collectively, these data provide evidence for a functional antagonism between the POPs and EXO domains.

The antagonism between exonucleolytic degradation and polymerization is unlikely explained by the canonical role of the Pol2 EXO domain in mutation correction, since *pol2-REL*

showed a wild-type level of mutation rates (Meng et al. 2020). Rather, it suggests that EXO-mediated polymerase backtracking impedes the forward movement of the enzyme when POPS is defective (**Figure 3.10**). How the POPS domain counterbalances such a role to favor forward polymerase movement is currently unclear. Our data suggest that it is unlikely that POPS tethers Mrc1 and Csm3 to the replisome or that it promotes PCNA binding. Given its peripheral localization (**Figure 3.2A**), one possibility is that POPS may enable a stable conformation that favors continued synthesis such as by blocking the transfer of the nascent primer to the EXO domain in the absence of mis-incorporated nucleotide. This model agrees with recent evidence supporting the importance of proper regulation of partition of the nascent primer between the EXO and CAT domains (Parkash et al. 2019; Xing et al. 2019). The role proposed by POPS may be particularly useful when the polymerase is paused by template barriers. Indeed, we found that *pol2-REL* growth was improved by reducing template barriers via overexpression of the Rhn1 protein. In addition, *pol2-REL* growth was worsened when the burden of template barriers was increased. Given that individual mutations studied in *pol2-REL* correspond to recurrent cancer-associated mutations in POLE and exhibited replication defects (Meng et al. 2020), the synthetic sickness profile uncovered for *pol2-REL*, such as with mutations inactivating checkpoint and repair proteins, may be able to inform strategies of selective killing cancer cells bearing POPS mutations.

3.4 Materials and Methods

Yeast strains, procedures, and cell cycle experiments. Yeast strains are listed in Supplementary Table 1. At least two strains per genotype were examined in each experiment. Standard PCR-based methods were used to generate integrated alleles and add tags to proteins at endogenous loci, followed by DNA sequencing verification. Primer information is available upon request. Cell cycle arrest and release were conducted following standard

protocol and described in detail in the supplemental materials.

Cell cycle arrest and release: Standard protocol was used for these experiments. In brief, log-phase cultures were treated with alpha-factor (5 μg ml^{-1}) until >90% of cells exhibited G1 arrest. For experiments involving temperature shift, G1-arrested cells were shifted to 34 °C for 1 h. 300 μg ml^{-1} pronase (Millipore) was then added into the G1-arrested culture to release the cells. Samples were collected at multiple time points for examination. Flow cytometry analyses were performed using a standard procedure. Briefly, ethanol fixed yeast cells were washed with and resuspended in sodium citrate solution. RNase and Proteinase K were added sequentially to remove RNAs and proteins. Sytox green was then used to stain DNA. Flow cytometry was performed using BD LSRII flow cytometer, and data were analyzed with the FlowJo software.

Co-immunoprecipitation experiments and Western blotting. Standard protocols were followed, G1 and S phase cells were harvested. Cells were disrupted by glass bead beating in lysis buffer, and Benzonase was added to digest nucleic acid before centrifugation for 30 min at 20,000 x *g* to obtain whole-cell extract (WCE). The lysis included 50 mM HEPES-KOH pH 7.5, 100 mM KOAc 1% TritonX-100, 2 mM MgOAc, 2 mM NaF, 2 mM beta-glycerophosphate, 10 mM beta-mercaptoethanol, and cOmplete™ Ultra EDTA free protease inhibitor (Roche). WCE was incubated with pre-washed anti-flag beads (A2220, Sigma-Aldrich), for 2 h at 4 °C. After washing the beads, bead-bound proteins were eluted with 2x Laemmli buffer without DTT. Proteins were boiled for 5 min before subjected to SDS-PAGE on 4-20% gradient gels (Bio-Rad) and transferred to nitrocellulose membrane (GE healthcare) for Western blotting. Antibodies used in probing western blots include anti-Psf1 (gift from K. Labib), anti-Flag (Sigma-Aldrich, F1804), anti-Mrc1 (gift from G. De Piccoli), anti-Csm3 (gift from K. Labib), and anti-HA (3F10).

GCR and mutation rate assays. GCR assays were performed using a standard protocol and rates were calculated (Myung et al. 2004; Putnam and Kolodner 2010). To ensure all experiments are in the same genetic background, we moved the GCR assay to W303 background (Wan et al. 2019). For each genotype, at least seven cultures were examined. Cells were plated on SC + 5-FOA + Can (FC) and SC plates to obtain colony numbers that lose the URA3-CAN1 cassette and total viable colonies, respectively. GCR rates were calculated as m/NT, wherein m (1.24 + ln[m]) – NFC = 0. m: mutational events, NFC: number of colonies on FC plates, NT: colonies formed on SC plates. The upper and lower 95% confidence intervals were then derived. The URA3-CAN1 cassette is inserted at YEL068c in GCR assay.

Replicon-seq data acquisition and analyses. Cells were grown to OD_{600}=0.35 in YPD, arrested in G1 with alpha factor for 150 minutes at 25ºC. 30 minutes before the end of the G1 arrest, BrdU was added to 400ug/mL. Cells were washed 3 times with 30ºC pre-warmed YPD media, and released from G1 arrest at 30ºC in YPD containing 400ug/mL of BrdU. DNA preparation for Replicon-seq was performed as described in Claussin et al, 2022. Briefly, cells were premetallized with digitonin in Zentner buffer A: 15mM of Tris-HCl pH 7.5, 80mM KCl, 0.1mM EGTA, 0.5mM spermidine, 0.2mM spermine and ½ of pierce mini-protease inhibitor EGTA-free, for 5 minutes at 30ºC. Shifted to an ice-cold water bath for 1 minute. Calcium was added to final concentration of 2mM for 10 seconds. The reaction was stopped with 400mM NaCl, 20mM EDTA and 8mM EGTA buffer. Cells were spheroplasted with 5 mg of Zymolyase 100T for 5 minutes at 30ºC. Spheroplasts were washed once with spheroplast buffer (1M sorbitol, 1mM EGTA, 10mM β-mercaptoethanol, and 50mM Tris-HCl pH 7.5) and twice with RNAse buffer (4mM EGTA, 50mM NaCl, 50mM Tris-HCl pH 8.5, 10mM DTT). 10uL of RNAse cocktail was added to the cells and incubated at 37ºC for 2 hours. Reaction was incubated at 55ºC for 1 hour with the addition of 50mg of chelex beads, 10uL of 20mg/mL of proteinase K, 0.5% SDS, 1% Sarkozyl. DNA was extracted with phenol chloroform and ethanol precipitated.

DNA pellet was resuspended in 50uL of ultra-pure water. Two biological replicates were performed for each mutant and sequenced using Oxford Nanopore Minion flow cells. Library preparation was performed as described in Claussin et al, 2022 using SQK-LSK-109 kit. 700ng to 1.2ug of DNA library was loaded into the Nanopore MinION FLO-MIN106 flow cell R9 and sequenced for over 48h. Replicates were inspected for consistency and then combined to increase for read number for analysis shown. BrdU positive reads were called using DNAscent V2 and mapped to a reference assembled yeast genome (Claussin et al. 2022). Only BrdU reads with a score >=0.2 were selected for the analysis. Data was plotted as described in Claussin et al, 2022.

3.5 Acknowledgments

We thank Drs. Erik Johannsson and Dirk Remus for helpful discussion X.L is supported by NIGMS grants R01GM131058 and R35GM145260. I. W. is supported by R01RGM102253 and R01GM129058. CC is supported by a Francois Wallace Monahan Fellowship.

Author contributions

All authors were involved in research design and data analyses, Xiangzhou Meng, Gemma Regan-Mochrie, and Clémence Claussin performed research. Gemma Regan-Mochrie, Iestyn Whitehouse, and Xiaolan Zhao wrote the paper with all authors' input. Specific contributions to individual figure panels are noted below each legend.

3.5 Figures

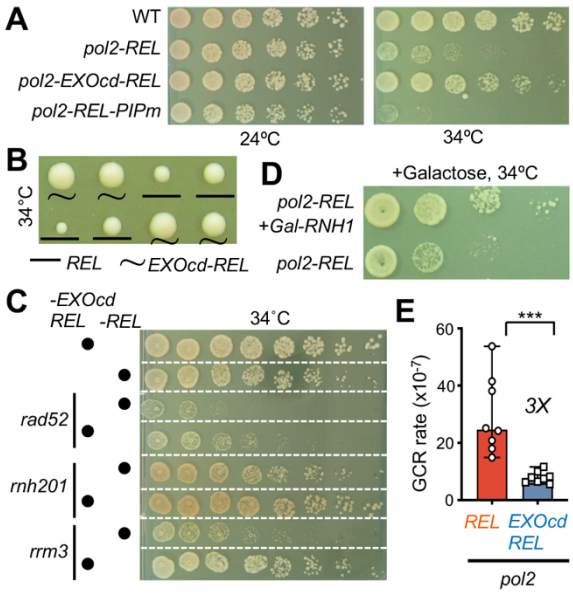


Figure 3.1 Inactivating the Pol2 exonuclease suppressed the growth defect and genomic instability of a POPS mutant.
A. *The growth of three different pol2 mutant strains. In each, the endogenous POL2 locus was replaced by the indicated pol2 mutation. While pol2-REL cells grew poorly at semi-permissive temperate (34˚C), further mutating two catalytic residues of Pol2 exonuclease domain (D290A/E292A) led to better growth. In contrast, mutation the PCNA binding motif (PIPm) worsened growth of pol2-REL cells. Cells were examined in three-fold series of dilution on plates.* ***B.*** *The growth of pol2-EXOcd-REL verse pol2-REL spore clones on dissection plates. The diploid strain carrying both copies of the indicated pol2 mutants were sporulated and dissected. Two representative tetrads are shown.* ***C.*** *Inactivating the Pol2 exonuclease rescued the sensitivity pol2-REL cells to the loss of the genes encoding for the recombinational repair factor Rad52, the RNA-DNA hybrid resolving protein Rnh201, or the Rrm3 helicase that removes protein barriers from the template DNA. Cells were examined in three-fold series of dilution on plates.* ***D.*** *The growth of the pol2-REL mutant cells was improved by overexpressing the Rnh1 protein that can resolve the RNA-DNA hybrids. Rnh1 overexpression was driven by the galactose promoter.*

Panels ***a, b, d,*** *and* ***e*** *were generated by the Xiangzhou Meng. Panel* ***c*** *was generated by the author of this thesis.*

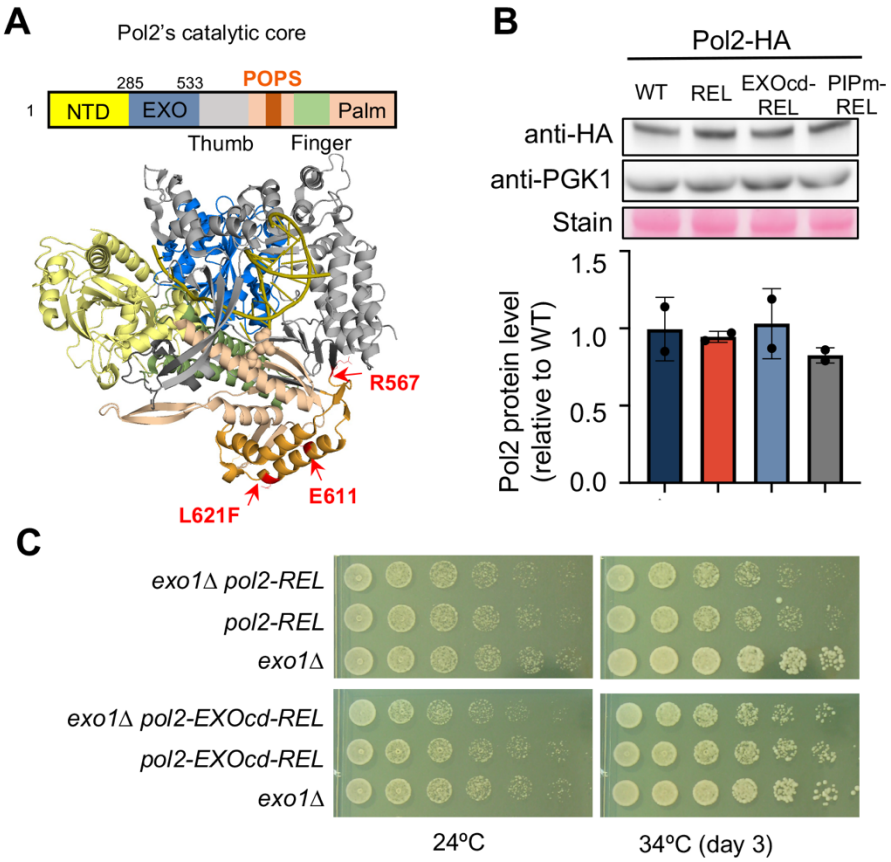


Figure 3.2 Examination of protein levels and growth of pol2 mutants.

A. *Top: Schematics of the domain organization of the Pol2 the catalytic core located in the N-terminal half of the protein. These include the Pol2 N-terminal domain (NTD), its exonuclease domain (EXO), three canonical catalytic sub-domains (Thumb, Finger and Palm), and the POPS. Bottom: Crystal structure of the Pol2 catalytic core regions with each domains marked and colored as shown in the schematic above (Hogg 2014). The three conserved residues mutated in the POPS to construct the pol2-REL allele, namely R567, L621F, and E611, are labeled.* ***B****. The examined pol2 mutations did not alter protein levels. Immunoblots of asynchronously cycling cells (top) and quantification of Pol2 levels in two replicates (bottom). PGK1 and stain were used as loading controls.* ***C****. Removal of the Exo1 exonuclease did not improve the growth of pol2-REL cells or that of pol2-EXOcd-REL cells. Cells were examined in three-fold series of dilution on plates.*

Panels ***a, and c*** *were generated by the Xiangzhou Meng. Panel* ***b*** *was generated by the author of this thesis.*

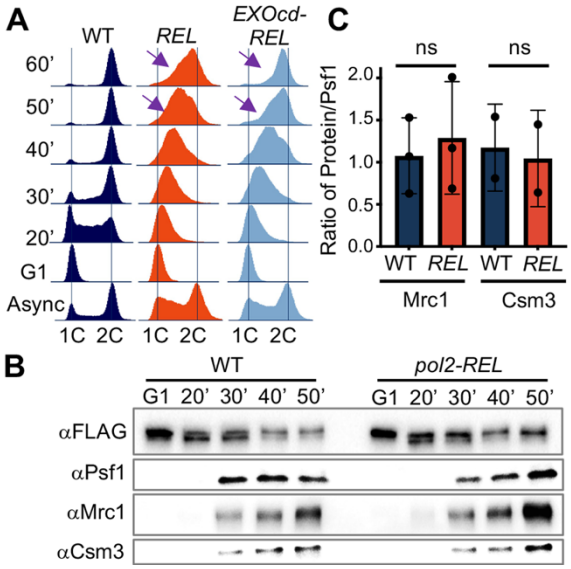


Figure 3.3 pol2-REL replication defect was improved by inactivating the Pol2 exonuclease and was not associated with reduced CMG-bound Mrc1 and Csm3.

A. *Flow cytometry analysis of cell cycle progression of WT, pol2-REL, and pol2-EXOcd-REL cells at 34˚C. Cells were arrested in G1 phase by alpha-factor treatment and then released into the cell cycle. Arrows highlight the different DNA contents between pol2-REL and pol2-EXOcd-REL at 60 min after G1 release.* ***B.*** *CMG-bound Mrc1 or Csm3 examined in pol2-REL and wild-type cells. Cells were synchronized in G1 and then released into S phase at 24ºC. At each examined timepoint, the Mcm4-Flag protein was immunoprecipitated. Co-immunoprecipitation of indicated proteins were examined by immunoblots using indicated antibodies.* ***C.*** *The quantification of Mrc1 and Csm3 levels relative to Mcm4-bound Psf1 based on the co-immunoprecipitation experiments shown in panel B. Average values and standard deviations of two to three biological replicates are shown. NS: not statistically significant based on two-tailed student t-test.*

Panels ***a, b and c*** *were generated by the Xiangzhou Meng, with assistance from the author of this thesis for panel a.*

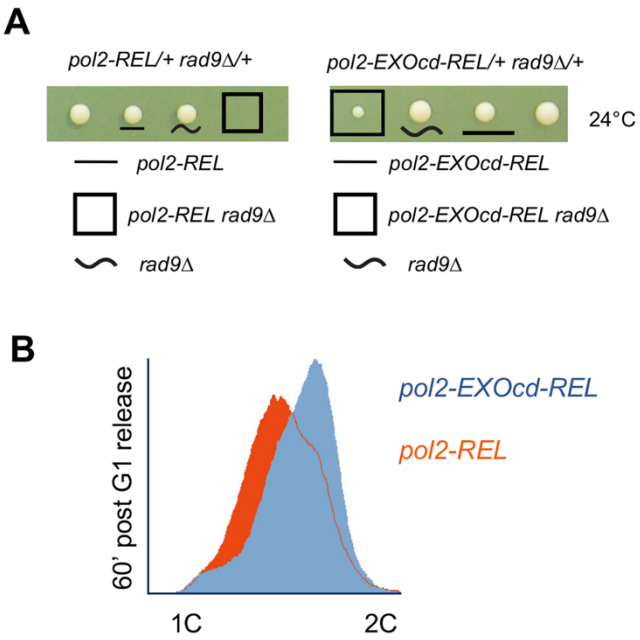


Figure 3.4 Examination of a genetic interaction and the FACS profile of pol2 mutants
A. *pol2-REL was synthetical lethal when combined with the loss of the DNA damage checkpoint mediator Rad9, and this was rescued by inactivating the Pol2 exonuclease.* ***B.*** *Overlay of FACS data between the pol2-REL and pol2-EXOcd-REL cells taken 60 minutes after release from G1, as shown in Figure 2A.*

Panels ***a*** *and* ***b*** *were generated by the Xiangzhou Meng and the author of this thesis*

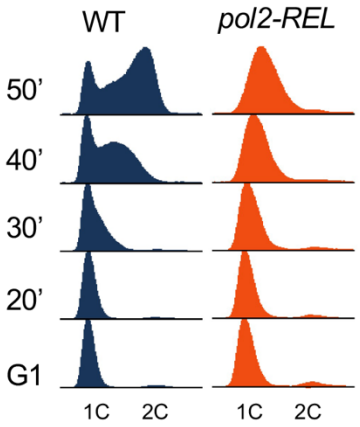


Figure 3.5 FACS profiles for wild-type and pol2-REL cells examined for Mrc1 and Csm3.

Wild-type and pol2-REL cells were grown at 24ºC, the permissive temperature of pol2-REL. Cells were arrested in G1 phase by alpha factor treatment and then released into S phase after pronase digestion of alpha factor. Samples from the indicated time points were examined by FACS analyses as shown here as well as by immunoblotting as shown in Figure 3.3.

This figure was generated by the Xiangzhou Meng with assistance from the author of this thesis.

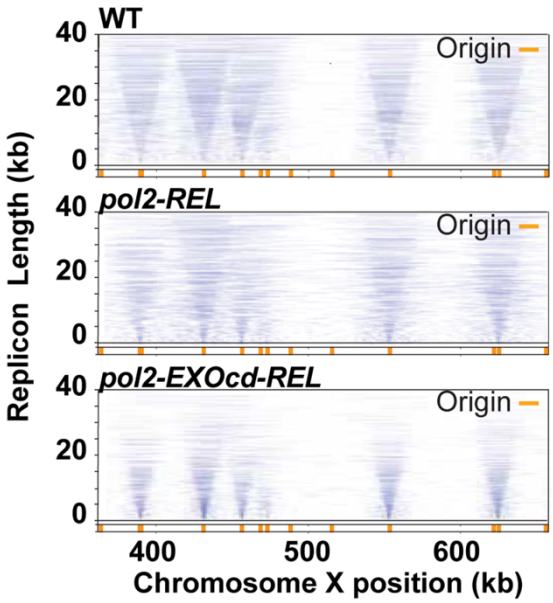


Figure 3.3.6 Replicon-seq analyses revealed genome-wide sister fork asymmetry of pol2-REL which was improved by EXO inactivation.
Examples of DNA replication mapped by replicon-seq for indicated strains are shown. The full length of BrdU containing sequencing reads based on unfiltered data are plotted as lines according to their position (x-axis) and length (y-axis) for Chromosome X. Top, middle, and bottom panels show data from WT, pol2-REL and pol2-EXOcd-REL cells, respectively. Replication origins are indicated by orange bars.

This figure was generated by Clémence Claussin and Iestyn Whitehouse

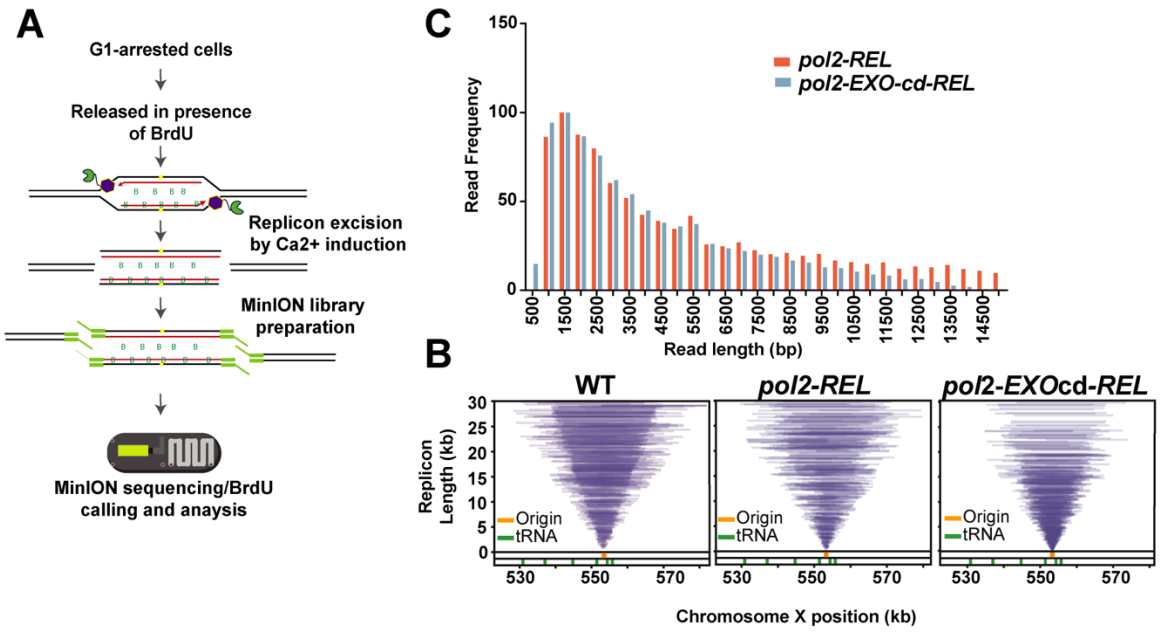


Figure 3.7 Analysis of DNA replication with Replicon-seq.
A. *Schematic of the experimental workflow for Replicon-seq. Mcm4-Mnase is represented by the purple hexagon. Two sister replication forks move from an origin of replication (yellow square) in the presence of BrdU, and synthesize a replicon. Upon calcium addition, the DNA is cleaved at the replisome (Mcm4-Mnase position) and replicons are excised from the genome. DNA fragments are processed to allow nanopore adaptor ligation and sequencing.* ***B.*** *Frequency plot of reads for pol2-REL and pol2-EXOcd-REL mutants for reads binned at different lengths.* ***C.*** *Tornado plots from replicon-seq experiments for WT, pol2-REL and pol2-EXOcd-REL mutants focusing on regions around ARS1018. Only reads with a BrdU score ≥ 0.2 and overlapping the selected ARS are shown. ARS (Origin) positions are shown in the lower part of the graph in orange, t-RNAs are green.*

This figure was generated by Clémence Claussin and Iestyn Whitehouse

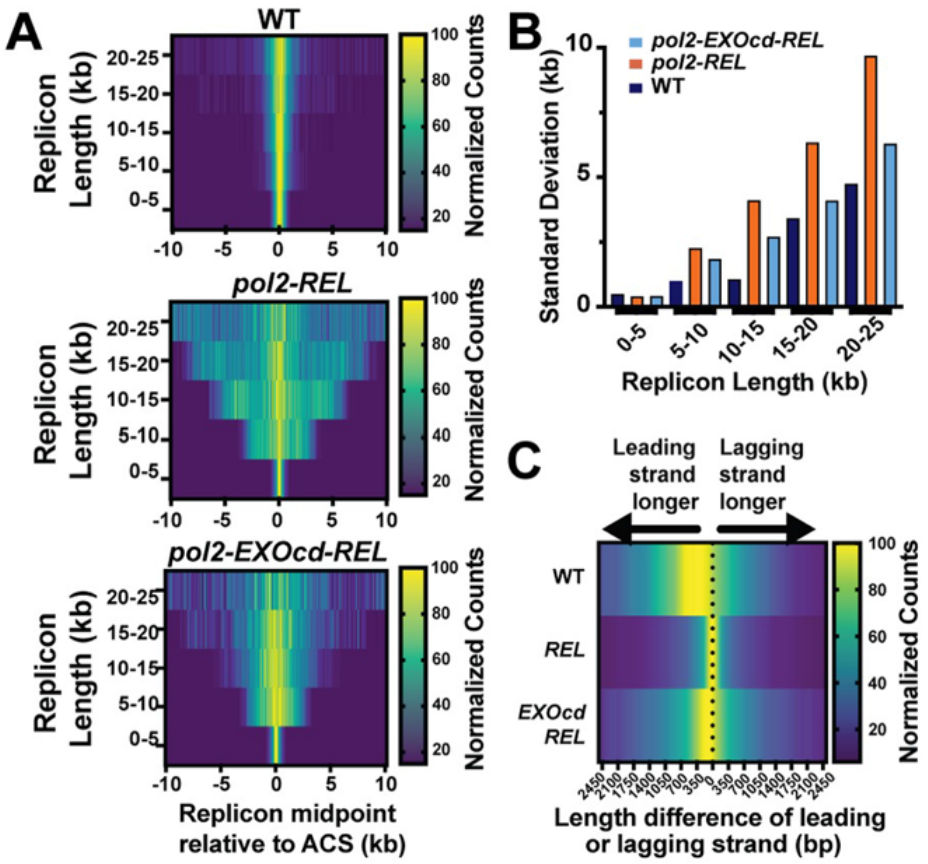


Figure 3.8 Sister replication fork asymmetry and leading strand DNA synthesis in pol2 mutants

A. *Genome wide analysis of sister replication fork progression from replication origins. All replicon-seq data was filtered to select BrdU-containing reads that overlap with defined ACS sequences (Eaton et al. 2010). Replicons were binned according to their length in 5 kb increments and total read counts in each bin were normalized to allow comparison between bins. The positions of the midpoint of selected reads were plotted relative to the position of the ACS. The data is shown as a heatmap. Top, middle and bottom panels show data from WT, pol2-REL, and pol2-EXOcd-REL cells, respectively.* ***B.*** *Standard deviations of the data shown in panel A was calculated by fitting gaussian distribution. Data from WT is shown as red, pol2-REL is green and pol2-EXOcd-REL is blue. Data were binned according to different sequencing read lengths indicated on x-axis.* ***C.*** *The relative differences in the lengths of leading vs lagging strands is shown as a heatmap for WT, pol2-REL and pol2-EXOcd-REL. The read orientation and the difference in position between the read midpoint and the ACS were used to define the fraction of each read that was synthesized by the leading or lagging strand polymerases. The relative difference in leading or lagging strand length was plotted on the x axis; 0 indicates no difference, whereas signal to the left or right of 0 indicates that the leading strand or lagging strand is longer, respectively.*

This figure was generated by Clémence Claussin and Iestyn Whitehouse

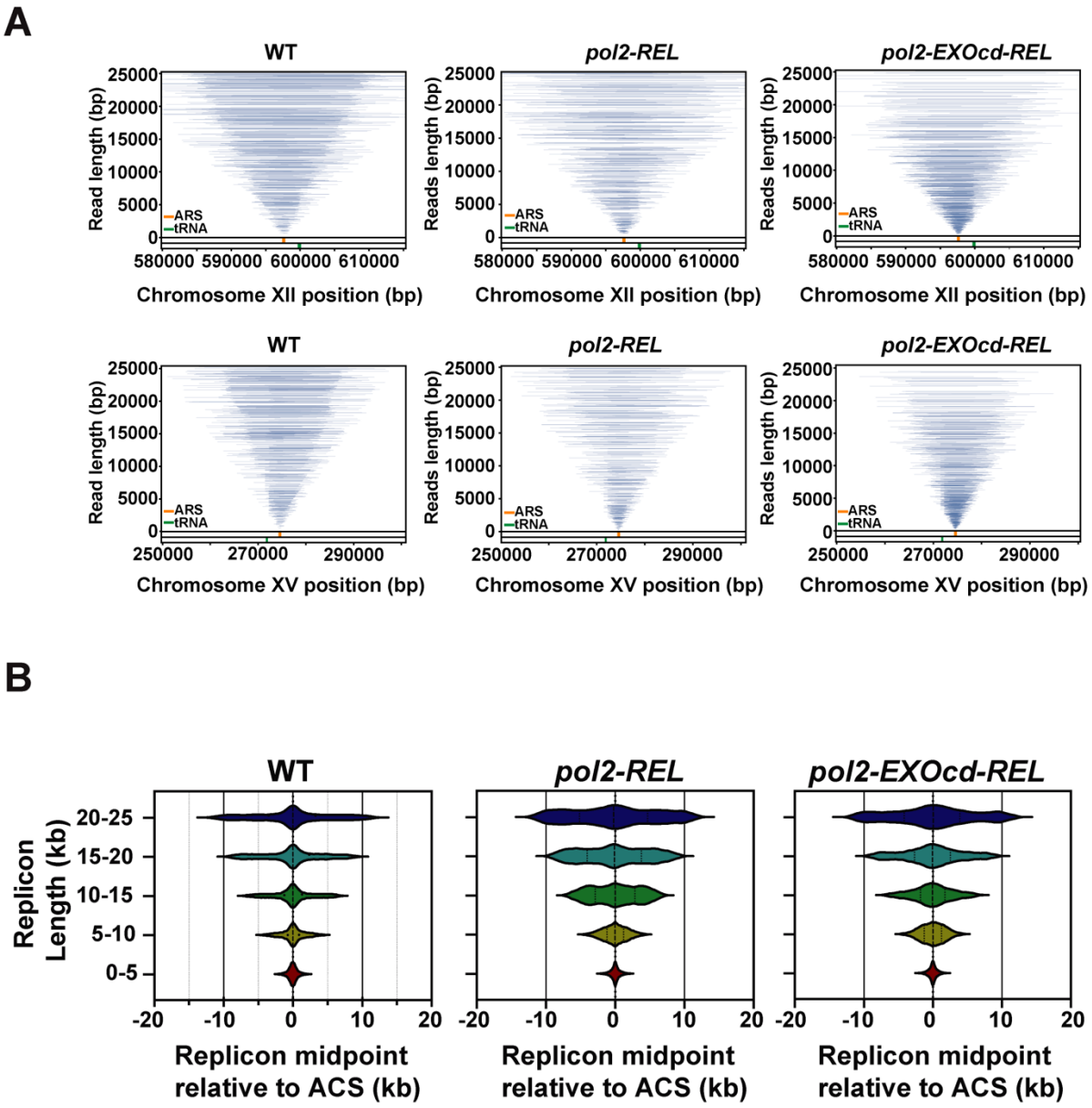


Figure 3.9 Analysis of sister replisome progression with Replicon-seq.
A. *Tornado plots of ARS1218 and ARS1511 from replicon-seq experiments for WT, pol2-REL and pol2-EXOcd-REL mutants. Only reads with a BrdU score ≥ 0.2 and overlapping the selected ARS are shown. ARS positions are shown in the lower part of the graph in orange, t-RNAs are green. Violin plots of the positions of read midpoints relative to the nearest ACS. Only reads overlapping an origin with an annotated ACS were analyzed. Reads were grouped according to their length (shown on y-axis) and the distance between the read midpoint and the middle of the ACS was calculated.* ***B.*** *Violin plots of the positions of read midpoints relative to the nearest ACS. Only reads overlapping an origin with an annotated ACS were analyzed. Reads were grouped according to their length (shown on y-axis) and the distance between the read midpoint and the middle of the ACS was calculated.*

This figure was generated by Clemence Claussin and Iestyn Whitehouse

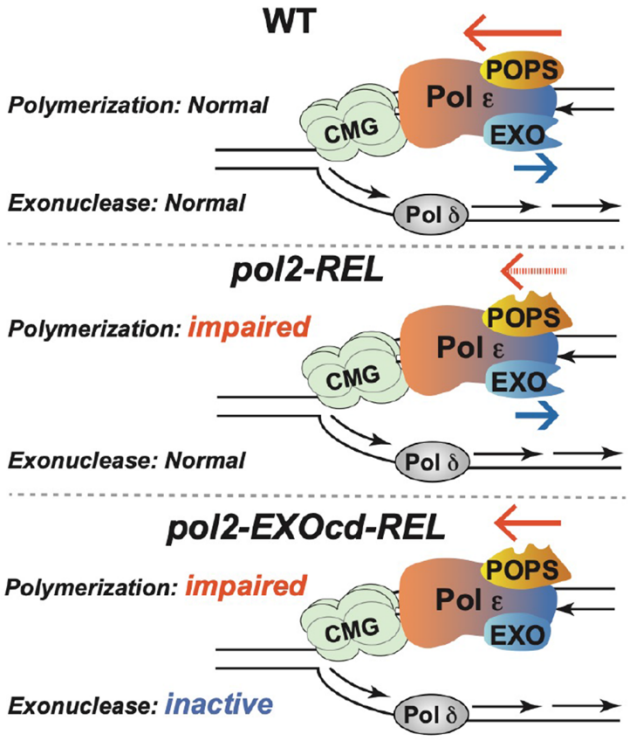


Figure 3.10 A model for the antagonistic action of the Pol2's POPS and EXO domains in leading strand synthesis.
During normal growth, processive leading strand synthesis is benefited from POPS counterbalance of EXO-mediated backtracking of the polymerase and/or nascent strand degradation. In POPS mutants, such as pol2-REL examined here, EXO-mediated backtracking can impair leading strand synthesis, generating sister fork asymmetry and multiple replication-associated defects. These defects can be rescued by inactivating Pol2's EXO function. See text for details.

This figure was generated by Xiaolan Zhao, Iestyn Whitehouse, and the author of this thesis

Table 2. Strains used in Chapter 3

Strain	Genotype
T1948-3a	*MATalpha POL2-3HA::KAN*
T1950-11a	*MATalpha pol2-R567C,E611K,L621F (REL)-3HA::KAN*
T2166-3d	*MATalpha pol2-REL-PIPm-3HA::KAN*
T2168-2b	*MATalpha pol2-EXOcd-REL-3HA::KAN*
T1950-11d	*MATa pol2-REL-3HA::KAN*
X8488-7c	*pol2-EXOcd-REL-3HA::KAN rrm3Δ::KAN*
X7992-1c	*pol2-REL-3HA::KAN rrm3::KAN*
X8489-5b	*pol2-EXOcd-REL-3HA::KAN rnh201Δ::KAN*
X7990-6a	*pol2-REL-3HA::KAN rnh201Δ::KAN*
X8490-1c	*pol2-EXOcd-REL-3HA::KAN Rad52Δ::TRP1*
X7260-3a	*pol2-REL-3HA::KAN Rad52Δ::TRP1*
X8351-5d	*pol2-REL-3HA::KAN pGal-CBP-RNH1:: LEU2*
X8351-8d	*pol2-REL-3HA::KAN*
X7776-5c	*ura3Δ can1Δ YEL068C::URA3-CAN1*
X8473-9a	*ura3Δ can1Δ YEL068C::URA3-CAN1pol2-EXOcd-REL::KAN*
X8370-1-15b	*exo1Δ::KAN*
X8370-1-3c	*pol2-EXOcd-REL-3HA::KAN exo1Δ::KAN*
X8446-1-11b	*pol2-REL-3HA::KAN exo1Δ::KAN*
X8679	*pol2-REL-3HA::KAN/+ rad9Δ::HIS3/+*
X8613	*pol2-exo-REL::KAN/+ rad9Δ::HIS3/+*
X7021-10d	*MATa POL2-3HA::KAN*
X7022-9a	*MATa pol2-REL-3HA::KAN*
T2168-11b	*MATa pol2-EXOcd-REL-3HA::KAN*
X7091-10a	*MATa POL2-3HA::KAN MCM4-3FLAG::KAN*
X7092-17c	*MATa pol2-REL-3HA::KAN MCM4-3FLAG::KAN*
CCY285-6	*MATa MCM4-N-ter-MNase LEU2::BrdU-Inc URA3::GPD-TK7*
X8676-1a	*MATa pol2-REL-3HA::Kan MCM4-N-ter-MNase LEU2::BrdU-Inc URA3::GPD-TK7*
X8677-22c	*MATa pol2-EXOcd-REL-3HA::Kan MCM4-N-ter-MNase LEU2::BrdU-Inc URA3::GPD-TK7*

CHAPTER 4 : CONCLUSIONS AND FUTURE DIRECTIONS

DNA replication is fundamental to life and mistakes made in any step in this process can lead to a severe reduction of fitness of cells or tumorigenesis in higher eukaryotes. Given its critical importance it must be carefully regulated. While we have a good understanding of many of the key players in this process from the past decades of study, especially in the budding yeast model system, their regulation and interplay has yet to be fully understood. In this thesis we sought to fill gaps in our knowledge of DNA replication by exploring a less well understood modification, sumoylation, of the origin recognition complex and investigating how a unique domain on Pol2 contributes to the balance between the polymerization and the exonuclease activity of the leading strand polymerase.

4.1 Summary

DNA replication begins with the origin recognition complex that can recruit the Mcm2-7 double hexamer, a helicase responsible for unwinding double stranded DNA to allow polymerases to synthesize new DNA strands. The careful timing and regulation of this process is crucial to ensure that the entire genome is replicated, and that this replication occurs at the right time. The temporal regulation of replication can be affected through modulation of origin activation during S-phase. In **Chapter 2** we found that the sumoylation of ORC helps to regulate origin firing timing. Specifically, sumoylation of Orc2 restrains early origin activation to allow for the proper activity of late origins and completion of late replicating regions. Replication, of course, entails the copying of the parental DNA by polymerases. In **Chapter 3** we examined the role for a highly conserved domain on the Pol2 subunit of Polε. Here we uncovered that mutation of three cancer-associated mutations in the POPS domain of Pol2 lead to increased sister replication fork asymmetry. This asymmetry could be partially rescued with the inclusion

of a Pol2 exonuclease mutation. This work suggests that a careful balance between polymerization and exonuclease activity, which is critical for leading strand synthesis and sister fork symmetry, is partially mediated by POPS.

4.2 ORC sumoylation – a new layer of control for origin usage

Previous studies have shown that SUMO regulates genome replication by targeting MCM and DDK (Wei and Zhao 2016b; Psakhye et al. 2019). This thesis reveals a new means by which SUMO limits replication initiation by preferentially affecting early origin usage through targeting the ORC complex. This sumo-based mechanism of origin firing regulation is in addition to known mechanisms that control the replication program. Multi-pronged strategies to regulate origin firing timing can render robustness to SUMO-based regulation and adding multiple regulatory points at different origins and under different growth conditions can fine tune this regulation.

For all the above SUMO substrates, sumoylated forms are only a small portion of the corresponding protein. This is consistent with almost all SUMO substrates examined so far and reflects the highly dynamic nature of the sumoylation and de-sumoylation cycles (Geiss-Friedlander and Melchior 2007; Zhao 2018). One general question is how a low level of overall detectable sumoylation at any given time can lead to a functional outcome?

We envision two non-mutually exclusive explanations for ORC sumoylation, which may be also generally applicable for other substrates. *First*, we suggest that selective sumoylation at a subset of genomic loci may exert a local effect, referred to as the “Biased Sumoylation” model (**Figure 4.1A**). It has previously been observed that the desumoylation enzyme responsible for most nuclear desumoylation is found at replication origins (Psakhye et al. 2019). Given that yeast ORC is associated with origins throughout the cell cycle, differential levels of the desumoylation enzyme at individual origins could permit only a subset of ORC complexes to

maintain relatively long-lasting sumoylation to affect replication. The selective effects of ORC sumoylation on early origins may suggest that these origins favor ORC sumoylation perhaps due to a relatively high potential to be sumoylated.

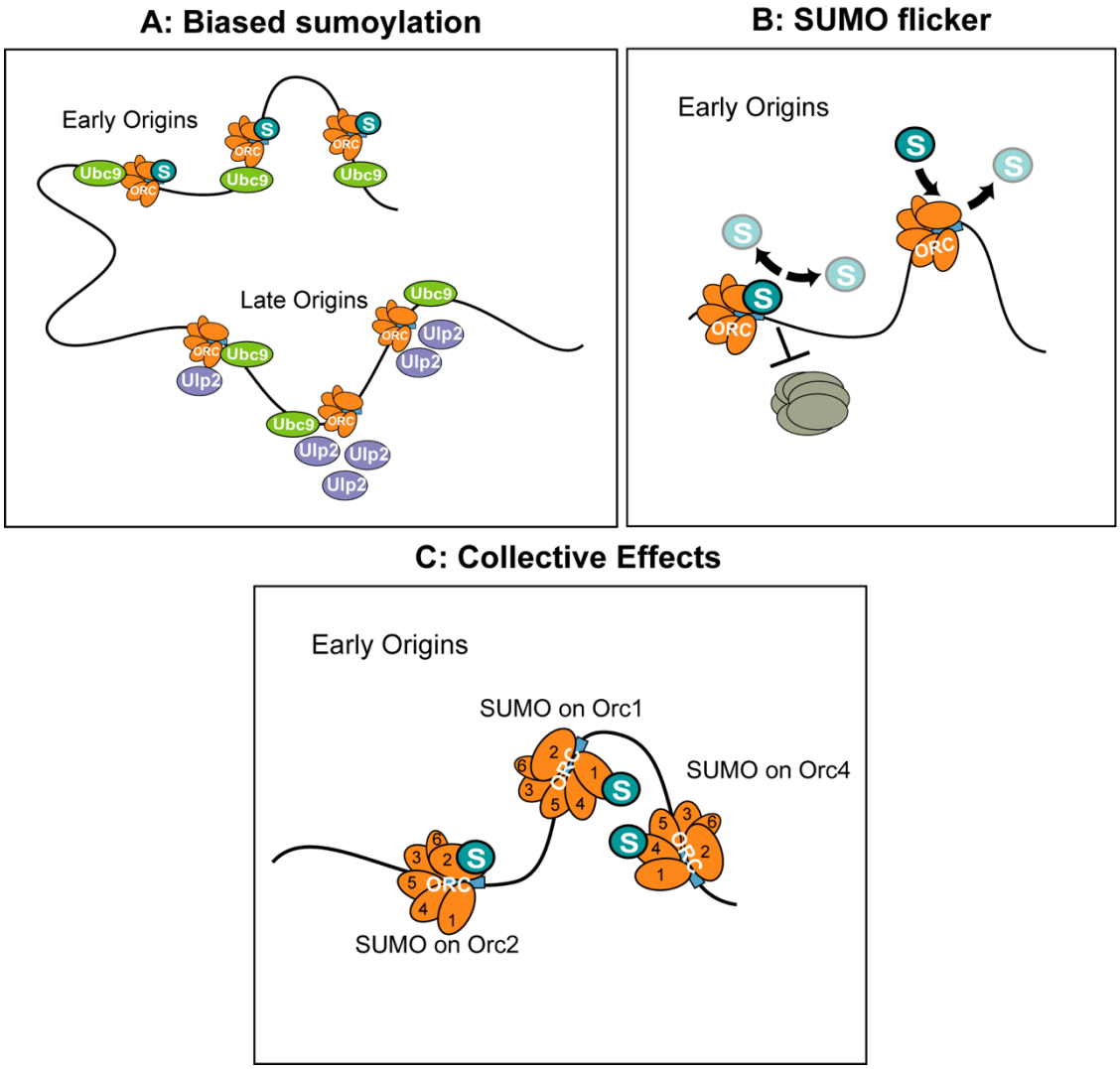


Figure 4.1 ORC sumoylation models

A. *Biased sumoylation in which early origins are sumoylated because there are differential levels of a desumoylation enzyme* ***B.*** *Sumo flicker model in which transient cycles of ORC sumoylation can prevent MCM loading, but only allow a small, modified portion to be observed at any time.* ***C.*** *Collective effects model in which sumoylation on different subunits works together to generate larger effects.*

Second, we envision a “SUMO flicker” model, wherein cycles of transient ORC sumoylation can reduce the time window for unmodified ORC to complete the multiple-step MCM loading process (**Figure 4.1B**). As described above, origin-bound ORC recruits Cdc6 followed by the MCM-Cdt1 complex to initiate MCM loading. Additional steps are required to generate two hexameric MCM rings capable of replisome assembly in S phase. These include ATP-hydrolysis mediated rejection of the first Cdc6 and Cdt1, followed by ‘ORC flipping’ wherein ORC dissociates from ACS while maintaining an interaction with MCM and binds to an adjacent B1 element to recruit a second set of Cdc6 and MCM-Cdt1(**Figure 1.1**) (Stillman 2022). Given the complexity of these steps, we envision that on-and-off cycles of ORC sumoylation could impede any one of these steps thus leading to overall negative effects on MCM loading.

We note that the two proposed models do not exclude means by which a small percentage of ORC sumoylation may achieve functional effects. For example, a “Collective Effects” model suggests that sumoylation of any one of the six ORC subunits could exert a negative effect on MCM recruitment (**Figure 4.1C**). This fits with our observation that increasing sumoylation of multiple ORC subunits by Orc2-SuON has a more profound effect than altering Orc2 sumoylation alone. Future work should determine the sumoylation sites on other ORC subunits to establish if these modifications are acting in concert with one another or have separable roles in regulating replication.

Similarly to what we observed here, a small percentage of human Orc2 is sumoylated in late S to G2/M phase. This modification prevents centromere re-replication by recruitment of the KDM5A histone deacetylase that favors the heterochromatic state of the centromere (Huang et al. 2016; Wang et al. 2017). Despite their mechanistic differences, human and yeast Orc2 both restrain origin usage, suggesting a general SUMO-based negative regulation of ORC functions and that appropriate sumoylation is important in both organisms for smooth progression through the cell cycle.

Together with previous findings, our data suggest that SUMO-based ORC and MCM sumoylation may provide a balance contrary to the DDK and CDK mediated positive regulation of origin activity. Similar complementary regulatory mechanisms have been seen in many other intricate processes such as the cell cycle and checkpoint regulation. Counter-balanced regulatory circuitry provides cells with the ability to fine-tune origin activity at different genomic loci, distinct times in S phase, and likely under different growth conditions and in distinct cell types. Future studies should test these models to determine how a low level of sumoylation can be functional and will help reveal how each could contribute to SUMO-mediated effects on origin firing.

4.3 Possible mechanisms by which ORC sumoylation interferes with origin usage

Examination of recently available structures of ORC bound to origin DNA alone, in complex with Cdc6, or in complex with Cdc6, Cdt1 and MCM suggests several not-mutually exclusive possibilities by which sumoylation at Orc2-K406 could affect MCM recruitment. *First*, given that K406 is located on the surface of ORC, and that SUMO is flexibly tethered to its substrate allowing the bulk of the modifier (~100 amino acids) to survey a large area, it could cause a steric hinderance for the incoming MCM complex. It has previously been shown that, for example, the SUMO that modifies the DNA clamp PCNA can adopt several extended flexible conformations (Tsutakawa et al. 2015). *Second*, comparing structures that capture snapshots of the different states of ORC revealed that the Orc2-WHD undergoes major conformational changes in order to accommodate the incoming MCM. It is possible, that sumoylation at K406 could interfere with one or more of these changes to prevent MCM loading. *Third*, in the 3D structure of the ORC-Cdc6, Orc2 K406 is located close to Cdc6, suggesting that sumoylation at this site could potentially interfere with Cdc6-ORC association, thus disfavoring MCM recruitment (**Figure 2.12e**). Future work should aim to test more directly, possibly through

further structural studies, how Orc2 sumoylation at this site could interfere with MCM loading, and which step of MCM double hexamer formation is impacted by sumoylation at this site.

4.4 Proper ORC sumoylation is required for optimal genomic stability and cell fitness

The multi-layered regulation of origin usage suggested by previous studies and the current work begs the question of whether each regulatory mechanism works independently to contribute to genome stability and cell fitness, or if they play largely redundant roles? In the case of ORC sumoylation, we found that up- or down-regulation of ORC sumoylation is sufficient to generate a dependence on other genome maintenance factors for cell fitness and increase the instability of the rDNA. Taken together, our data suggest that early origin suppression, conferred by even a single site sumoylation on Orc2 is enough affect genomic stability and cell fitness.

The increased early firing caused by loss of Orc2 sumoylation is similar to that seen in *sir2*Δ cells, which leads to delayed replication of late replicated regions (Hoggard et al. 2020). This shift is due to the fact that replication origins compete for a limiting amount of replisome assembly factors, thus when early origin usage is increased, late regions suffer from further delay in initiation, increasing the probably of not properly completing replication within the S phase (Hoggard et al. 2020). This imbalance in origin firing can be a source of DNA damage, and Sir2 mutants also show a negative interaction with Rad9 mutants similar to what we showed here (Mohanty et al. 2009). It is possible that these two mechanisms work concertedly to suppress early origin firing to enable a proper distribution of replisome assembly factors at late origins and thus complete whole genome replication. However, which early origins are affected by each of these proteins and whether there is overlap in the mechanisms involved has yet to be determined. Future work should test the interplay between Orc2 sumoylation and other

regulatory mechanisms that impact origin firing, including Sir2, but also Rpd3, the Orc1-BAH domain and the Fkh1/2 proteins.

Finally, we observed that Orc2 sumoylation is altered based on cell cycle stage. The levels of Orc2 sumoylation increase as cells progress through S-phase and decrease in G2 (**Figure 2.15**). These data are consistent with our model that sumoylation of ORC plays a regulatory role in origin control during S phase. However, this increase in late S-phase sumoylation also suggested to us that ORC sumoylation could play a role in regulating re-replication. In the future it will be important to conduct similar tests for the other ORC subunits to gain a comprehensive understanding of the timing for all ORC subunit sumoylation. It will also be interesting to examine if this sumoylation plays a role regulating re-replication by combining this orc2-K406R mutant with other alleles that disrupt the CDK controls of re-replication as described in 1.3.5.

4.6 Pol2 POPS contributes to the balance between polymerase and exonuclease function

The division of labor between the leading and lagging strand polymerases during genome duplication is fundamentally important for the stable inheritance of genetic information each time cells divide. The striking difference in the length of nascent DNA synthesized by the two polymerases argues that they must have distinct strategies optimized for their specific tasks. In particular, the leading strand polymerase Pol ε faces a challenge during continuous synthesis of long stretches of DNA. Previous work has revealed three Pol ε intrinsic mechanisms that stabilize its association with template DNA, its unique P-domain, its structural subunits, and its physical tethering to the CMG helicase (Hogg et al. 2014; Goswami et al. 2018). Collectively, these features can allow Pol ε to persistently associate with a template thus favoring continuous synthesis of long stretches of DNA. I contributed to a collaborative project

that reveals an additional strategy that also contributes to Pol ε mediated leading strand synthesis. This strategy is enabled by a highly conserved intrinsic feature of Pol ε not found in other polymerases and entails this POPS domain counterbalancing the action of its exonuclease domain and promoting replication elongation. Our data also adds to other recent data that support the importance of regulating the partition of the nascent primer between the POL and EXO domains of Pol ε in productive leading strand synthesis (Parkash et al. 2019; Xing et al. 2019).

Previous studies have noted that the highly conserved POPS domain is uniquely possessed by Pol2 family proteins; however, the function of this domain in replication elongation has remained unclear (Meng et al. 2020). Our single molecule replicon-seq analyses elucidated the features of sister replication forks across the entire genome. This data revealed a genome wide defect in replication fork symmetry caused by cancer-associated mutations in the POPS domain. In addition, the replicon-seq data indicate a leading strand synthesis defect in *pol2-REL* mutant cells. These findings provide strong evidence for an important role for POPS in supporting processive Pol ε-mediated DNA synthesis and replication fork symmetry across the genome. Future work should examine the interplay between this domain, its sumoylation and the exonuclease domain.

4.7 Mechanisms of POPS's roles in processive leading strand synthesis?

We showed in **Chapter 3** that replication progression factors, including Mrc1 and Csm3, remain associated with the replicative helicase in *pol2-REL* mutants. In addition, deleting the PCNA binding motif on Pol2 worsened *pol2-REL* growth, thus providing a genetic argument that POPS is unlikely to act via a PCNA or Mrc1/Csm3-mediated mechanism. In contrast to these negative findings, the suppression of multiple *pol2-REL* defects was conferred by abolishing

Pol2's exonuclease activity shedding important light onto the functions of both the POPS and the EXO domains. Using replicon-seq, we showed that replication fork asymmetry and impaired leading strand syntheses of *pol2-REL* cells were both improved by EXO inactivation. This observation supports a functional antagonism between the two domains.

This antagonism is unlikely to pertain to the canonical role of the Pol2 EXO domain in mutation avoidance, since *pol2-REL* showed a wild-type level of mutation rates (Meng et al. 2020). Rather, the observed functional relationship between the POPS and EXO domains can be explained by the recently suggested role of EXO in regulating the movement of the nascent primer from the catalytic sites to the EXO site even in the absence of mis-incorporated nucleotides (Parkash et al. 2019; Xing et al. 2019). In principle, EXO-mediated polymerase backtracking by itself could impede the forward movement of the enzyme thus retarding processive synthesis. When this occurs in a low dNTP situation, it could also lead to nascent strand degradation. Our data suggest that EXO-mediated Pol ε backtracking is countered by the POPS domain. How the POPS domain counterbalances such a role to favor forward polymerase movement is currently unclear and should be explored in future work. Given its peripheral localization seen in the currently available Pol ε structures (**Figure 3.2A**), POPS may influence the structural changes in the catalytic region to favor a conformation that can continue to engage with the nascent primer even when the polymerase is paused by template barriers. Future structural and biophysical studies will be needed to test this and other possible models.

It is also important to consider whether this balancing effect is specific to these two domains or if EXO inactivation generally increases fork progression. Previous studies have focused on the mutation phenotypes of the EXO mutation but less work has focused on if mutating this domain increases fork progression. If the EXO mutant generally leads to increased fork progression it could also rescue progression defects of other progression promoting factors such as Mrc1 or Csm3. There is some evidence that the Pol2 EXO mutant does not suppress

the slow growth of mrc1Δ cells but further study is necessary to explore this possibility (Williams et al. 2015).

Our data suggest that the proposed role of POPS in countering nascent primer movement to the EXO domain may be the most important when Pol ε is stalled by template barriers, such as RNA-DNA hybrids. The data presented do not exclude the possibility that the delicate balance between POPS and EXO functions occurs even without template barriers. It will be interesting to test this possibility in the future. Regardless, the significance of this functional interplay is evidenced by the suppression of other defects of *pol2-REL* by EXO inactivation, including slow genome replication and S phase progression, increased genome instability, and synthetic sickness with checkpoint and repair mutants.

The individual mutations studied in *pol2-REL* correspond to recurrent cancer-associated mutations in POLE and individually showed similar but milder defects as compared to the combined *pol2-REL* mutations (Meng et al. 2020). Our finding suggests that some of these cancer-associated mutations affect genome stability beyond the canonical model of POLE influencing cancer genomes via mutation avoidance. Future studies should examine the mechanisms by which these mutations, perhaps through larger genomic rearrangements, can contribute to cancer formation. Furthermore, the synthetic sickness profile uncovered for *pol2-REL*, such as with mutations inactivating checkpoint and repair proteins, may be able to inform strategies to selectively kill cancer cells bearing mutations in the POPS domain.

Overall, this work uncovers two novel mechanisms by which DNA replication is regulated, ORC sumoylation that affects replication initiation and the means by which the novel POPS domain contributes to replication elongation. This work also suggests lines of further study that can expand our understanding of how replication is regulated.

REFERENCES

Almedawar S, Colomina N, Bermúdez-López M, Pociño-Merino I, Torres-Rosell J. 2012. A SUMO-dependent step during establishment of sister chromatid cohesion. *Curr Biol* **22**: 1576-1581.

Aparicio OM, Weinstein DM, Bell SP. 1997. Components and dynamics of DNA replication complexes in S. cerevisiae: redistribution of MCM proteins and Cdc45p during S phase. *Cell* **91**: 59-69.

Attali I, Botchan MR, Berger JM. 2021. Structural mechanisms for replicating DNA in eukaryotes. *Annu Rev Biochem* **90**: 77-106.

Batrakou DG, Muller CA, Wilson RHC, Nieduszynski CA. 2020. DNA copy-number measurement of genome replication dynamics by high-throughput sequencing: the sort-seq, sync-seq and MFA-seq family. *Nat Protoc* **15**: 1255-1284.

Bayer P, Arndt A, Metzger S, Mahajan R, Melchior F, Jaenicke R, Becker J. 1998. Structure determination of the small ubiquitin-related modifier SUMO-1. *J Mol Biol* **280**: 275-286.

Bell SP, Labib K. 2016. Chromosome Duplication in Saccharomyces cerevisiae. *Genetics* **203**: 1027-1067.

Bell SP, Stillman B. 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**: 128-134.

Bernad A, Blanco L, Lazaro JM, Martin G, Salas M. 1989. A conserved 3'----5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell* **59**: 219-228.

Bleichert F, Botchan MR, Berger JM. 2017. Mechanisms for initiating cellular DNA replication. *Science* **355**.

Bonner JN, Choi K, Xue X, Torres NP, Szakal B, Wei L, Wan B, Arter M, Matos J, Sung P et al. 2016. Smc5/6 mediated sumoylation of the Sgs1-Top3-Rmi1 complex promotes removal of recombination intermediates. *Cell Rep* **16**: 368-378.

Breier AM, Chatterji S, Cozzarelli NR. 2004. Prediction of Saccharomyces cerevisiae replication origins. *Genome Biol* **5**: R22.

Brutlag D, Kornberg A. 1972. Enzymatic synthesis of deoxyribonucleic acid. 36. A proofreading function for the 3' leads to 5' exonuclease activity in deoxyribonucleic acid polymerases. *J Biol Chem* **247**: 241-248.

Burgers PMJ, Kunkel TA. 2017. Eukaryotic DNA Replication Fork. *Annu Rev Biochem* **86**: 417-438.

Capson TL, Peliska JA, Kaboord BF, Frey MW, Lively C, Dahlberg M, Benkovic SJ. 1992. Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4. *Biochemistry* **31**: 10984-10994.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E et al. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**: 401-404.

Chen S, Bell SP. 2011. CDK prevents Mcm2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. *Genes Dev* **25**: 363-372.

Cheng J, Li N, Huo Y, Dang S, Tye BK, Gao N, Zhai Y. 2022. Structural Insight into the MCM double hexamer activation by Dbf4-Cdc7 kinase. *Nat Commun* **13**: 1396.

Chilkova O, Stenlund P, Isoz I, Stith CM, Grabowski P, Lundstrom EB, Burgers PM, Johansson E. 2007. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res* **35**: 6588-6597.

Claussin C, Vazquez J, Whitehouse I. 2022. Single-molecule mapping of replisome progression. *Mol Cell* **82**: 1372-1382 e1374.

Clyne RK, Kelly TJ. 1997. Identification of autonomously replicating sequence (ARS) elements in eukaryotic cells. *Methods* **13**: 221-233.

Cosgrove AJ, Nieduszynski CA, Donaldson AD. 2002. Ku complex controls the replication time of DNA in telomere regions. *Genes Dev* **16**: 2485-2490.
Coster G, Frigola J, Beuron F, Morris EP, Diffley JFX. 2014. Origin Licensing Requires ATP Binding and Hydrolysis by the MCM Replicative Helicase. *Molecular Cell* **55**: 666-677.
Crabbe L, Thomas A, Pantesco V, De Vos J, Pasero P, Lengronne A. 2010. Analysis of replication profiles reveals key role of RFC-Ctf18 in yeast replication stress response. *Nat Struct Mol Biol* **17**: 1391-1397.
Cremona CA, Sarangi P, Yang Y, Hang LE, Rahman S, Zhao X. 2012. Extensive DNA damage-induced sumoylation contributes to replication and repair and acts in addition to the Mec1 checkpoint. *Mol Cell* **45**: 422-432.
Czajkowsky DM, Liu J, Hamlin JL, Shao Z. 2008. DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. *J Mol Biol* **375**: 12-19.
Daigaku Y, Keszthelyi A, Muller CA, Miyabe I, Brooks T, Retkute R, Hubank M, Nieduszynski CA, Carr AM. 2015. A global profile of replicative polymerase usage. *Nat Struct Mol Biol* **22**: 192-198.
Das SP, Borrman T, Liu VW, Yang SC, Bechhoefer J, Rhind N. 2015. Replication timing is regulated by the number of MCMs loaded at origins. *Genome Res* **25**: 1886-1892.
Das SP, Rhind N. 2016. How and why multiple MCMs are loaded at origins of DNA replication. *Bioessays* **38**: 613-617.
de Albuquerque CP, Liang J, Gaut NJ, Zhou H. 2016. Molecular Circuitry of the SUMO (Small Ubiquitin-like Modifier) Pathway in Controlling Sumoylation Homeostasis and Suppressing Genome Rearrangements. *J Biol Chem* **291**: 8825-8835.
De Ioannes P, Leon VA, Kuang Z, Wang M, Boeke JD, Hochwagen A, Armache KJ. 2019. Structure and function of the Orc1 BAH-nucleosome complex. *Nat Commun* **10**: 2894.
Debatisse M, Le Tallec B, Letessier A, Dutrillaux B, Brison O. 2012. Common fragile sites: mechanisms of instability revisited. *Trends Genet* **28**: 22-32.
Deshpande AM, Newlon CS. 1992. The ARS consensus sequence is required for chromosomal origin function in Saccharomyces cerevisiae. *Mol Cell Biol* **12**: 4305-4313.
Deutsch EW, Bandeira N, Sharma V, Perez-Riverol Y, Carver JJ, Kundu DJ, Garcia-Seisdedos D, Jarnuczak AF, Hewapathirana S, Pullman BS et al. 2020. The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res* **48**: D1145-D1152.
Devbhandari S, Remus D. 2020. Rad53 limits CMG helicase uncoupling from DNA synthesis at replication forks. *Nat Struct Mol Biol* **27**: 461-471.
Dhingra N, Wei L, Zhao X. 2019. Replication protein A (RPA) sumoylation positively influences the DNA damage checkpoint response in yeast. *J Biol Chem* **294**: 2690-2699.
Donaldson AD, Fangman WL, Brewer BJ. 1998. Cdc7 is required throughout the yeast S phase to activate replication origins. *Genes Dev* **12**: 491-501.
Duda DM, van Waardenburg RC, Borg LA, McGarity S, Nourse A, Waddell MB, Bjornsti MA, Schulman BA. 2007. Structure of a SUMO-binding-motif mimic bound to Smt3p-Ubc9p: conservation of a non-covalent ubiquitin-like protein-E2 complex as a platform for selective interactions within a SUMO pathway. *J Mol Biol* **369**: 619-630.
Dumas LB, Lussky JP, McFarland EJ, Shampay J. 1982. New temperature-sensitive mutants of Saccharomyces cerevisiae affecting DNA replication. *Mol Gen Genet* **187**: 42-46.
Duncker BP, Chesnokov IN, McConkey BJ. 2009. The origin recognition complex protein family. *Genome Biol* **10**: 214.
Early A, Drury LS, Diffley JF. 2004. Mechanisms involved in regulating DNA replication origins during the cell cycle and in response to DNA damage. *Philos Trans R Soc Lond B Biol Sci* **359**: 31-38.
Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM. 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev* **24**: 748-753.

Fang D, Lengronne A, Shi D, Forey R, Skrzypczak M, Ginalski K, Yan C, Wang X, Cao Q, Pasero P et al. 2017. Dbf4 recruitment by forkhead transcription factors defines an upstream rate-limiting step in determining origin firing timing. *Genes Dev* **31**: 2405-2415.
Feng X, Noguchi Y, Barbon M, Stillman B, Speck C, Li H. 2021. The structure of ORC-Cdc6 on an origin DNA reveals the mechanism of ORC activation by the replication initiator Cdc6. *Nat Commun* **12**: 3883.
Ferguson BM, Brewer BJ, Reynolds AE, Fangman WL. 1991. A yeast origin of replication is activated late in S phase. *Cell* **65**: 507-515.
Ferguson BM, Fangman WL. 1992. A position effect on the time of replication origin activation in yeast. *Cell* **68**: 333-339.
Forsburg SL. 2004. Eukaryotic MCM Proteins: Beyond Replication Initiation. *Microbiology and Molecular Biology Reviews* **68**: 109-131.
Friedel AM, Pike BL, Gasser SM. 2009. ATR/Mec1: coordinating fork stability and repair. *Curr Opin Cell Biol* **21**: 237-244.
Friedman KL, Brewer BJ, Fangman WL. 1997. Replication profile of Saccharomyces cerevisiae chromosome VI. *Genes Cells* **2**: 667-678.
Friedman KL, Diller JD, Ferguson BM, Nyland SV, Brewer BJ, Fangman WL. 1996. Multiple determinants controlling activation of yeast replication origins late in S phase. *Genes Dev* **10**: 1595-1607.
Fritze CE, Verschueren K, Strich R, Easton Esposito R. 1997. Direct evidence for SIR2 modulation of chromatin structure in yeast rDNA. *EMBO J* **16**: 6495-6509.
Gabrielse C, Miller CT, McConnell KH, DeWard A, Fox CA, Weinreich M. 2006. A Dbf4p BRCA1 C-terminal-like domain required for the response to replication fork arrest in budding yeast. *Genetics* **173**: 541-555.
Gan H, Yu C, Devbhandari S, Sharma S, Han J, Chabes A, Remus D, Zhang Z. 2017. Checkpoint Kinase Rad53 Couples Leading- and Lagging-Strand DNA Synthesis under Replication Stress. *Mol Cell* **68**: 446-455 e443.
Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E et al. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**: pl1.
Geiss-Friedlander R, Melchior F. 2007. Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* **8**: 947-956.
Golebiowski F, Matic I, Tatham MH, Cole C, Yin Y, Nakamura A, Cox J, Barton GJ, Mann M, Hay RT. 2009. System-wide changes to SUMO modifications in response to heat shock. *Sci Signal* **2**: ra24.
Goswami P, Abid Ali F, Douglas ME, Locke J, Purkiss A, Janska A, Eickhoff P, Early A, Nans A, Cheung AMC et al. 2018. Structure of DNA-CMG-Pol epsilon elucidates the roles of the non-catalytic polymerase modules in the eukaryotic replisome. *Nat Commun* **9**: 5061.
Green BM, Morreale RJ, Ozaydin B, Derisi JL, Li JJ. 2006. Genome-wide mapping of DNA synthesis in Saccharomyces cerevisiae reveals that mechanisms preventing reinitiation of DNA replication are not redundant. *Mol Biol Cell* **17**: 2401-2414.
Gros J, Kumar C, Lynch G, Yadav T, Whitehouse I, Remus D. 2015. Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. *Mol Cell* **60**: 797-807.
Gupta S, Friedman LJ, Gelles J, Bell SP. 2021. A helicase-tethered ORC flip enables bidirectional helicase loading. *Elife* **10**.
Hawkins M, Retkute R, Muller CA, Saner N, Tanaka TU, de Moura AP, Nieduszynski CA. 2013. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep* **5**: 1132-1141.
Hiraga S, Alvino GM, Chang FJ, Lian HY, Sridhar A, Kubota T, Brewer BJ, Weinreich M, Raghuraman MK, Donaldson AD. 2014. Rif1 controls DNA replication by directing

Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex. *Gene Dev* **28**: 372-383.
Hizume K, Araki H. 2019. Replication fork pausing at protein barriers on chromosomes. *FEBS Lett* **593**: 1449-1458.
Hoang ML, Leon RP, Pessoa-Brandao L, Hunt S, Raghuraman MK, Fangman WL, Brewer BJ, Sclafani RA. 2007. Structural changes in Mcm5 protein bypass Cdc7-Dbf4 function and reduce replication origin efficiency in Saccharomyces cerevisiae. *Mol Cell Biol* **27**: 7594-7602.
Hogg M, Osterman P, Bylund GO, Ganai RA, Lundstrom EB, Sauer-Eriksson AE, Johansson E. 2014. Structural basis for processive DNA synthesis by yeast DNA polymerase varepsilon. *Nat Struct Mol Biol* **21**: 49-55.
Hoggard T, Chang F, Perry KR, Subramanian S, Kenworthy J, Chueng J, Shor E, Cosgrove M, Boeke J, Fox CA et al. 2018. Yeast heterochromatin regulators Sir2 and SIr3 act directly at euchromatic DNA replication origins. *PLOS genetics*: 1-40.
Hoggard T, Hollatz AJ, Cherney RE, Seman MR, Fox CA. 2021. The Fkh1 Forkhead associated domain promotes ORC binding to a subset of DNA replication origins in budding yeast. *Nucleic Acids Res* **49**: 10207-10220.
Hoggard T, Muller CA, Nieduszynski CA, Weinreich M, Fox CA. 2020. Sir2 mitigates an intrinsic imbalance in origin licensing efficiency between early- and late-replicating euchromatin. *Proc Natl Acad Sci U S A* **117**: 14314-14321.
Hoggard T, Shor E, Muller CA, Nieduszynski CA, Fox CA. 2013. A link between ORC-origin binding mechanisms and origin activation time revealed in budding yeast. *Plos Genetics* **9**.
Huang C, Cheng J, Bawa-Khalfe T, Yao X, Chin YE, Yeh ETH. 2016. SUMOylated ORC2 recruits a histone demethylase to regulate centromeric histone modification and genomic stability. *Cell Rep* **15**: 147-157.
Hung CW, Martinez-Marquez JY, Javed FT, Duncan MC. 2018. A simple and inexpensive quantitative technique for determining chemical sensitivity in Saccharomyces cerevisiae. *Sci Rep* **8**: 11919.
Jackson LP, Reed SI, Haase SB. 2006. Distinct mechanisms control the stability of the related S-phase cyclins Clb5 and Clb6. *Mol Cell Biol* **26**: 2456-2466.
Jain R, Rajashankar KR, Buku A, Johnson RE, Prakash L, Prakash S, Aggarwal AK. 2014. Crystal structure of yeast DNA polymerase epsilon catalytic domain. *Plos One* **9**: e94835.
Johnson ES. 2004. Protein modification by SUMO. *Annu Rev Biochem* **73**: 355-382.
Khare V, Eckert KA. 2001. The 3' --> 5' exonuclease of T4 DNA polymerase removes premutagenic alkyl mispairs and contributes to futile cycling at O6-methylguanine lesions. *J Biol Chem* **276**: 24286-24292.
Knott SR, Peace JM, Ostrow AZ, Gan Y, Rex AE, Viggiani CJ, Tavare S, Aparicio OM. 2012a. Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae. *Cell* **148**: 99-111.
Knott SRV, Peace JM, Ostrow AZ, Gan Y, Rex AE, Viggiani CJ, Tavaré S, Aparicio OM. 2012b. Forkhead Transcription Factors Establish Origin Timing and Long-Range Clustering in *S. cerevisiae*. *Cell* **148**: 99-111.
Kobayashi T. 2011. Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cell Mol Life Sci* **68**: 1395-1403.
Kobayashi T, Sasaki M. 2017. Ribosomal DNA stability is supported by many 'buffer genes'-introduction to the Yeast rDNA Stability Database. *FEMS Yeast Res* **17**.
Kunkel TA, Roberts JD, Sugino A. 1991. The fidelity of DNA synthesis by the catalytic subunit of yeast DNA polymerase alpha alone and with accessory proteins. *Mutat Res* **250**: 175-182.

Kwan EX, Alvino GM, Queitsch C, Brewer BJ, Raghuraman MK. 2020. Small but greedy: a minimal rDNA array outcompetes the genome for replication initiation. *Faseb J* **34**.
Kwan EX, Foss EJ, Tsuchiyama S, Alvino GM, Kruglyak L, Kaeberlein M, Raghuraman MK, Brewer BJ, Kennedy BK, Bedalov A. 2013. A natural polymorphism in rDNA replication origins links origin activation with calorie restriction and lifespan. *PLoS Genet* **9**: e1003329.
Labib K. 2010. How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes Dev* **24**: 1208-1219.
Lai B. 1993. Pulsed Field Gel Electrophoresis: A Practical Guide. Academic Press.
Langston LD, Zhang D, Yurieva O, Georgescu RE, Finkelstein J, Yao NY, Indiani C, O'Donnell ME. 2014. CMG helicase and DNA polymerase epsilon form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc Natl Acad Sci U S A* **111**: 15390-15395.
Lee CSK, Cheung MF, Li J, Zhao Y, Lam WH, Ho V, Rohs R, Zhai Y, Leung D, Tye BK. 2021. Humanizing the yeast origin recognition complex. *Nat Commun* **12**: 33.
Lei M, Kawasaki Y, Young MR, Kihara M, Sugino A, Tye BK. 1997. Mcm2 is a target of regulation by Cdc7-Dbf4 during the initiation of DNA synthesis. *Genes Dev* **11**: 3365-3374.
Lengronne A, Pasero P. 2020. Sir2 takes affirmative action to ensure equal opportunity in replication origin licensing. *Proc Natl Acad Sci U S A* **117**: 16723-16725.
Li N, Lam WH, Zhai Y, Cheng J, Cheng E, Zhao Y, Gao N, Tye BK. 2018. Structure of the origin recognition complex bound to DNA replication origin. *Nature* **559**: 217-222.
Li S, Wasserman MR, Yurieva O, Bai L, O'Donnell ME, Liu S. 2022. Nucleosome-directed replication origin licensing independent of a consensus DNA sequence. *Nat Commun* **13**: 4947.
Liang C, Weinreich M, Stillman B. 1995. ORC and Cdc6p interact and determine the frequency of initiation of DNA replication in the genome. *Cell* **81**: 667-676.
Lynch KL, Alvino GM, Kwan EX, Brewer BJ, Raghuraman MK. 2019. The effects of manipulating levels of replication initiation factors on origin firing efficiency in yeast. *PLoS Genet* **15**: e1008430.
Mantiero D, Mackenzie A, Donaldson A, Zegerman P. 2011. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J* **30**: 4805-4814.
Meng X, Wei L, Devbhandari S, Zhang T, Xiang J, Remus D, Zhao X. 2020. DNA polymerase epsilon relies on a unique domain for efficient replisome assembly and strand synthesis. *Nat Commun* **11**: 2437.
Meng X, Wei L, Peng XP, Zhao X. 2019. Sumoylation of the DNA polymerase epsilon by the Smc5/6 complex contributes to DNA replication. *PLoS Genet* **15**: e1008426.
Miller TCR, Locke J, Greiwe JF, Diffley JFX, Costa A. 2019. Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. *Nature* **575**: 704-710.
Mohanty BK, Bairwa NK, Bastia D. 2009. Contrasting roles of checkpoint proteins as recombination modulators at Fob1-Ter complexes with or without fork arrest. *Eukaryot Cell* **8**: 487-495.
Muller CA, Nieduszynski CA. 2017. DNA replication timing influences gene expression level. *J Cell Biol* **216**: 1907-1914.
Muller P, Park S, Shor E, Huebert DJ, Warren CL, Ansari AZ, Weinreich M, Eaton ML, MacAlpine DM, Fox CA. 2010. The conserved bromo-adjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin. *Genes Dev* **24**: 1418-1433.

Myung K, Smith S, Kolodner RD. 2004. Mitotic checkpoint function in the formation of gross chromosomal rearrangements in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **101**: 15980-15985.
Natsume T, Muller CA, Katou Y, Retkute R, Gierlinski M, Araki H, Blow JJ, Shirahige K, Nieduszynski CA, Tanaka TU. 2013. Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. *Mol Cell* **50**: 661-674.
Newlon CS, Petes TD, Hereford LM, Fangman WL. 1974. Replication of yeast chromosomal DNA. *Nature* **247**: 32-35.
Nguyen VQ, Co C, Irie K, Li JJ. 2000. Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. *Curr Biol* **10**: 195-205.
Nguyen VQ, Co C, Li JJ. 2001. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* **411**: 1068-1073.
Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PM, Kunkel TA. 2008. Division of labor at the eukaryotic replication fork. *Mol Cell* **30**: 137-144.
Ogawa T, Okazaki T. 1980. Discontinuous DNA replication. *Annu Rev Biochem* **49**: 421-457.
Papouli E, Chen SH, Davies AA, Huttner D, Krejci L, Sung P, Ulrich HD. 2005. Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2p. *Mol Cell* **19**: 123-133.
Parkash V, Kulkarni Y, Ter Beek J, Shcherbakova PV, Kamerlin SCL, Johansson E. 2019. Structural consequence of the most frequently recurring cancer-associated substitution in DNA polymerase epsilon. *Nat Commun* **10**: 373.
Pellicano G, Al Mamun M, Jurado-Santiago D, Villa-Hernandez S, Yin X, Giannattasio M, Lanz MC, Smolka MB, Yeeles J, Shirahige K et al. 2021. Checkpoint-mediated DNA polymerase epsilon exonuclease activity curbing counteracts resection-driven fork collapse. *Mol Cell* **81**: 2778-2792 e2774.
Perez-Riverol Y, Bai J, Bandla C, Garcia-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M et al. 2022. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* **50**: D543-D552.
Perkins G, Diffley JF. 1998. Nucleotide-dependent prereplicative complex assembly by Cdc6p, a homolog of eukaryotic and prokaryotic clamp-loaders. *Mol Cell* **2**: 23-32.
Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605-1612.
Pohl TJ, Brewer BJ, Raghuraman MK. 2012. Functional centromeres determine the activation time of pericentric origins of DNA replication in Saccharomyces cerevisiae. *PLoS Genet* **8**: e1002677.
Poli J, Tsaponina O, Crabbe L, Keszthelyi A, Pantesco V, Chabes A, Lengronne A, Pasero P. 2012. dNTP pools determine fork progression and origin usage under replication stress. *EMBO J* **31**: 883-894.
Powers KT, Lavering ED, Washington MT. 2018. Conformational flexibility of ubiquitin-modified and sumo-modified PCNA shown by full-ensemble hybrid methods. *J Mol Biol* **430**: 5294-5303.
Pozo PN, Matson JP, Cole Y, Kedziora KM, Grant GD, Temple B, Cook JG. 2018. Cdt1 variants reveal unanticipated aspects of interactions with cyclin/CDK and MCM important for normal genome replication. *Mol Biol Cell* **29**: 2989-3002.
Psakhye I, Castellucci F, Branzei D. 2019. SUMO-Chain-Regulated proteasomal degradation timing exemplified in DNA replication initiation. *Mol Cell* **76**: 632-645 e636.
Pursell ZF, Isoz I, Lundstrom EB, Johansson E, Kunkel TA. 2007. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* **317**: 127-130.

Pursell ZF, Kunkel TA. 2008. DNA polymerase epsilon: a polymerase of unusual size (and complexity). *Prog Nucleic Acid Res Mol Biol* **82**: 101-145.
Putnam CD, Kolodner RD. 2010. Determination of gross chromosomal rearrangement rates. *Cold Spring Harb Protoc* **2010**: pdb prot5492.
Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL. 2001. Replication dynamics of the yeast genome. *Science* **294**: 115-121.
Reha-Krantz LJ. 2010. DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochim Biophys Acta* **1804**: 1049-1063.
Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JF. 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139**: 719-730.
Rhind N, Gilbert DM. 2013. DNA replication timing. *Cold Spring Harb Perspect Biol* **5**: a010132.
Ryba T, Battaglia D, Pope BD, Hiratani I, Gilbert DM. 2011. Genome-scale analysis of replication timing: from bench to bioinformatics. *Nat Protoc* **6**: 870-895.
Salim D, Bradford WD, Freeland A, Cady G, Wang J, Pruitt SC, Gerton JL. 2017. DNA replication stress restricts ribosomal DNA copy number. *PLoS Genet* **13**: e1007006.
Sanchez JC, Kwan EX, Pohl TJ, Amemiya HM, Raghuraman MK, Brewer BJ. 2017. Defective replication initiation results in locus specific chromosome breakage and a ribosomal RNA deficiency in yeast. *PLoS Genet* **13**: e1007041.
Saner N, Karschau J, Natsume T, Gierlinski M, Retkute R, Hawkins M, Nieduszynski CA, Blow JJ, de Moura AP, Tanaka TU. 2013. Stochastic association of neighboring replicons creates replication factories in budding yeast. *J Cell Biol* **202**: 1001-1012.
Santocanale C, Diffley JF. 1998. A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication. *Nature* **395**: 615-618.
Santocanale C, Sharma K, Diffley JF. 1999. Activation of dormant origins of DNA replication in budding yeast. *Genes Dev* **13**: 2360-2364.
Sarangi P, Zhao X. 2015. SUMO-mediated regulation of DNA damage repair and responses. *Trends Biochem Sci* **40**: 233-242.
Schepers A, Diffley JF. 2001. Mutational analysis of conserved sequence motifs in the budding yeast Cdc6 protein. *J Mol Biol* **308**: 597-608.
Schimmel J, Eifler K, Sigurethsson JO, Cuijpers SA, Hendriks IA, Verlaan-de Vries M, Kelstrup CD, Francavilla C, Medema RH, Olsen JV et al. 2014. Uncovering SUMOylation dynamics during cell-cycle progression reveals FoxM1 as a key mitotic SUMO target protein. *Mol Cell* **53**: 1053-1066.
Seki T, Diffley JF. 2000. Stepwise assembly of initiation proteins at budding yeast replication origins in vitro. *Proc Natl Acad Sci U S A* **97**: 14115-14120.
Sengupta S, van Deursen F, de Piccoli G, Labib K. 2013. Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Curr Biol* **23**: 543-552.
Shah K, Krishnamachari A. 2012. Nucleotide correlation based measure for identifying origin of replication in genomic sequences. *Biosystems* **107**: 52-55.
Shcherbakova PV, Pavlov YI, Chilkova O, Rogozin IB, Johansson E, Kunkel TA. 2003. Unique error signature of the four-subunit yeast DNA polymerase epsilon. *J Biol Chem* **278**: 43770-43780.
Sheu YJ, Stillman B. 2006. Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. *Mol Cell* **24**: 101-113.
-. 2010. The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature* **463**: 113-117.
Skene PJ, Henikoff S. 2015. A simple method for generating high-resolution maps of genome-wide protein binding. *Elife* **4**: e09225.

Smith L, Plug A, Thayer M. 2001. Delayed replication timing leads to delayed mitotic chromosome condensation and chromosomal instability of chromosome translocations. *Proc Natl Acad Sci U S A* **98**: 13300-13305.
Spacciapoli P, Nossal NG. 1994a. Interaction of DNA polymerase and DNA helicase within the bacteriophage T4 DNA replication complex. Leading strand synthesis by the T4 DNA polymerase mutant A737V (tsL141) requires the T4 gene 59 helicase assembly protein. *J Biol Chem* **269**: 447-455.
-. 1994b. A single mutation in bacteriophage T4 DNA polymerase (A737V, tsL141) decreases its processivity as a polymerase and increases its processivity as a 3'-->5' exonuclease. *J Biol Chem* **269**: 438-446.
Sriramachandran AM, Dohmen RJ. 2014. SUMO-targeted ubiquitin ligases. *Biochim Biophys Acta* **1843**: 75-85.
Stillman B. 2022. The remarkable gymnastics of ORC. *Elife* **11**.
Sun J, Evrin C, Samel SA, Fernandez-Cid A, Riera A, Kawakami H, Stillman B, Speck C, Li H. 2013. Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. *Nat Struct Mol Biol* **20**: 944-951.
Sun J, Kawakami H, Zech J, Speck C, Stillman B, Li H. 2012. Cdc6-induced conformational changes in ORC bound to origin DNA revealed by cryo-electron microscopy. *Structure* **20**: 534-544.
Syed S, Desler C, Rasmussen LJ, Schmidt KH. 2016. A Novel Rrm3 Function in Restricting DNA Replication via an Orc5-Binding Domain Is Genetically Separable from Rrm3 Function as an ATPase/Helicase in Facilitating Fork Progression. *PLoS Genet* **12**: e1006451.
Takayama Y, Kamimura Y, Okawa M, Muramatsu S, Sugino A, Araki H. 2003. GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev* **17**: 1153-1165.
Tanaka S, Araki H. 2011. Multiple regulatory mechanisms to inhibit untimely initiation of DNA replication are important for stable genome maintenance. *PLoS Genet* **7**: e1002136.
Tanaka S, Araki H. 2013. Helicase Activation and Establisment of Replication Forks at Chromosomal Origins of Replication. *Cold Spring Harbor Perspectives in Biology* **5**.
Tanaka S, Nakato R, Katou Y, Shirahige K, Araki H. 2011. Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Curr Biol* **21**: 2055-2063.
Tanny RE, MacAlpine DM, Blitzblau HG, Bell SP. 2006. Genome-wide analysis of re-replication reveals inhibitory controls that target multiple stages of replication initiation. *Mol Biol Cell* **17**: 2415-2423.
Taylor MRG, Yeeles JTP. 2019. Dynamics of Replication Fork Progression Following Helicase-Polymerase Uncoupling in Eukaryotes. *J Mol Biol* **431**: 2040-2049.
Theis JF, Newlon CS. 1994. Domain B of ARS307 contains two functional elements and contributes to chromosomal replication origin function. *Mol Cell Biol* **14**: 7652-7659.
-. 1997. The ARS309 chromosomal replicator of Saccharomyces cerevisiae depends on an exceptional ARS consensus sequence. *Proc Natl Acad Sci U S A* **94**: 10786-10791.
Thu YM, Bielinsky AK. 2013. Enigmatic roles of Mcm10 in DNA replication. *Trends Biochem Sci* **38**: 184-194.
Tsutakawa SE, Yan C, Xu X, Weinacht CP, Freudenthal BD, Yang K, Zhuang Z, Washington MT, Tainer JA, Ivanov I. 2015. Structurally distinct ubiquitin- and sumo-modified PCNA: implications for their distinct roles in the DNA damage response. *Structure* **23**: 724-733.
Tuduri S, Tourriere H, Pasero P. 2010. Defining replication origin efficiency using DNA fiber assays. *Chromosome Res* **18**: 91-102.
Ulrich HD. 2009. **497**.

Ulrich HD, Davies AA. 2009. In vivo detection and characterization of sumoylation targets in Saccharomyces cerevisiae. *Methods Mol Biol* **497**: 81-103.
Vogelauer M, Rubbi L, Lucas I, Brewer BJ, Grunstein M. 2002. Histone acetylation regulates the time of replication origin firing. *Mol Cell* **10**: 1223-1233.
Vujcic M, Miller CA, Kowalski D. 1999. Activation of silent replication origins at autonomously replicating sequence elements near the HML locus in budding yeast. *Mol Cell Biol* **19**: 6098-6109.
Wan B, Wu J, Meng X, Lei M, Zhao X. 2019. Molecular Basis for Control of Diverse Genome Stability Factors by the Multi-BRCT Scaffold Rtt107. *Mol Cell* **75**: 238-251 e235.
Wang R, Liu F, Zhao Y, Wu D, Chen L, Yeh ETH, Huang C. 2017. Reversible regulation of ORC2 SUMOylation by PIAS4 and SENP2. *Oncotarget* **8**: 70142-70155.
Watanabe Y, Maekawa M. 2010. Spatiotemporal regulation of DNA replication in the human genome and its association with genomic instability and disease. *Curr Med Chem* **17**: 222-233.
Wei L, Zhao X. 2016a. A new MCM modification cycle regualted DNA replication inititation. *Nature Structural and Moecular Biology* **23**: 209-216.
-. 2016b. A new MCM modification cycle regualted DNA replication inititation *Nat Struct Mol Biol* **23**: 209-216.
-. 2017. Roles of SUMO in Replication Initiation, Progression, and Termination. *Advances in Experimental Medicine and Biology* **1042,**: 371-393.
Williams LN, Marjavaara L, Knowels GM, Schultz EM, Fox EJ, Chabes A, Herr AJ. 2015. dNTP pool levels modulate mutator phenotypes of error-prone DNA polymerase epsilon variants. *Proc Natl Acad Sci U S A* **112**: E2457-2466.
Wu PY, Nurse P. 2009. Establishing the program of origin firing during S phase in fission Yeast. *Cell* **136**: 852-864.
Xing X, Kane DP, Bulock CR, Moore EA, Sharma S, Chabes A, Shcherbakova PV. 2019. A recurrent cancer-associated substitution in DNA polymerase epsilon produces a hyperactive enzyme. *Nat Commun* **10**: 374.
Xu J, He Y, Qiang B, Yuan J, Peng X, Pan XM. 2008. A novel method for high accuracy sumoylation site prediction from protein sequences. *BMC Bioinformatics* **9**: 8.
Yabuki N, Terashima H, Kitada K. 2002. Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* **7**: 781-789.
Yeeles JT, Deegan TD, Janska A, Early A, Diffley JF. 2015. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**: 431-435.
Yeeles JTP, Janska A, Early A, Diffley JFX. 2017. How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication. *Molecular Cell* **65**: 105-116.
Yoshida K, Bacal J, Desmarais D, Padioleau I, Tsaponina O, Chabes A, Pantesco V, Dubois E, Parrinello H, Skrzypczak M et al. 2014. The histone deacetylases sir2 and rpd3 act on ribosomal DNA to control the replication program in budding yeast. *Mol Cell* **54**: 691-697.
Yoshida K, Poveda A, Pasero P. 2013. Time to be versatile: regulation of the replication timing program in budding yeast. *J Mol Biol* **425**: 4696-4705.
Yuan Z, Georgescu R, Bai L, Zhang D, Li H, O'Donnell ME. 2020. DNA unwinding mechanism of a eukaryotic replicative CMG helicase. *Nat Commun* **11**: 688.
Yuan Z, Riera A, Bai L, Sun J, Nandi S, Spanos C, Chen ZA, Barbon M, Rappsilber J, Stillman B et al. 2017. Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat Struct Mol Biol* **24**: 316-324.
Zegerman P, Diffley JF. 2007. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* **445**: 281-285.
-. 2010. Checkpoint-dependent inhibition of DNA replication initiation by Sld3 and Dbf4 phosphorylation. *Nature* **467**: 474-478.

Zhao X. 2018. SUMO-mediated regulation of nuclear functions and signaling processes. *Mol Cell* **71**: 409-418.
Zhao X, Blobel G. 2005. A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. *Proc Natl Acad Sci U S A* **102**: 4777-4782.
Zhou JC, Janska A, Goswami P, Renault L, Abid Ali F, Kotecha A, Diffley JFX, Costa A. 2017. CMG-Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome. *Proc Natl Acad Sci U S A* **114**: 4141-4146.